Mental Representations Underlying Syllogistic Reasoning

Catherine Lucie Ardin

Ph.D.

University of Edinburgh
1991

To Alec

Declaration

I declare that this thesis has been composed by myself and that the research reported therein has been conducted by myself unless otherwise indicated.

Cath Ardin

Edinburgh, March 26, 1991

Acknowledgements

There are a number of people to whom thanks is more than due.

Keith Stenning for his patience and forebearance in the early, difficult stages and advice later on, when it could do some good. Ken Gilhooley, Norman Wetherick and Guy Politzer for their comments and discussion.

Many people in and around the Center have allowed me to bore them with practical and theoretical issues, especially Judy Delin, Carol Foster, Denise Neapolitan, Mike Oaksford and Joe Levy. Special thanks to the Computing Support team and Robert for answering all my irritating little queries and problems and keeping the system up in those last crucial stages. Peter Yule and Morten Christiansen were invaluable data-collectors and extra thanks to Peter for his help in collating the data. Thank you also to Margaret Rennex who had the job of putting hundreds of data cases onto disc. I salute all my subjects, many of whom were friends, for doing all those syllogisms for me.

The contribution of my family and friends cannot be overestimated, especially Margaret who was a wonderful Granny to Harry and gave me some much needed breaks. A special thank you also to Harry for being very patient with Mummy when she had to work so hard.

And thank you (in dog) to Alec. I couldn't have done it without you.

Abstract

This thesis uses the data from two experiments to investigate the mental representations underlying syllogistic reasoning. The first takes the standard syllogisms task but adds a recall component, the second consists of an entirely new syllogistic reasoning paradigm: seeking valid individuals.

The main effects found centre on the explanation for the figural effect and the status of agglomerated representations as an explanation for responses to syllogisms. Subjects who do not use an agglomerated representation show a figural effect - this throws doubt on explanations of this phenomenon based on the transitive arrangement of an agglomerated representation. Analysis reveals that subjects seem to interpret the syllogism at a linguistic level initially and derive a likely topic for the argument of the syllogism.

The representation must also include information about the potential of the syllogism to "cancel" the middle term. If the syllogism does not cancel, then this will either cause effects similar to those found for indeterminate texts or with the individuals task help subjects to determine whether syllogisms possess a valid conclusion or not.

Strong order effects for positive and negative information are found in both experiments and this preference also has an effect on the reasoning data, similar to the figural effect. The information in the syllogism must therefore be reordered before a conclusion is given, to satisfy this preference.

Many subjects do appear to use an agglomerated representation plus cycles of testing to solve syllogisms, but the representation they use must contain linguistic information about topic, and will be reordered so that positive information precedes negative. Agglomerated representations are probably equivalent at the level of coding common to all subjects, the differences apparent to introspection will depend on the particular instantiation chosen by the subject.

Contents

1.1 Human Reasoning 1
1.1.1 Reasoning and resource limitation 2
1.1.2 Reasoning and context 3
1.2 Investigating Human Reasoning 5
1.3 Syllogisms 7
1.4 Theories of Syllogistic Reasoning 8
1.4.1 Non-Logical Mechanisms and Errors of Interpretation 9
1.4.2 The deductive approach 13
1.4.3 Evaluating the hypotheses 17
1.4.4 The explanation of the figural effect. 22
1.4.5 Explaining quantifier misinterpretation 25
1.5 Logical and non-logical strategies 26
1.6 Memory and Syllogistic Reasoning 29
1.7 Agenda 33

2 Method **35**
2.1 Experiment One 35
2.1.1 Rationale 35
2.1.2 Subjects 36
2.1.3 Materials 36
2.1.4 Design 37
2.1.5 Presentation 38
2.1.6 Summary of the data collected 41
2.2 Experiment Two 41
2.2.1 Subjects 42
2.2.2 Materials 42
2.2.3 Design 42
2.2.4 Presentation 42
2.2.5 Analysis 43

3 Preliminary Investigation of the Recall Data **44**
3.1 Recall of the Quantifiers 44
3.1.1 Quantifier pairs by recall errors 48
3.1.2 Quantifier errors and the recall menu 52
3.1.3 Summary of the quantifier errors 55
3.2 Recall of Noun Order 55
3.2.1 Recall variables used 55
3.2.2 A preliminary model of recall of noun order 61

3.3 Processing the Figure 65
3.4 Interactions Between Recall Variables 67
3.5 Summary 69

4 The Reasoning Data **72**
4.1 General Findings: Experiment One 72
4.1.1 Syllogisms with no valid conclusion 73
4.2 General Effects: Experiment Two 75
4.2.1 Individuals where no valid conclusion exists 76
4.3 The Figural Effect 79
4.4 The Atmosphere Effect 80
4.4.1 Matching 81
4.4.2 The relevance of the atmosphere effect 82
4.4.3 Conclusion types and quantifier order 85
4.4.4 The figural effect - a summary 87
4.5 Strategies for Syllogisms Without a Conclusion 88
4.5.1 The individuals experiment 90
4.5.2 The standard task 91
4.6 Summary 94

5 Testing Hypotheses **97**
5.1 General Recall Effects 97
5.2 The Figural Effect 101
5.2.1 The Mental Models explanation of the figural effect. 101
5.2.2 Predictions for recall 102
5.2.3 Other explanations for the figural effect. 107
5.2.4 Comparing the mental models and grammatical roles explanations 110
5.2.5 Other strategies affecting the choice of conclusion. 111
5.3 Misinterpretation of the Quantifiers 114
5.3.1 Conversions 114
5.3.2 Other types of misinterpretation 117
5.4 Recall and Features of the Representation 118
5.5 Summary 121

6 Implications for Syllogistic Reasoning **125**
6.1 Reasoning and Recall 125
6.2 Ordering of Information in the Syllogism 128
6.3 Positive and Negative Information 134
6.3.1 Negative information and the recall data 134
6.3.2 Negative information and choosing individuals 135
6.3.3 The overall effect of negative information 136
6.4 Processing the Information in the Syllogism 138
6.4.1 Recall of the Figures and Types of Conclusion 141
6.5 The Nature of the Representation 144
6.6 Summary 147

7 Conclusions **149**
7.1 Reasoning effects 149

7.2 The Figural Effect 150
7.3 The move to positive information followed by negative information . . . 152
7.4 The Nature of the Representation 153
7.5 Implications for Future Research 154
7.5.1 The Figural Effect 154
7.5.2 Positive and Negative Information 155
7.5.3 Matchers Versus Non-matchers 156

A Experiment One **162**

B Experiment Two **167**

C Conclusions, Cancellation and Double-Hatching **172**

D Instructions to Subjects - Experiment One **175**

E Instructions to Subjects - Experiment Two **180**

List of Figures

1.1 An example syllogism 8
1.2 Possible components of a syllogism 8
1.3 Politzer's formulation: an example 10
1.4 Drawing a conclusion by Politzer's method 11
1.5 The four quantifiers in Euler's circles notation 15
1.6 Part of Revlis' flow chart for syllogistic reasoning 18
1.7 Stenning's version of Euler's circles 22
1.8 Combining premises with Euler's circles 23
1.9 Possible interpretation of "All of the A are B, some of the B are C" . . 27
1.10 Mental Models representation of "All of the B are A, all of the B are C 28
1.11 Combining Stenning's variation on Euler's circles 30
1.12 Syllogisms that are not double-hatched but that have valid conclusions . 31
1.13 Syllogisms that have a valid individual but no valid conclusion 31
1.14 Syllogisms that cancel 32
1.15 Syllogisms that do not cancel 32

2.1 First example syllogism 37
2.2 Second example syllogism 37
2.3 A sample conclusion 39
2.4 A sample recall menu 39
2.5 A sample menu at delayed recall 40

3.1 Mean recall errors for the quantifiers 45
3.2 Mean errors for the first quantifier 46
3.3 Mean errors for the second quantifier 47
3.4 Actual quantifier by chosen quantifier 49
3.5 Errors for quantifier pairs, immediate recall 50
3.6 Recall errors for quantifier pairs, delayed recall 50
3.7 Feature change: absolute numbers and percentage 52
3.8 Percentage choices for quantifiers 53
3.9 Percentage of overall choice 55
3.10 Mean recall errors for noun order 56
3.11 Mean error for noun assignment 58
3.12 Mean error for figure 59
3.13 Actual figure by recalled figure 60
3.14 Percentage of total error for Type 2 errors 62
3.15 Percentage of total errors for Type 3 errors 63
3.16 Percentage of total errors that are Type 4 64

3.17 Information preserved or lost by noun order error types 65
3.18 Percentage noun order errors by quantifier error 67
3.19 Percentage of correctly recalled texts by recalled quantifiers 68
3.20 Percentage of recalls with each noun as subject by total, recall without quantifier error and negative to positive changes 70

4.1 Percentage correct conclusions by model: Experiment One 73
4.2 Correct responses for syllogisms with no valid conclusion 74
4.3 Johnson-Laird and Steedman's data 74
4.4 Number of correct responses for None/Some..not syllogisms 74
4.5 Types of individual for Experiment Two 77
4.6 Percentage correct responses by number of models: Experiment Two . . 77
4.7 Syllogisms with no valid conclusion but a valid individual 78
4.8 Percentage correct responses for syllogisms with no valid conclusion, but where in some cases an individual does exist 78
4.9 Choice of conclusion by figure: Experiment One 79
4.10 Percentage of each noun in first position by figure: Experiment Two . . 79
4.11 Atmosphere response as a percentage of incorrect responses 80
4.12 The figural effect by matching and non-matching subjects 84
4.13 Error rate for conclusions by matcher and non-matchers 85
4.14 Frequencies of conclusion type 86
4.15 Percentage of responses by quantifier order 87
4.16 Stenning's version of Euler's circles 89
4.17 Syllogisms with a valid conclusion but no double hatching 90
4.18 Percentage of incorrect responses by cancellation 91
4.19 Percentage of incorrect responses by cancellation and presence of a valid individual 91
4.20 Frequencies of response types for selected syllogisms 92
4.21 Percentage incorrect responses by double hatching for syllogisms with no valid conclusion 92
4.22 Reading times of the second premise by type of syllogism 94

5.1 Percentage error by first and second premise 98
5.2 Percentage of recall errors by models for correctly answered syllogisms . 100
5.3 Percentage of recall errors by models for correctly answered syllogisms of AE and EA type 100
5.4 Mental models processing for the figural effect 103
5.5 Percentage of noun errors by figure 104
5.6 Percentage of each figure type chosen when an error in the figure was made at delayed recall 105
5.7 Processing time by figure 106
5.8 Percentage premise reversals 106
5.9 Percentage of recalls where the noun order in one or both premises was reversed by figure 108
5.10 Percentage recall by conclusion type 110
5.11 Predictions for mental models and grammatical roles accounts 111
5.12 Percentage recall errors for the nouns of the syllogism by conclusion type and figure 112

5.13 Mean choice of role for nouns 113
5.14 Mean errors for conforming and non-conforming syllogisms 114
5.15 Percentage of type 2 errors for each quantifier 115
5.16 Percentage conversions for each quantifier 116
5.17 Percentage of confusions for quantifiers 118
5.18 Recall errors and reading times for the second premise by the presence or absence of cancellation 119
5.19 Recall errors and reading times for the second premise by the presence or absence of double-hatching 119
5.20 Examples of syllogisms with no valid conclusion with combinations of cancelling and double-hatching 122
5.21 Recall errors, Incorrect responses and reading times for the second premise by the presence or absence of double-hatching and cancellation for syllogisms with no valid conclusion 123
5.22 An example of a non-cancelling syllogism 123

6.1 Percentage error by types of response 126
6.2 First position noun for the individuals task 129
6.3 First position noun by figure for the individuals task 129
6.4 Most popular types of individuals for AE/IE and EA/EI syllogisms by figure 132
6.5 Possible noun orders by individuals 132
6.6 Percentage of recalls with A in the second premise 133
6.7 Percentage of recalls with C in the first premise 133
6.8 Percentage of recalls with B in first position for the EA/EI syllogisms . 133
6.9 Errors in recall by preferred and non-preferred quantifier order 135
6.10 Percentage of type of noun order for each individual 137
6.11 Position of nouns by negative terms 137
6.12 Processing time by figure 140
6.13 Choice of topic for each figure according to the factors identified by Keiras 141
6.14 Percentage of correctly recalled noun orders by figure and conclusion given 142
6.15 Percentage of type of figure chosen by response given to syllogism 143
6.16 A comparison of Euler's circles and mental models representations for the quantifiers "some" and "some..not" 146

A.1 Responses for the figure AB–BC: Experiment One 163
A.2 Responses for the figure BA–CB: Experiment One 164
A.3 Responses for the figure AB–CB: Experiment One 165
A.4 Responses for the figure BA–BC: Experiment One 166

B.1 Responses for the figure AB–BC: Experiment Two 168
B.2 Responses for the figure BA–CB: Experiment Two 169
B.3 Responses for the figure AB–CB: Experiment Two 170
B.4 Responses for the figure BA–BC: Experiment Two 171

C.1 Features of syllogisms with figure AB–BC 173
C.2 Features of syllogisms with figure BA–CB 173
C.3 Features of syllogisms with figure AB–CB 174

C.4 Features of syllogisms with figure BA–BC 174

List of Tables

3.1 Type 1 recall 61
3.2 Type 2 error 62
3.3 Type 3 error 63
3.4 Type 4 error 64

1.1 Human Reasoning

The study of mental processes is as old as the discipline of psychology. Early work on the structure of memory and some of the earliest strictly experimental work in psychology was carried out by Ebbinghaus (1964, originally published in 1885) who laid the foundations for much of the later work in the field and who first yielded the observation that familiar and meaningfully connected material is easier to commit to memory than material which is meaningless and unconnected. Later work by Bartlett (1932) demonstrated that memory is constructive - that when attempting to recall rather abstruse material, people will make errors that demonstrate that they are actively reconstructing what they recall having read or seen to fit the information in better with their own conceptions or experience.

Theories of learning, the process of which memory is the result, have been strongly influenced by the behaviourist school of psychology which originates at least in part from the work of Pavlov on conditioned reflexes in dogs. These theories have in common a conception of human behaviour as being the predictable response to an external stimulus. The early behaviourists (Skinner for example) professed themselves interested only in the observable elements of behavious and not in the part between stimulus and response, the element of cognition. As experimental techniques have been refined it has become possible to make more and more of that hidden process observable and more and more complex descriptive models, such as mathematical models for example, have been used to characterise the learning process, but the underlying model of cognition as a process of taking information (stimulus) and transforming it (response) has persisted. The similarity of this to the way information is processed by a computer has led to efforts to formulate human thinking based on the methods used to program computer systems.

The information-processing conception of human reasoning relies on elements of computer programming to explain how humans might go about solving problems, understanding language or even negotiate social exchanges. The first significant attempt to simulate a human cognitive process in terms of computer architecture was made by Newell and Simon (1956) who developed an information processing model to prove theorems in symbolic logic and subsequent research has found much of value in the computer metaphor, as an example of information processing that is transparent to research in the best behaviourist tradition, by virtue of the fact that it is a human artefact.

1.1.1 Reasoning and resource limitation

As the metaphor of the serial computer has become popular in cognitive psychology, so its limitations as an explanatory model have become apparent. There are many aspects of human cognitive performance that simply cannot be captured on a computer of this type. One of the main reasons for this lies in one of the most basic features of the machine: its memory. A serial computer will forget nothing that it is not programmed to delete and possesses no intrinsic facility for processing the information it receives, so as to create a memory store that is meaningfully structured. Although humans also possess a possibly infinite store for long term events, the information within it is arranged so as to be accessible from a number of different points and as the result of a number of different searches. The part of our memory that does not act as a permanent store - working memory - has a very limited size. In a now-famous paper, Miller (1957) established that working memory could contain seven units +/- two units of information at any given time. It was already known that working memory spontaneously starts to decay after a few seconds without rehearsal to maintain the information in it. These phenomena can only be replicated in a serial computer by artificially creating limits on memory size and duration and no way has yet been found of simulating the transfer from working to long-term memory in such a way as to structure the information in the appropriate form.

The importance of the limitations on the computer metaphor becomes very obvious when attempts are made to model human reasoning on a serial machine[1]. There are certain tasks that humans find particularly difficult:- complex mental arithmetic, for example, and the reason for this difficulty appears to lie in the limits on working memory. In the case of mental arithmetic, we are incapable of computing the next stage in the process before the information from the preceding stage has started to fade. As a contrast, a serial computer, once properly programmed has no difficulty at all with such tasks as there are no limits on its memory, hence the popularity of pocket calculators. Conversely there are human functions such as understanding speech, which we perform

[1]A serial computer is taken to mean a computer programmed as a serial computer

effortlessly, but which have resisted attempts at computer simulation to the present day. While we may have developed special perceptual mechanisms for such processes. the form of our storage and retrieval mechanisms in long-term memory must also play a part. This should come as no surprise given that the earliest research by Bartlett. already referred to above, found exactly that human recall is not a simple matter of certain items just being deleted from memory. perhaps by interference, but of elaboration and reconstruction. Further, work such as that by Sachs (1967) and Jarvella (1971) demonstrates that not all information is treated equally in memory - that some information is more readily lost than others and this in turn has led in part to a model of language processing and memory being undertaken at a series of levels (for example Kintsch (1974)) with syntactic information being lost from memory rather readily but the essential semantic content of the information being preserved. This taken with the finding of Bransford and Franks (1971) that language is only readily understandable when it is presented in a context that makes it meaningful tends to work against the idea that human processing works similarly to simple computer models which take in information without surrounding context, treating all parts of the message with equal importance.

It is clear that an understanding of memory is essential for a proper understanding of cognitive functions. Further, to understand why humans find certain reasoning tasks difficult, we must appeal to the notion of resource limitation, the operation of context and the prioritisation of information for storage in memory as sources of error.

1.1.2 Reasoning and context

In the previous section reference was made to the fact that a language function such as speech recognition was difficult to model on a serial computer and that this was probably due to some characteristic features of human memory. It may be the case that in order to successfully carry out the functions of language, which are immensely complicated, we have developed particular kinds of mental systems, which function not on the basis of logical operations, but on the basis of complex heuristics derived from repeated experience. Language, being our most effective means of communication has developed in such a way as to reduce ambiguity in the message to be transmitted. However the complexity of the situations we attempt to communicate about means that the receiver of language must attempt to exclude ambiguity as much as the deliverer. This can be done by making certain assumptions; these might be about the communicative intentions of the speaker, derived from a knowledge of the speaker; the situation one is in (the context of the speech) and certain conventions all users of language assume to be in operation, summarised by Grice (1975) in the form of conversational implicatures. All this information and the inferences that can be drawn from it must be built up from repeated language experiences and stored in such a way that everything relevant to a conversation can be activated once speech begins.

Because language has such a central importance to humans, it is reasonable to suppose that the processes we use for language function affect other aspects of our mental functioning that, while not directly derived from language, are related to it. Theories of reasoning have found that, apart from error due to resource limitation, humans adopt a particular perspective to some logical problems that lead them to draw non-logical inferences. One example of this is the selection task extensively studied by Wason (1965). This problem consists of subjects being presented with four cards each of which has something printed on both sides. But they are only able to see one side of each card. They are asked to verify a rule which states a condition about the nature of the items printed on the cards. They are asked to do this by indicating the smallest number of cards that they would need to turn over to verify or falsify the rule. In the abstract condition, the rule might be "if there is a vowel on one side of the card, there must be an odd number on the other". Subjects have four cards, one with a vowel; one an odd number; one an even number and one a consonant on the visible side. The overwhelming response from naive subjects is to choose the card with the vowel on one side and the card with the odd number. Clearly subjects are making the inference that the rule is reversible, that if p->q then q->p. This is not warranted by the rule. Subjects also fail to make the inference that not q-> not p and therefore fail to check this by turning over the card with the even number to see that it does not have a vowel on its reverse. Once subjects have discovered the correct response they are able to perform the task correctly without difficulty, but their initial approach to the problem is very often incorrect.

This finding suggests that subjects are unable to use standard logic appropriately. Evans (1973) found that rather than using a verficationist strategy, subjects were choosing cards on the basis of "matching" the items in the rule to those on the card indicating that subjects are using a non-logical strategy to solve the problem. The effect of the content of the problem on performance has also been studied, with the finding that, although using real-life materials aids reasoning it does not do so in every paradigm. A comprehensive study of this problem and conditional reasoning in general by Oaksford (1988) has shown that subjects' failure to grasp the requirements of the task and the relationship of content to reasoning can be explained by a model of the reasoner as searching for a suitable context in which to situate the rule[2]. Falsification is not used as a strategy unless taught as a method because it is not a useful means of communicating information. The ability to use falsification is increased by situating the rule in context where the reasoner has had the opportunity to experience the negated implication of the rule

Human reasoning does not proceed along the lines of predicate logic, therefore, partly because we are unable to keep in mind all the ramifications that such problems require and partly because we bring to reasoning problems a particular set of strategies and priorities that have been developed for successful language use. To understand human performance at a reasoning problem it is crucial to identify the effect these strategies might have and to bear in mind that any given reasoning problem will first-and-foremost be interpreted as an attempt at communication.

1.2 Investigating Human Reasoning

The discussion above should serve to illustrate that the interface between natural language processing and reasoning is deserving of further investigation. Firstly there is the observation that people do not always adopt a rational strategy when faced with some forms of logical problem, but that a significant proportion of them make predictable errors when attempting to solve such problems. Secondly there is the observation that features of natural language processing appear to provide a possible explanation for subjects' behaviour, as has been the case with the four-card problem. At this point it is appropriate to set the agenda for the work which follows in terms of the proposition that human reasoning is to some extent dependent and affected by strategies adopted

[2]Similar effects have been found for syllogisms, beginning with Wilkins (1928) and recently Oakhill et al (1989).

for language comprehension. To take the matter further, it could be argued that important insights into natural language processing could be gained by observing people's performance on reasoning problems given that the errors they make will be at least in part the consequence of the natural language strategies they use for everyday communication. The evidence already outlined above and some of the discussion below (see especially Newstead's work) support such an argument and pave the way for a detailed examination of these issues.

With these aims in mind, the way forward would seem to be to build on the findings from the four-card problem and investigate human performance in relation to other reasoning tasks. A view of human reasoning is being adopted which assumes that cognitive processes are highly "transparent", that is that they are open to examination directly and it could be argued that this is not the case, however it seems unreasonable to argue that two functions of human cognition, language processing and deductive reasoning, could be conducted by two completely separate and unrelated sets of procedures. It also seems unreasonable to believe that these processes are completely or even largely opaque and that experimental results can give us nothing beyond superficial information about how these processes are carried out. It is essential to a "top-down" approach to take on trust that these operations can be described meaningfully at the macro-level and the success of other studies of this kind in using such methods to explain aspects of human reasoning should bear testament to the persuasiveness of this approach.

To proceed a reasoning problem must be chosen that has certain specific features. These can be listed as being:-

- The problems must require reasoning to be solved correctly
- There must not be too high or too low a ceiling on the difficulty of the problem - subjects must be able to solve some of the problems but have a reasonable risk of making an error
- The problems must to stateable in language used for everyday communication - mental arithmetic would not be suitable for example
- There must be sufficient numbers of the problems to obtain a suitable number of cases per subject

- Specifically with reference to the examination of memory effects and reasoning. the problems must be similar enough that they can be confused in memory during a recall task, but not so similar that they cannot be distinguished. There should preferably be some part of the problem that can be used as a cue for recall and therefore is unique to each problem.

For these reasons the set of reasoning problems known as syllogisms were chosen as a means to study how the way in which humans approach reasoning problems might be bound up in the requirements of language comprehension.

This brief description of the structure of syllogisms demonstrates why this subset of logical problems have been chosen for study in line with the expressed aims of this work. The criteria laid out in the preceding section are all fulfilled: these problems are not of uniform difficulty, therefore subjects have a reasonable chance of answering some but not all of the problems correctly and reasoning is required to solve the problems, so that the performance of subjects not using logic to solve the problems can be distinguished from those that are using logic. Syllogisms as will be shown in the section that follows, can be stated in language that might be used to convey information in other settings and therefore should activate the language strategies of the subjects and they can contain unique information that can be used as a cue for recall, by replacing the abstract "A", "B" and "C" with nouns signifying distinct groups, such as professions, nationalities or pastimes.

1.3 Syllogisms

Syllogisms are a set of logical problems invented by Aristotle. They were used as a form of argument during the middle ages and during this century have been of interest to cognitive psychologists for a variety of reasons. Each syllogism consists of a predicate (P), middle (M) and subject (S) term which are arranged into two premises, each of which contains a quantifier describing the relationship between the nouns. An example is given in Figure 1.1

To correctly solve a syllogism the reader must either decide whether a valid conclusion follows from the premises and give a correct valid conclusion if one does, or determine whether a given conclusion follows from the premises. A valid conclusion is a relationship

"All the M are P."

"All the S are M."

Figure 1.1: An example syllogism

Figures	Quantifiers	Conclusions
AB–BC	All	A–C
AB–CB	Some	C–A
BA–BC	None	No Valid Conclusion
BA–CB	Some..not	

Figure 1.2: Possible components of a syllogism

between the subject and predicate terms which cannot be falsified by any interpretation of the syllogism. Some early studies only allowed one form of conclusion to be considered correct, that of the form "All S are P" for example. Later work has allowed conclusions of the form "All P are S" to be considered correct also[3]. Figure 1.2 gives the four ways the nouns can be arranged in each syllogism, the four quantifiers and the types of conclusion that can be drawn.

There are sixteen possible pairs of quantifiers that could be used in each syllogism and given the four possible figures, there are sixty-four possible syllogisms that can be evaluated. Of these, twenty-seven have valid conclusions of the form A–C or C–A and thirty-seven have no valid conclusion.

1.4 Theories of Syllogistic Reasoning

Syllogisms would be of little interest to psychologist were it not observable on casual inspection that they are not uniform in their difficulty. The very easiest syllogisms are very easy indeed, the very hardest very rarely solved correctly. From a logical point of view there is nothing to justify these differences, it seems to be a uniquely human feature of this kind of reasoning. Much work has therefore been carried out in an attempt to

[3]Following recent work I will adopt a slightly different terminology from that used above and call the predicate term A the middle term B and the subject term C

characterise subjects' attempts at syllogistic reasoning. Below these will be described. firstly those that emphasise the non-logical aspects of subjects' performance, then those descriptions that do rely on some notion of rationality. The strengths and weaknesses of each theory will be evaluated.

1.4.1 Non-Logical Mechanisms and Errors of Interpretation

The first studies on syllogisms adopted the idea that subjects drew conclusions on the basis of non-logical strategies or misinterpreted the premises of the syllogism before drawing a conclusion. Woodworth and Sells (1935) theorised that subjects, when forced to make errors would choose conclusions that had something in common with the syllogism itself. This they called the "atmosphere effect" and they formulated rules to show which type of conclusion the subject would be most likely to draw given any particular syllogism. These, as stated by Begg and Denny (1969), are the following:

- When one quantifier is negative, then the accepted conclusion will tend to be negative.
- When one quantifier is particular then the conclusion will tend to be particular.
- In general the tendency is to choose universal or affirmative conclusions.

Rather than use logic to solve the syllogisms, the argument runs, subjects can choose a conclusion to the syllogism by using the atmosphere of the premises.

There are several problems with this approach as it stands. Firstly it gives no account of how subjects can correctly answer syllogisms, except more or less by chance. Secondly, it fais to explain how subjects can ever give the response "no valid conclusion" as every syllogism has a conclusion that can be derived from atmosphere. Thirdly, it fails to account for effects found in the choices of conclusion that subjects make - the figural effect. Stated briefly the effect concerns the relative frequencies of different types of conclusion, whose frequencies vary according to the order of the nouns in the conclusion. Syllogisms with the figure AB–BC tend to be followed by conclusions of the form A–C and syllogisms of figure BA–CB are followed by C–A conclusions. The atmosphere effect has been criticised for failing to provide any means by which this effect can be explained.

Interest in this approach has continued to the present day and many of the claims of these two theories have been formalised and tested in other ways. Revlis (1975) con-

$+A \; +B$

$+C \; -B$

Figure 1.3: Politzer's formulation: an example

structed models that were elaborated versions of the atmosphere effect and conversion theory. This study found that 88.9% of the data collected could be explained by a feature-matching model based on the atmosphere effect. Politzer (1989) has reformulated the atmosphere effect to explain the responses of his subjects. Recognising that the atmosphere effect as conventionally stated is a good descriptive theory but a poor explanatory or predictive one he has retained the four principles that make it up and added some others. In this revised theory, he includes a principle preventing subjects from creating a conclusion containing the middle term and one that asserts that subjects will always choose a conclusion that is as general or less general than the least general premise of the syllogism. The most crucial new factor however is his introduction of a new means for combining the information carried by the quantifiers and the nouns in each premise. Politzer cites precedents that suggest that the information relating to quantity conveyed by the quantifiers (whether it is universal or particular) is carried by the subject of the premise in which the quantifier appears and that the information relating to quality (whether it is positive or negative) is carried by the predicate. Thus each syllogism can be expressed in terms of a feature matrix, no two syllogisms carrying the same combination of nouns and features. Figure 1.3 demonstrates the consequences of this for the syllogism "All the A are B, none of the C are B". Politzer then uses the principles of the atmosphere effect to select a quantifier for a candidate conclusion, opens this quantifier into a two feature frame as was done for the quantifiers of the syllogism and fills it using the major and minor terms of the syllogism. Rather than chance determining the order of these nouns in the conclusion, Politzer suggests that nouns will be attracted into the available slots by the possession of a feature that matches that of the slot.

By this means Politzer introduces the possibility that different syllogisms will have a preponderance of certain conclusions and that in certain cases the response "No Valid Conclusion" is likely to be given. Figure 1.4 will demonstrate this for the syllogisms

	1	3	4
	+B +A	+B +A	+A +B
	+C -B	+B -C	+B -C
conclusion frame	(+ -)	(+ -)	(+ -)
first filling	(+C -*)	(+* -C)	(+* -C)
conclusion	+C -A	+A -C	+A -C
second filling	(+* -*)	(+* -C)	+A -C
conclusion	+C -A	+A -C	+A -C

Figure 1.4: Drawing a conclusion by Politzer's method

containing the quantifiers "all" and "some" in the first and second premises respectively, as in figure 1.3 with the figures 1, 3 and 4. When attempting to fill a frame for a conclusion using the nouns bound to the feature in question in the syllogism, the subjects will often find that the middle term is the noun most suitable for a slot in the frame. This term cannot play a part in the conclusion and therefore when making a filling where this occurs an asterix (*) is used in its place. The subject must then find another term for the slot. In some cases s/he will decide that there is no conclusion to the syllogism. Politzer states that this decision is made with a probability that "is a parameter of the population", which is taken to mean that individuals will differ markedly in their decision under these circumstances and that predicting their response is not possible. The subjects may also decide to substitute another term for the middle term, this will always be the term not already filling a slot in the frame. In certain cases the frame can be filled directly, without the involvement of the middle term, this is the case with the syllogism AE1[4] . The prediction can therefore be made that the incidence of NVC[5] responses will be low for this syllogism as the necessity to choose a non-ideal noun for the conclusion will never arise.

Once the first fillings are completed, a stage is allowed where any other fillings for the conclusion frame are considered and the conclusions that can arise from this are

[4]Following an earlier convention the quantifiers will be occasionally referred to as A for "all", I for "some", E for "none" and O for "some.not. Pairs of quantifiers can be expressed in this way - AE refers to the pair "All, none"

[5]This acronym is used for the phrase "no valid conclusion"

given. In cases where both the major and minor terms are bound to the appropriate feature in the same premise, Politzer proposes that the term that takes precedence will depend on factors such as the similarity of the conclusion frame to the premise each term appears in and the temporal proximity of each term to the other feature of the conclusion frame.

In this way Politzer shows that his updated version of the atmosphere effect can be made to account for two major phenomenen that previously it was unable to explain: the figural effect and the existence of "no valid conclusion" responses. A full analysis of the 64 syllogisms in this way gives a prediction for the conclusions that closely mirrors the performance of his subjects.

Another early approach was to explain error as the result of misinterpretation of the quantifiers of the syllogism. Chapman and Chapman (1959) identified conversion as a main source of such error, meaning by this that subjects would interpret a statement such as "All the A are B" to mean "A is equivalent to B" and therefore suppose "All the B are A" also to be true. In this way, the subject would misunderstand the logic of the premises and draw an incorrect conclusion. Newstead and Griggs (1983) have followed this precedent and have looked more specifically at subjects' ability to make inferences on the basis of the logical meaning of the four quantifiers used in syllogisms. In their experiments they found that subjects were able to make the contradictory and contrary inferences (if "all" is true then "not some...not" is also true; if "all" is true then "none" is false) but had difficulty with the subcontrary inference ("some" is not equivalent to "some..not") and the subaltern ("all" implies "some" and "none" implies "some..not" and vice versa). Conversions of premises with "all" or "some..not" as the quantifier were made in 34% and 60% of cases respectively[6].

Newstead (1988) evaluated predictions generated by the conversion hypothesis, Gricean implicatures and Erickson's set theoretic approach using Euler's circles and immediate inference tasks. Subjects' choices in these paradigms provide evidence relating to these theories in a number of ways.

[6]For two of the quantifiers used in syllogisms, it is logically possible to infer the converse of a given sentence to be true: "None of the A are B" implies "None of the B are A" and similarly "Some of the A are B" implies "Some of the B are A". However the quantifers "all" and "Some...not" do not imply the converses of the sentences in which they appear.

Newstead found the following:-

1. The conversion hypothesis, stated as a belief of the subject that all quantifiers are symmetrical in their logic, would predict that subjects choose Euler's circles or make immediate inferences to the effect that if "All A are B" then "All B are A" for example. Newstead finds significant numbers of these errors for the immediate inference task but only for "all" with the Euler's circles paradigm.

2. Given that subjects are following Gricean implicatures when interpreting the syllogisms, they should tend not to make the inference that "Some A are B" implies "All A are B" and "Some A are not B" implies "No A are B". In other words they will interpret "some" as meaning "some but not all" and "Some...not" as meaning "some but not none". There is evidence that subjects perform in this way for the Euler's circles task but not in the immediate inference paradigm.

3. The difficulty of the interpretation of the quantifiers is related to their complexity, as predicted by Erickson's (1974) set theoretic approach, but although this theory specifically rests on subjects choosing only one representation for each quantfier, the evidence suggests that this is rare and the difference in difficulty for "some" and "Some...not" is not found to be great enough to unequivocally support Erickson's hypothesis.

Newstead concludes that both conversions and Gricean implicatures play an important role in the interpretation of quantifiers, but that their relative importance depends upon the sort of task the subject is required to perform[7].

1.4.2 The deductive approach

The theories described above tend to ignore the deductive component of this process and ascribe errors in drawing conclusions to processes outside reasoning. Some researchers have become uneasy with using these approaches as whole explanations, as they appear to overlook the fact that subjects can and do behave rationally when attempting to solve syllogisms. Various attempts have therefore been made to create a description of the deductive processes subjects use when solving syllogisms.

[7] See also Politzer(1990)

Guyote and Sternberg (1978). for example use a combinatorial symbolic language to capture subjects reasoning performance. They assume that the combinatorial rules are applied without error, and that subjects produce incorrect responses by failing to consider all the possible representations that can be considered. Newell (1981) takes Venn diagrams as a starting point, with a heuristic process combining the premises to provide a possible conclusion to be evaluated.

Erickson's (1974) set-theoretic approach uses Euler's circles as the basis of the representational device, with limitations on the numbers of interpretations that can be considered. The reason that this limit must exist is that the traditional version of Euler's circles requires several representations to be constructed to fully capture the logic of the quantifiers. Figure 1.5 demonstrates that only "none" has only one interpretation under this representation, "all" has two, "some..not" three and 'some" four.

Erickson (1979) tests two models based on this notation, one of which allows only one interpretation of the syllogism to be entertained and one which allows all interpretations to be considered in turn - a random combination model. The former model is the most sucessful explanation of the data, but does not allow the response "no valid conclusion". Erickson suggests that a compromise model that can consider more than one but not all combinations of the syllogism would overcome this difficulty.

Major criticisms of these theories, made by Johnson-Laird (1983) are that they assume that subjects possess a mental logic which is in some way flawed so as to produce the errors observed. He argues that theories of this kind cannot explain why some syllogisms appear to be harder than others. Theories such as Erickson's he criticises for allowing a combinatorial explosion of models and having to be artificially constrained as a result. Further he states, the fact that "some" and "some..not" have an equivalent representation in an Euler's circles notation, forces Erickson to invoke the atmosphere effect to explain why subject will choose one quantifier in some situations and the other under different circumstances.

Johnson-Laird's solution to these problems is to propose a different representational device based on individuals rather than sets and to appeal to the limitations and properties of working memory to explain subjects' errors and preferences. Firstly, as did Erickson he identifies the fact that certain syllogisms require a number of mental operations to be performed before the presence or absence of a valid conclusion for them

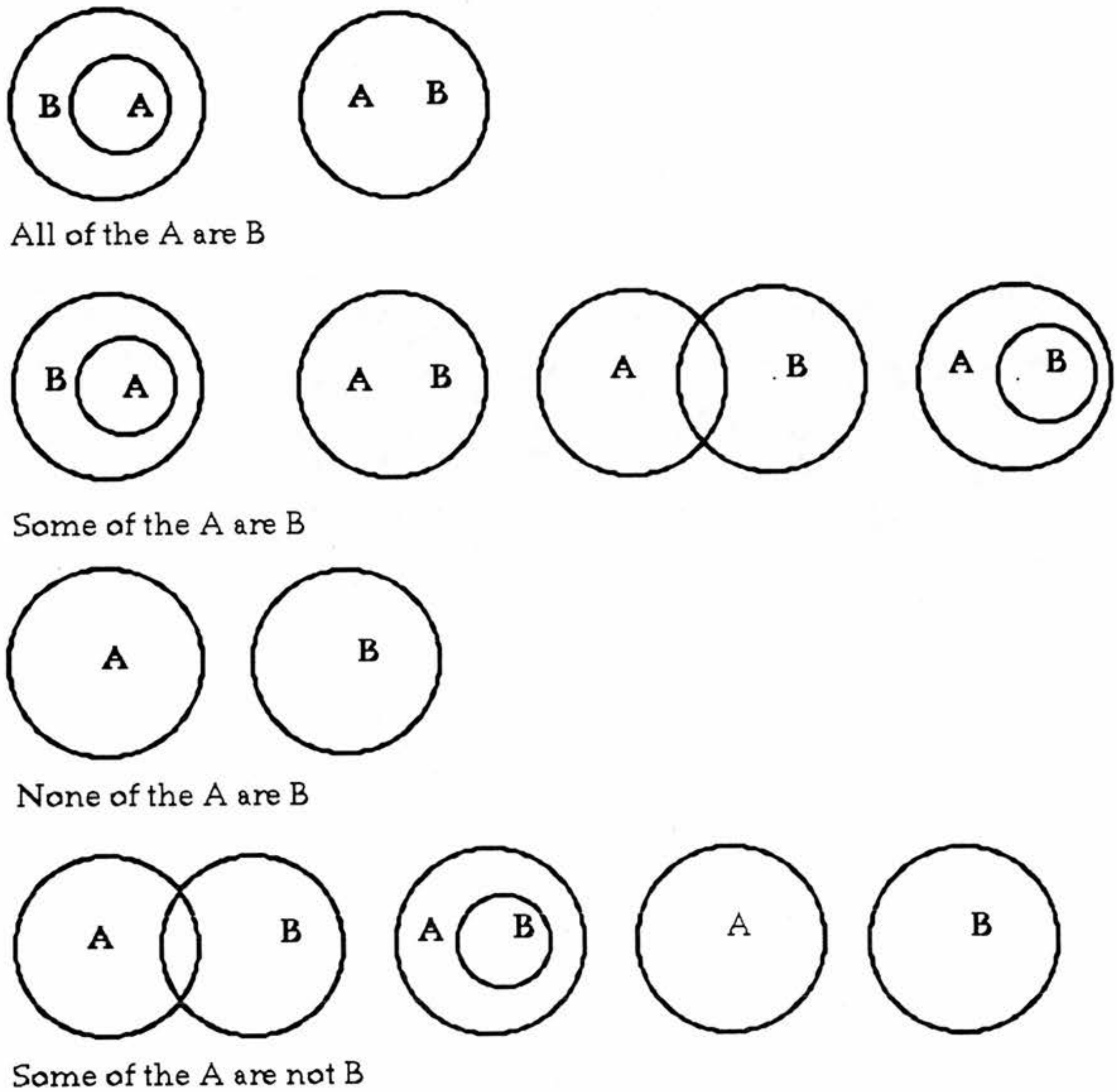


Figure 1.5: The four quantifiers in Euler's circles notation

can be ascertained. Other syllogisms require only one representation before it becomes clear that there is a valid conclusion. Johnson-Laird argues that the more operations a syllogism takes to solve, the less likely it is that the subject will be able to produce the answer as, following Sternberg (1981), the limits on working memory make it difficult for her/him to entertain all the possibilities simultaneously. In addition, Johnson-Laird proposes that subjects tend not to attempt to negate the conclusions they have drawn and are thus led into error. Examining the error rates of subjects allowed to freely draw conclusions, Johnson-Laird (1978) finds high numbers of errors where two or three operations are required, low error rates where only one model is needed.

His second criticism is that these theories of syllogistic reasoning fail to account for subjects' marked preference for certain forms of conclusion - the figural effect. To elaborate on what was mentioned earlier: it was established by Frase (1968), Pezzoli and Frase (1968) that the four figures of the syllogism resemble the four stage mediational paradigm and on the basis of this, differences in the difficulty of the syllogisms were predicted according to figure[8]. It was found by Dickstein (1978) that these differences do exist; that subjects make most errors when the figure is AB-BC intermediate errors on AB-CB and BA-BC and fewest on BA-CB. In addition, the number of conclusions of the form A-C exactly reverses the trend. If conclusions of either form are allowed, this finding shows that for the figure AB-BC' conclusions of the form A-C are preferred and C-A for figure BA-CB, while the other two figures have no bias.

To explain this effect Johnson-Laird states that when interpreting the syllogism subjects must arrange the nouns in it so that they can be related into a single model. For the figure AB-BC this is straightforward: A is first related to B and then B to C and the relationship between A and C tested. With BA-CB this cannot the case. To represent syllogisms of this figure, subjects must find a way of rearranging the nouns in the syllogism so that the middle term stands between the two other terms. Johnson-Laird hypothesises that the premises are interpreted in reverse order of presentation, giving a figure of the form CB-BA. With the figure AB-BC, the term A enters working memory before the term C. Thus, when a conclusion has to be drawn, this term is the first to be retrieved, given a "first in-first out" principle is operating in working memory.

[8]Here reference to difficulty is only meaningful given that correct conclusions must be of the form C-A.

Therefore conclusions of the form A–C will be preferred when this figure is encountered. With the figure BA–CB, the term C will be the first to be retrieved from working memory and therefore conclusions of the form C–A will be chosen. For the other two figures. reversal of the nouns within one premise to give one of the two arrangements required is thought to be the mechanism used. If the nouns are reversed at random then a 50-50 distribution on types of conclusions should be found. and this is in fact the case with the data in question.

1.4.3 Evaluating the hypotheses

A brief description of a number of major theories has been given along with what have been identified as their main problems. Johnson-Laird uses these arguments as arguments in favour of his mental models approach and offers evidence of its applicability to other areas of human reasoning to advance the argument that a mechanism of this kind forms the basis of a large part of human verbal reasoning and other language processes.

However, it may be that he is too ready to dismiss other theories. sometimes perhaps because they present problems that his approach cannot resolve. In the following sections, it will be discussed whether certain theories really are incompatible with Johnson-Laird's claims and ways of reconciling what may appear to be at first sight conflicting approaches considered. It will also be of interest to discuss whether some of these theories have offered insights that Johnson-Laird connot easily explain.

Quantifier misinterpretation

Misinterpretation of the quantifiers as a source of error in reasoning. is an established finding, as already shown. For example in Fisher's (1981) model. misinterpretation. particularly conversion, is found to be necessary to explain typical performance. A model depending on correctly interpreted premises provides a poorer fit to the data then one including misinterpretations. Fisher also finds that introducing a factor resembling a limit on working memory capacity by constraining the number of interpretations of the quantifiers that the system can entertain must be included, as must a deductive reasoning component which produces errors through the use of alogical and incomplete strategies.

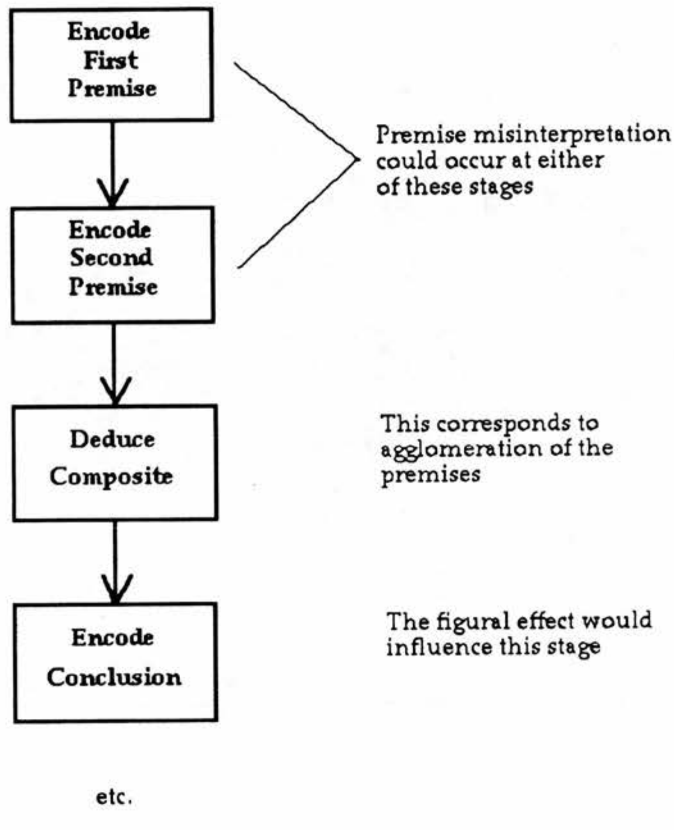


Figure 1.6: Part of Revlis' flow chart for syllogistic reasoning

Most formalisations of syllogistic reasoning use a flow chart to demonstrate the process. To take as an example Revlis' (1975) model (see Figure 1.6), this takes as separate parts of the process the interpretation of the premises and the manipulation of the representation derived from the premises to produce an answer to the problem. In principle then, there is no reason why a hypothesis such as the conversion hypothesis or Gricean implicatures should conflict with a theory such as that of mental models. In fact Johnson-Laird himself separates the reasoning process into two such parts - premise combination and then cycles of testing to establish the correct response. Thus hypotheses that explain error through misinterpretation of the quantifiers are found not to be exclusive of theories incorporating deductive reasoning. It is quite plausible that a subject might incorrectly interpret the syllogism and go on to reason with it logically. This seems even more acceptable once it is remembered that the conception of conversions, for example, has changed with time from being an automatic and inevitable process (Revlis,1975), to it being a natural propensity of the system (Revlin and Leirer, 1980). In this way, subjects do not have to make conversions, but may make them, for example, by failing to disallow a representation which includes an illicit conversion.

The atmosphere effect

The atmosphere effect is not a theory of premise misinterpretation, but rather a mechanism that allows subjects to choose conclusions without using deductive reasoning processes. Given that subjects are observed to behave rationally, it is hard to defend a theory that does not make any allowance for deductive reasoning. Revlis hypothesises that in circumstances where rationality cannot be used, such as in Woodworth and Sells' experiment where subjects were asked to draw conclusions for syllogisms where no valid conclusion exists, feature-matching will be used as a strategy to generate conclusions. Politzer, whose neo-atmospheric approach resolves many of the arguments against the atmosphere effect as originally conceived by developing a mechanism whereby both the figural effect and "no valid conclusion" can be generated without the need for deduction, argues that some syllogisms indeed exceed the capacity of the subject. It might also be the case that some subjects fail to use deductive reasoning at all and use nothing but non-logical strategies to generate conclusions. Politzer points out that there is a group of syllogisms who appear quite difficult to solve in that correct answers are given to them rarely and a larger group that get correct responses only part of the time. He infers from this that many syllogisms are beyond the capabilities of a large group of subjects who therefore use other means to draw a conclusion to them. This view conflicts with the deductive approach that holds that underlying all syllogism solving is a reasoning mechanism common to all who attempt syllogisms and probably to the whole class of reasoning problems. Rather than differing in their ability to hold a number of possibilities in mind, rather as we differ in our ability to perform mental arithmetic, according to Politzer subjects differ in their approach to the problems themselves. Few if any, he supposes, reach the stage of using deductive reasoning to solve these problems and instead resort to the method he describes. It is also possible that subjects use an atmosphere approach to first generate conclusions to the syllogisms and then gradually learn to test the conclusions thus arrived at by means of a interpreted model.

Theories of deduction

It has been shown that theories based on non-logical mechanisms do not necessarily conflict with the idea that subjects can treat syllogisms rationally, but may simply be capturing the performance of subjects who do not fully understand the logic of the quantifiers, even if the rest of their performance is immpeccably deductive or that in circumstances where the subject finds a syllogism exceeds their capacity, non-logical approaches are used to generate an answer.

The explanations for the deductive part of the process have also been criticised however. The three main problems found and introduced above are: that theories based on rules of inference express in syntactic form that which is semantic, that theories based on Euler's circles notation cause combinatorial explosion and that none of these theories adequately explains the figural effect.

First let us consider the criticism that human deductive reasoning is semantic rather than syntactic and that therefore systems of reasoning that are based on proofs,being syntactic in nature connot incorporate that which is distinctive about human processing of this kind - its reliance on semantic networks in the shape of context and real world knowledge. Intuitively it seems that there is a very real difference between these two things and it is this argument that Johnson-Laird uses to motivate his theory. Stenning(1990) makes clear however that there is a conceptual muddle about the use of the terms syntactic and semantic in this sense. He argues that there is no syntactic arrangement of a logical process that cannot be restated as a piece of semantics (say as a theorem being restated as a truth table and vice versa) Rather than these two levels being two sides of the same coin, he conceptualises them as two points along a continuum along which human reasoners travel, the poles of which are agglomerative strategies versus analytical ones. At some level of coding it may well be true that human reasoning is analytical, but if this is so, it is a level not available either to introspection or experimentation. What Johnson-Laird has asserted about human reasoning is not that it is semantic, but that it is agglomerative - that human beings tend to combine all the information available into a single representation which they then use to assess the truth or falsity of statements. The problem with syllogisms is that assessing their truth or falsity involves the kind of abstract manipulations that we find places too great a burden on memory.

If it is agglomerativeness that is important rather than "semanticness", what advantages do mental models have to offer above other types of representation that are also agglomerative? Erickson's set-theorectic approach is certainly agglomerative - Euler's circles almost beg to be combined, but this approach has been criticised because of the combinatorial explosion it is supposed to provoke, as discussed above and the point has already been made that tolerating more than one representation is exactly what people are bad at. Mental models get round this problem by using individuals rather than sets and including a notational device that can allow the possibility of a relation existing, without asserting it to necessarily exist. Stenning and Oberlander (1990) note however that the inclusion of a shading notation allows Euler's circles to capture the information in each quantifier with one diagram as Johnson-Laird's tableau do. Figure 1.7 shows that shading the area in an Euler's circles representation where individuals can be asserted to definitely exist gives an equivalent representation to tableau, while also falling in with subject's intuitions about the type of structures they are using.

These diagrams can be combined just in the same way that Johnson-Laird's are, but with the interesting effect that in some cases there are areas that become shaded twice as a result (see figure 1.8). This feature does not correspond exactly with the existence or otherwise of a valid conclusion, although in many cases where double shading exists, a valid conclusion does exist and the absence of shading is often found where no valid

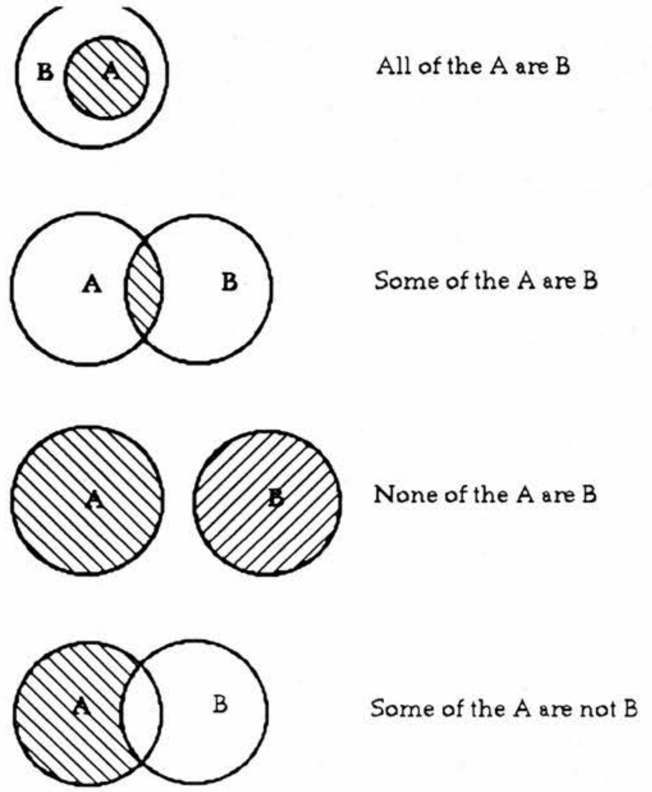


Figure 1.7: Stenning's version of Euler's circles

conclusion can be found. The fact that there is no exact overlap between double shading and the existence of valid conclusions leads to interesting predictions about subjects' behaviour should they be using this method to select conclusions.

It can be seen that in both cases several possible representations must be considered before the correct response can be generated, so Johnson-Laird's intuitions about the way memory interacts with the reasoning process can also be preserved.

1.4.4 The explanation of the figural effect.

The important claim about mental models than is that they are agglomerative and parsimonious, this has been shown to be the case for an Euler's circles notation too. The claim might then be made that mental models stil have some greater claim to explanatory power because they are able to provide an explanation for the figural effect. As already shown, Johnson-Laird claims that this effect occurs because the nouns in the syllogism must be presented in a certain order for a representation to be constructed. This is not a necessary feature of all representations of syllogisms, but follows from mental models because they incorporate an element of order. An Euler's circles representation, for example, does not require nouns to be presented in any particular order.

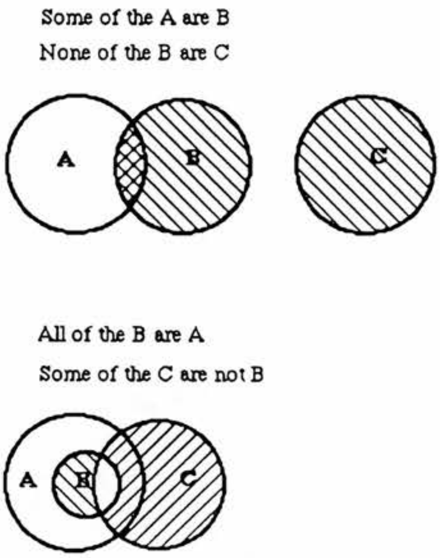


Figure 1.8: Combining premises with Euler's circles

any information can be included in the representation at any point, but for a mental model to be constructed, information entering the system must be related to objects that have already been included in the model, leading to the necessity for a particular order of processing to be adopted. It is this, the argument runs that creates the figural effect, because of the way in which memory must deal with this ordering.

Thus the figural effect falls out of mental models theory as part of the process of reasoning. But as a consequence, Johnson-Laird is forced to make a very strong claim about the form of the representation subjects use. A subject that produces a figural effect must use a mental representation that involves the use of individuals rather than sets. In other words, according to Johnson-Laird, subjects cannot use Euler's circles as a means for solving syllogisms and still display a figural effect, despite the fact that subjects report the spontaneous discovery of Euler's circles as a reasoning device as they become proficient at syllogistic reasoning (Inder 1986) and that Euler's circles are widely used as a means of teaching syllogistic reasoning (this, incidentally is one of Johnson-Laird's prerequisites for a successful theory of syllogistic reasoning). Johnson-Laird might argue that underlying an Euler's circles representation is one based on individuals, or that the preference created by mental models remains even if the representation changes. This leads to a characterisation of the process that is unecessarily complex.

Johnson-Laird makes the assumption that the figural effect must be a consequence of the representation adopted by the subject, hence his success at criticising other theories for failing to explain it. However, what he is doing in fact is providing a description of mental models rather than an explanation. Euler's circles could be given a notation to specify order of processing, without this necessarily explaining why the order has to be so. A conceptual leap is made in assuming that because the nouns in mental models have to be arranged in a particular way this implies that they have to be processed in that order. If what is required of a theory is that it provides a description of the figural effect then this paves the way for any explanation of the figural effect to be acceptable so long as it can be incorporated into existing theories.

From another point of view then it is possible that syllogisms are seen by reasoners as an attempt at communication (there is evidence cited above from study of other reasoning problems to suggest that this is a likelihood), then the most likely interpretation of syllogisms will be that they are a form of argument. This is the position adopted by Gilhooley and Wetherick (1989) who go on to point out that if syllogisms are interpreted as arguments, then it is natural to suppose that the subject will identify some topic for the argument to be about. This topic will be the subject of the conclusion, as the point of an argument must be to make some statement about its topic. The person solving the syllogism must use some criterion to decide what item in the syllogism the argument is about. There is ample evidence from linguistics[9] to suggest that one such might be the grammatical roles of the nouns in the syllogism, with the subject term being taken as the subject of the argument. When there are two subject terms or no subject terms in the syllogism then other factors might come into play, some of which will be discussed in chapter 4.

If the requirement of a theory is that it provides a framework in which a description of the figural effect can be couched then the force of the argument in favour of mental models is weakened. Any factor could be responsible for creating the effect, a respectable substitute has been proposed above, so it is not necessarily the case that the representation itself must be the cause of the effect.

[9] see Chafe (1976) for a summary

1.4.5 Explaining quantifier misinterpretation

Factors such as premise misinterpretation and the atmosphere effect are well supported experimentally as sources of error in syllogistic reasoning and have been shown not to exclude in principle the possibility of subjects using deductive reasoning as another part of the process. In fact the experimental finding that subjects show preferences for certain interpretations of the quantifiers which accord with Gricean implicatures creates a difficulty for Johnson-Laird's choice of representation. As noted above, using a mental models notation has the advantage that one representation can be used to denote each quantifier, whereas with the original formulation of Euler's circles between one and four diagrams are needed. Stenning and Oberlander's (1991) additional notation solves this problem, but allows a measure of ambiguity that may capture why logically inexperienced subjects fail to fully understand the meaning of the quantifiers. Other researchers have found that "some..not" and "some" are often confused in meaning. With the Euler's circles notation it is easy to see how this could happen - misplace the shading and the two become equivalent. With mental models, quite radical restructuring is required before the two resemble each other and it is hard to see how subjects can ever make such errors if they are using such a representation.

The same is true of the conversion hypothesis. If subjects make conversions as a result of only choosing a particular interpretation of the quantifier (in the case of "all" for example they would choose the interpretation "A is equivalent to B") then they must have a representational device that does not necessarily allow for the full logical meaning of the quantifier to be included. Although Johnson-Laird can achieve this by excluding some of the information from the mental model for each syllogism, he would render his tableau equivalent to Euler's circles by doing so. Gricean implicatures provide a basis for arguing that subjects choose an impoverished version of the meaning of the quantifiers "some" and "some..not" as these implicatures are claimed to underlie language processing and will affect our comprehension and therefore representation of information *a priori*. However, conversions do not appear to result from any preexisting preference, rather they appear to be a consequence of the representation itself. The same is true of the common confusion between "some" and "some..not", although this may result from a Gricean interpretation of their meaning leading to equivalent representations for these two quantifiers. Thus, as quantifier misinterpretation is supposed to result from the representation itself and as there is no principled reason for any information to be excluded from a mental model , to do so in order to allow quantifier misinterpretation to occur would be a device designed to suit the data. Johnson-Laird claims that Erickson's theory is inadequate because it cannot capture the figural effect due to the symmetricality of Euler's circles and that the inclusion of some device to generate a figural effect would be "...an ad hoc manouevre designed to save (the theory) from falsification". Now it is found that to capture the effects of quantifier misinterpretation, mental models would have to make a similar manouevre. On the other hand, quantifier misinterpretation "drops out" of an Euler's circle representation as a result of one of the criticisms Johnson-Laird makes of them:- the memory load on subjects produced by the multiplication of figures needed to capture all the information about the quantifiers means that not all the figures are produced and those chosen are those that conform to factors such as Gricean implicatures.

1.5 Logical and non-logical strategies

Theories of reasoning generally assume that all subjects exercise some degree of deduction when solving syllogisms. One of the criticisms of the atmosphere effect for example is that it makes no provision for deductive reasoning strategies. It is certainly the case that some subjects display a grasp of the logic underlying syllogisms and use this with a greater or lesser degree of success to obtain an answer. However, there is evidence that some subjects fail to grasp the logical significance of syllogisms altogether and use other strategies to generate conclusions as argued by Politzer. Gilhooley and Wetherick (1989) find that subjects choose a conclusion that has the same quantifier as one of the quantifiers in the syllogism. By isolating a group of six syllogisms that have both quantifiers the same and for which there is no valid conclusion, they identified a group of subjects that were reporting conclusions for these syllogisms containing the same quantifier as that found in both premises. The performance of their subjects overall correlated highly with the number of such conclusions the subject gave - the higher the number of these conclusions, the higher the number of errors overall. Gilhooley and Wetherick concluded that certain subjects were using what they called a "matching" strategy to solve the syllogisms, rather than deductive reasoning. In a situation

A–B

A–B

(B)–C

(B)–C

(C)

All of the A are B

Some of the B are C

Figure 1.9: Possible interpretation of "All of the A are B, some of the B are C"

where the subjects had two different quantifiers in the syllogism to choose from for their conclusion, a principle such as the atmosphere effect might be used.

These responses also occur in Johnson-Laird's own data, but his theory fails to take account of the possibility of non-logical performance altogether. In fact, the claim is that mental models are a version of the representation that is the basis of all human inferencing - one of the arguments advanced for this is that mental models can be used to represent quasinumerical quantifiers and multiply-quantified assertions. If this is the case, then it must also be the case that simply to read and understand the syllogism involves the construction of a mental model of the type described, as this is the way the semantics of the quantifiers are understood. In this way all subjects can produce at least one model of a syllogism and generate a conclusion from it.

However this on its own does not explain why subjects when encountering the syllogism in figure 1.9 do not choose the response "None of the A are C" as a result of creating one model alone. To explain the instances in his own data where subjects produce matching responses where the response "no valid conclusion" would be correct, Johnson-Laird devises two heuristics that guide the initial choice of representation.

- Heuristic 1 — with affirmative premises ensure that the intersection between sets is never empty.
- Heuristic 2 — with negative premises ensure that the intersection between sets is always empty.

In this case, subjects who can only produce one mental model can generate a conclusion

A–B–C

A–B–C

(A) (C)

All of the B are A

All of the B are C

Figure 1.10: Mental Models representation of "All of the B are A, all of the B are C"

on the basis of the model that they have.

As demonstrated, these heuristics can generate the "matching" responses in a number of cases. However, there are other syllogisms where matching responses cannot be generated by these heuristics (see Figure 1.10). The response "All the A are C" or "All the C are A" to the example in figure 1.10 would not follow from the mental model drawn in accordance with Heuristic 1. As shown, it should be immediately obvious that the correct response is "Some of the A are C" or "Some of the C are A". In Johnson-Laird and Steedman (1978) 50% of subjects gave a response with "all" as the quantifier for the conclusion.

Given that this example cannot be accounted for, a theory of syllogistic reasoning must include the possibility of subjects using non-logical strategies. As subjects appear to sometimes choose conclusions that cannot be generated from a single mental model, even with a heuristically determined starting point, it must be supposed that some subjects are reading and understanding the syllogism without constructing a representation of the sort that Johnson-Laird proposes. A reason this might be possible is that syllogisms are first-and-foremost texts and, as Begg and Harris (1982) point out, they can be treated as attempts to communicate. If subjects fail to recognise the logical possibilities of the syllogisms then they will use pragmatic inferences to generate conclusions for them, of which matching is an example and the atmosphere effect another. The representation of syllogisms at the textual level is therefore unlike that used by subjects who are attempting deductive reasoning. The claim made by Johnson-Laird that mental models are the automatic means of representing syllogisms is breached.

1.6 Memory and Syllogistic Reasoning

As discussed in the first section, an understanding of memory is crucial for an understanding of human reasoning as its structure underpins all our cognitive processing. In particular a study of what information survives in memory after a particular task has been performed can give quite subtle insights into the nature of the processing that preceded the memory test. Successful use has been made of such a method to investigate the nature of textual processing by Stenning, Shepherd and Levy (1987), with the result that information has been gained about how people can associate properties as belonging to one individual and not another and thus build up stable descriptions of individuals. This question is also of importance to theories about syllogistic representations as these also require the processing of collections of individuals. Stenning has pointed out that there are seven distinct types of individuals that can be found in any syllogism: the three nouns of the syllogism can be combined in any one of seven ways to assert that an individual either has or does not have that particular property. Syllogisms can describe situations in which up to all seven of these individuals exist, but only in a limited number of situations will it be the case that a particular individual must exist. This is equivalent to a syllogism possessing a valid conclusion although the overlap between conclusions and individuals is not exact. This can be found to relate to the reformulated Euler's circles notation in an interesting way: as shown before in figure 1.11 certain syllogisms when represented in this way have areas that are shaded twice, or "double-hatched" and these are always syllogisms that have individuals that must exist. There are in addition two cases where syllogisms that are not double-hatched have valid individuals and some cases of syllogisms that have individuals but no valid conclusion (figures 1.12 and 1.13). In addition to this Stenning has noted that representing the quantifiers as expressing a simple affirmative or negated relation between two nouns[10] and combining the two premises gives a simple structure for the syllogism in which the B or middle term is affirmed or negated in both premises or negated in one and affirmed in the other (figure 1.14 and figure 1.15) depending on the figure of the syllogism. The cases that give B a different role in each premise are syllogisms that never, bar two cases, have a valid individual or conclusion. Should subjects be aware of such a mechanism,

[10]The quantifiers "all" and "some" have both nouns positve, "none" negates the non-repeated term and "some..not" negates the predicate

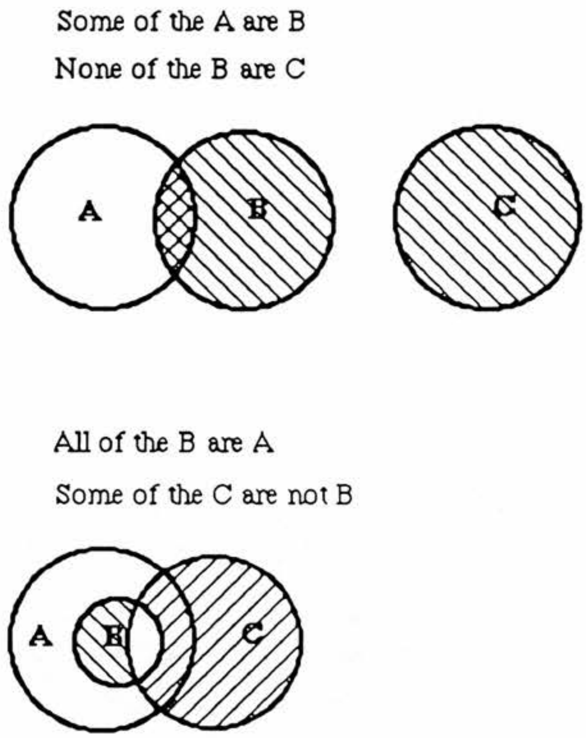


Figure 1.11: Combining Stenning's variation on Euler's circles

they would have a quick and easy way to decide if certain syllogisms have a conclusion or individual. The two exceptions are the two that have no double-hatching yet possess a valid conclusion and individual.

Asking subjects to find individuals rather than draw conclusions has never been done before. It is of interest therefore to try this variation on the syllogistic theme to see if subjects are able to find the individuals that exist where no conclusion does and to look for evidence that they use the presence of double-hatching and cancellation to make decisions about the presence or absence of individuals. These results can be compared to those obtained in the usual task to see if these factors play a part in the strategies adopted by subjects to find valid conclusions.

Another way in to the reasoning task though, is through memory. As mentioned above studies of recall data have already provided information about the ways in which properties and individuals are kept together and modelled in distributed systems. Memory could play a similar part in gathering information relating to the way properties are combined in representations. But memory also plays a crucial part in the explanation

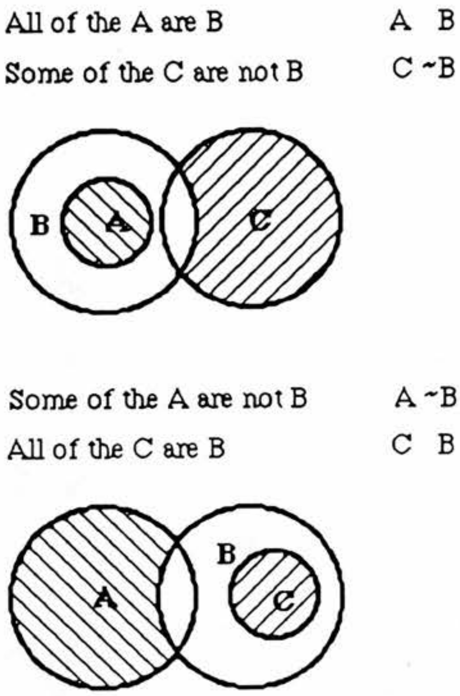


Figure 1.12: Syllogisms that are not double-hatched but that have valid conclusions

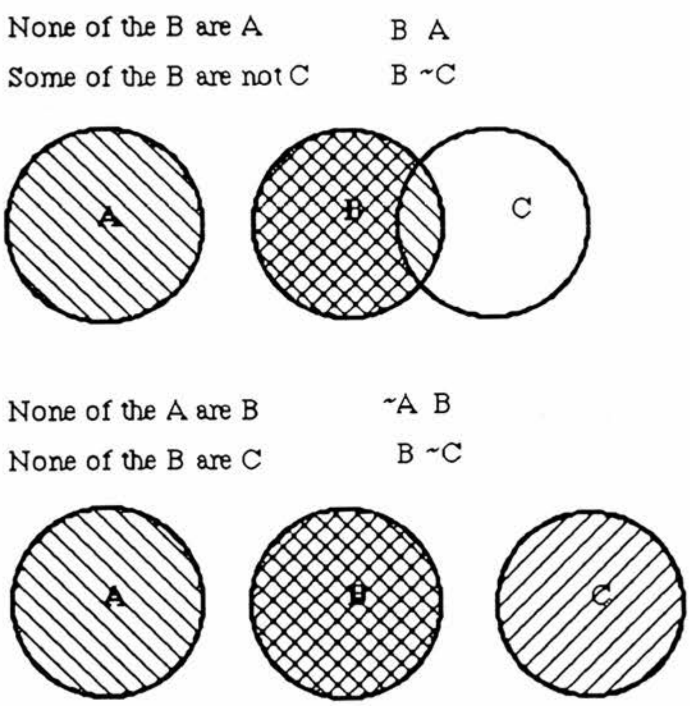
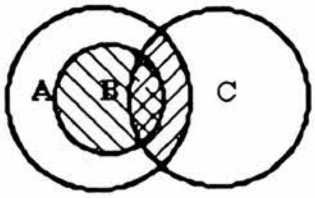


Figure 1.13: Syllogisms that have a valid individual but no valid conclusion

All of the B are A B A

Some of the C are B C B



All of the A are B A B

None of the B are C B ~C

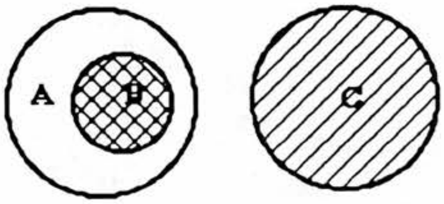
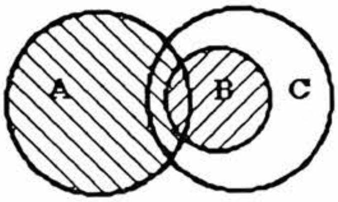


Figure 1.14: Syllogisms that cancel

Some of the A are not B A ~B

All of the B are C B C



Some of the B are A B A

Some of the C are not B C ~B

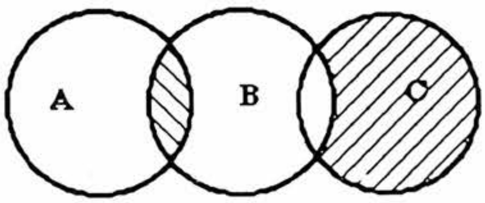


Figure 1.15: Syllogisms that do not cancel

of syllogistic phenomena: the mental models account of the figural effect and, most importantly, the errors created by cycles of testing exceeding memory capacity.

These relationships between memory and syllogistic reasoning could affect recall data in a number of general ways: phenomena such as the depth of processing effect reported by Craik and Lockhart (1972) may have relevance if memory for particular syllogisms can be shown to vary according to the amount of processing required to correctly solve them. It has also been demonstrated that the greater the difficulty of a task that follows a piece of text or the greater the processing capacity required to perform the task, the better remembered the text that was originally read[11]. It has been found that increases in difficulty beyond a moderate level do not improve memory and can in fact degrade it and that tasks that integrate information within a text aid recall after a delay in particular. Again these effects might be expected to transfer to the syllogisms paradigm in a similar way to depth of processing, but perhaps predicting a drop off in recall performance for the most complicated solutions and accentuated effects of amount of processing at delayed recall. Kintsch and van Dijk's (1983, 1990) work on levels of representation within texts gives a similar view: that the better intergrated the text, the better the recall of that text will be. Ratcliffe and McCoon (1989) suggest that different kinds of information may be available to recall at different points in retrieval. They have found that early in the time course of recognition tasks, surface similarity of the texts plays an important factor, and other considerations, related to deeper processing for delayed recognition. Again, cued recall of syllogisms may be affected by similar factors, with errors due to the formation of mental representations only appearing in any numbers at delay although the work of Lea et al (1990) finds that subjects are able to discriminate in memory between information that was actually present in a given text and information that must have been inferred from the text to correctly answer a question about it. The effect of reasoning on syllogisms may not therefore be to directly introduce inferred material into the recall.

1.7 Agenda

The general motivation for this work has already been outlined as being the desire to examine in detail the way in which natural languge processing affects human deductive

[11]reveiwed in Einstein and McDaniel 1990.

reasoning and further what can be said about language processing itself on the basis of this. A number of more specific questions have also been raised with respect to the interaction between memory and syllogistic reasoning and also issues about combining information, the operation of the figural effect and the whole question of what kind of representation is available to subjects of different levels of expertise.

In an attempt to decide these questions and to collect information about the general issues that form the theme of this work an experiment where recall data for syllogisms has been collected and analysed has been carried out. The second part of this work has been to collect data regarding subjects' ability to find individuals. This experiment it is hoped will demonstrate the existence of strategies for solving syllogisms similar to those found for the four-card problem and in general to aspects of language processing. Because the data collected from the second experiment relates to the reasoning and recall data collected from the first experiment, it has not been possible to adopt a conventional structure for this work, in which a first experiment is described and analysed and then a second. Instead the two experiments have very much been taken together and there are many points at which data from both experiments can be taken together to furnish a convincing argument which would otherwise be lost.

Chapter 2

Method

2.1 Experiment One

2.1.1 Rationale

This experiment aimed to investigate the relationship between language processing and reasoning by asking subjects to perform a reasoning task and then examining the recall errors they made when asked to remember the content of the syllogism. From these errors it should be possible to build up a picture of the strategies, priorities and assumptions subjects use when attempting to solve these reasoning problems and relate these to errors of reasoning. Ultimately a model of how humans reason might be constructed from this data which would probably have much in common with earlier formulations, but which will be enriched by the information collected here. Cued recall was used as the means to elicit the memory for the syllogism, as free recall under circumstances such as these is likely to be too poor to provide much of interest. Two recall conditions were included, one where recall was immediately after the subject read the syllogism, one where the recall was at 32 texts delay. The aim here was to ensure that a recall condition of sufficient difficulty was included - there are few precedents for this type of recall task, so little was known about the difficulty of the task. Should immediate recall prove too straightforward, delayed recall would certainly be more productive of errors. In addition, it is also possible to study the effects of decay with time on memory and further gain a picture of the most and least persistent aspects of the information the subject has stored.

2.1.2 Subjects

Twenty-five subjects were used, aged between 18 and 40, all graduates or undergraduates. Most were paid for their participation, some were volunteers. It was felt necessary to choose subjects who were likely to be familiar with some aspects of logical reasoning if not with syllogisms themselves. People who are exposed to syllogisms without any prior experience of this type of task are apt to find it difficult to grasp the requirements of the experiment without substantial coaching, for example the distinction between a valid conclusion and one that is only true in certain situations can be difficult for very naive subjects to understand. This can lead to large numbers of non-standard responses being given and although this experiment was constructed so as to prevent subjects from giving anything but standard answers, it was felt that inexperienced subjects would be confused to the point at which little of value would be obtained from their responses.

It was also important to use subjects who were not very familiar with syllogisms and also to exclude people who may be familiar with psychological theories of syllogistic reasoning. For that reason, students who were undertaking studies at the Centre for Cognitive Science were not asked to participate, as most if not all of these students had taken courses specifically concerned with syllogistic reasoning and its psychological implications.

2.1.3 Materials

This experiment requires subjects to read and evaluate syllogisms. Syllogisms consist of two sentences, each begins with one of four possible quantifiers. These quantifiers are "all"; "some"; "none" and "some not" known as the positive universal; positive particular, negative universal and negative particular quantifiers respectively. As there are two sentences and four possible quantifiers for each sentence, the number of possible quantifier pairs is sixteen. The rest of each sentence consists of two nouns. An example syllogism is given in Figure 2.1.

One noun is repeated in both sentences - this is known as the middle term. The unrepeated item in the first sentence is the called the major term and that in the second sentence the minor term. The nouns in the two sentences can be arranged in four different ways relative to each other. The syllogism in Figure 2.1 illustrates one such order, another is given in Figure 2.2

All of the busdrivers are cricketers.
None of the cricketers are German.

Figure 2.1: First example syllogism

Some of the French are not butchers.
All of the French are footballers.

Figure 2.2: Second example syllogism

Using the convention that the major term is A, the middle term B and the minor term C the four figures can be represented as AB–BC; BA–CB; AB–CB and BA–BC, giving the order of the nouns in the two premises of the syllogism. Combining the order of the nouns or figure of the syllogism with the number of pairs of quantifiers possible, the total number of syllogisms that can exist is sixteen times four or sixty-four.

Twenty-seven of these syllogisms have a valid conclusion - that is the situation described in the syllogism implies a relationship between the major and minor terms that cannot be falsified by another interpretation of the syllogism. Thirty-seven of the syllogisms have no valid conclusion, that is to say that there is no relationship between the major and minor terms that cannot be falsified by another interpretation of the syllogism.

2.1.4 Design

All sixty-four syllogisms were used. Each was presented to the subject once, in two sessions of thirty-two syllogisms each. The syllogisms were randomly assigned to these two sessions and were presented in a random order so that no two subjects read the syllogisms in the same order, or had the same group of syllogisms in either the first or second session. Each syllogism had a unique set of nouns: a nationality, a profession and a hobby were used in each syllogism and never repeated. The order of noun type was randomised so that, for example the major term was not always a profession, but might be a hobby or a nationality.

2.1.5 Presentation

The syllogisms were presented on a BBC microcomputer. Each subject first completed a practice session of 4 syllogisms, accompanied by a set of printed instructions, which explained how to use the microcomputer and gave four syllogisms with the correct answers worked out. Two of these had a valid conclusion, two did not. No particular strategy for representing or solving the syllogisms was discussed, but the correct conclusion for the syllogism was given with some indication as to why this had to be the case. All the subjects completed the same practice session and so all had already solved the same four syllogisms when the experiment proper began. The practice syllogisms were also included in the experimental session.

Once the practice session was completed, the subject proceeded to the main experiment. The subject read the first premise of the syllogism, then pressed the space bar on the computer keyboard to read the second premise. Once they had finished reading the second premise they pressed the space bar again and were asked whether the syllogism had a valid conclusion or not. They were required to answer "yes" or "no": they had been warned that once they had made a decision as to whether a valid conclusion existed or not and pressed the appropriate key, they could not change their mind. If they chose "no", the computer would print the message "No valid conclusion" and proceed. If they chose "yes", then the computer generated a menu consisting of the four quantifiers and the major and minor terms of the syllogism they had just read, the major and minor terms being in random order. An example response for the syllogism shown in Figure 2.1 is given in Figure 2.3

From this the subjects were instructed to choose a conclusion consisting of a quantifier and then the two terms, in the order in which they thought they should appear. For example the conclusion "none of the Germans are busdrivers" would be represented as "none Germans busdrivers", as shown above. Once the subject had chosen a conclusion, he or she pressed the RETURN key. The question "are you sure?" would then appear and the subject had the chance to change the answer. If the subject was sure of the answer he or she could respond "yes". If the subject attempted to choose more or less than three items for the conclusion, the message "try again, you should give three items" would appear.

menu	response
all	none
some	Germans
none	busdrivers
some not	
Germans	
busdrivers	

Figure 2.3: A sample conclusion

menu	response
all	all
some	busdrivers
none	cricketers
some not	none
busdrivers	cricketers
Germans	Germans
cricketers	

Figure 2.4: A sample recall menu

Once the conclusion was completed, the recall menu appeared. This consisted of seven items, the four quantifiers and the three nouns that the syllogism just read had contained. Figure 2.4 demonstrates what the subject would be shown and would have to produce having been given the syllogism in Figure 2.1.

The nouns were not in the same order as they had appeared in the syllogism, for example the middle term might appear first, the minor second and the major last. As noted above, these nouns were unique for each syllogism and acted as a cue for the syllogism concerned, so that the subject did not have to explicitly recall the nouns used in the syllogism. He or she was required to list six items from the menu, two groups of three consisting of a quantifier followed by two nouns, each group representing one premise of the syllogism. For example, if the syllogism contained the premise "All

menu	response
all	all
some	busdrivers
none	German
some not	none
cricketers	cricketers
Germans	German
busdrivers	

Figure 2.5: A sample menu at delayed recall

the butchers are French" the correct group of items to choose would be "All butchers French", however the subject was at liberty to choose any quantifier and any order for the nouns. Each premise had twenty-four different possible quantifier-noun combinations. Once the list was complete, the subject pressed the RETURN key and again was given the opportunity to correct the reply. Again, if there were fewer or more than six items in the list, the computer would prompt the subject to give the correct number. Once the recall was completed, the instruction to press the SPACE bar appeared and the subject could begin reading the next syllogism. This procedure was repeated for all thirty-two syllogisms in the session.

As soon as the recall for the thirty-second syllogism was completed, the instruction "recall the first syllogism" appeared, accompanied by the same menu that was used for the recall immediately after the first syllogism had been read, with the exception that the nouns were presented in a different order, as shown in Figure 2.5.

The subject could use the nouns in the menu as a cue for the syllogism in question as no other syllogism had contained these nouns. As before the subject had to choose six items from the menu which would capture all the information in the syllogism and the order in which the items occurred. Once the first syllogism had been recalled then a second menu appeared for the second syllogism and so on, until all thirty-two syllogisms had been recalled a second time. The subject was then told that the session was complete.

Exactly the same procedure was followed for the second session, which contained the thirty-two syllogisms not yet encountered in the first session and a separate set of nouns.

2.1.6 Summary of the data collected

The microcomputer used recorded the data for each subject. This consisted of an identifying code for each syllogism; a note of the exact nouns used in each syllogism; the conclusion given; a recall profile for immediate and delayed recall and the reading time for each premise of the syllogism. The data was analysed using the BMDP statistical analysis package.

The conclusion data consisted of the quantifier in the conclusion and the two nouns, or a code corresponding to the response "No Valid Conclusion". Each possible type of conclusion was therefore represented. The recall data consisted, as shown in Figures 2.4 and 2.5, of six items, the first three being the recall of the first premise, headed by the first quantifier and followed by the two nouns recalled as being in the premise, the second three corresponding to the second premise. All possible versions of the syllogism were captured in this way. There were a small number of responses that could not be classified, these consisted entirely of instances where the same noun was repeated in one premise. Errors of this type were excluded from the analysis.

The reading time for each premise of the syllogism was also recorded. In general the reading time for the second premise was much longer than for the first. The experiment was designed so that the syllogism would no longer be available when the subject reached the point of being asked for the solution to the problem. This resulted in the subjects processing the syllogism and reaching a conclusion for it when the second premise was still available. Thus the time taken to read the second premise corresponds to the time taken to process the syllogism and find an answer to it. The time taken to respond to the first premise relates to the time taken to form a simple representation of the information in the syllogism. Both measures of processing time will be used to test hypotheses about the syllogistic reasoning.

2.2 Experiment Two

In this experiment, subjects were given all sixty-four syllogisms, as in Experiment One, but rather than being asked to perform the standard task, were instructed to seek valid individuals, possessing all three properties in the syllogism in some respect, to report the individual should they exist and if one could not be found, to note this too.

2.2.1 Subjects

Twenty-one subjects were used, aged between 18 and 40, all graduates or undergraduates. Some were paid for their participation, others were volunteers. The same criteria for selected subjects were used for this experiment as for the first and for the same reasons.

2.2.2 Materials

The syllogisms were printed each on a separate piece of paper, with enough room for the subject to write her/his conclusion at the bottom.

2.2.3 Design

Every subject was intended to answer all sixty-four syllogisms. A fault in the program producing the syllogisms meant that the IE syllogism of figure AB–CB was replaced with the EI syllogism of the same figure. Not every session was faulty however, and the analysis has been adjusted to allow for this discrepancy. The syllogisms contained distinct sets of nouns, as in Experiment One. Each syllogism contained a nationality, an occupation and a hobby, pastime or other attribute. The syllogisms were presented in random order.

2.2.4 Presentation

Each subject was given a general explanation of what syllogisms consist of. They were given an example syllogism and the concept of a valid individual introduced. They were invited to attempt the first syllogism and discuss their response. An example without a valid individual was given and the subject asked to attempt it. Again their response was discussed until the subject felt confident that s/he understood the requirements of the task. The examples were chosen randomly from the subject's session and their responses not altered. Once the subject felt confident the nature of the task was grasped s/he was invited to begin solving the syllogisms. When these had been completed, the finished papers were collected and the subject thanked for his/her participation.

2.2.5 Analysis

The responses were transfered by hand to a master sheet, and this was then processed to a computer file, giving the subject number, response type and a score as to whether the response was correct or not. Three subjects had success rates beneath 50%. These were excluded from the analysis on the basis that their grasp of the task was too poor to provide meaningful results. Response type was divided into the individual actually chosen by the subject, represented by a series of positives and negatives and the order the nouns in the individual appeared in. A category for "no valid individual" was also used. Certain responses were unclassifiable, usually because the subject had given two responses without indicating his/her final choice or because only two of the three categories were actually used. These responses were excluded from the analysis. Further description of the types of possible individuals and other response data is given in chapter 4.

Chapter 3

Preliminary Investigation of the Recall Data

3.1 Recall of the Quantifiers

When recalling the syllogisms, the subjects indicate which quantifier they remember as having occurred in the first and second premises of the syllogism. The errors made when doing this are investigated in the following section. The first question to be addressed is whether there are any differences in the number of recall errors for the different types of quantifiers and whether there are differences in error rates between premises. Figure 3.1 shows the mean error rates for each premise, each type of quantifier and for immediate and delayed recall. The maximum number of errors for each category is sixteen.

A two-way Anova shows the significant differences between these means. There is a significant difference in numbers of errors for immediate and delayed recall ($F=118.33$, $p<0.01$). The type of quantifier is significant, for the first premise ($F=19.95$, $p<0.01$) and is significant for the second premise at delayed recall ($F=3.58$, $p<0.05$ from a one-way ANOVA). The interaction between the type of quantifier and immediate versus delayed recall is also significant for the first premise but not for the second ($F=8.12$, $p<0.01$; $F=1.96$, n.s.). This implies that the pattern of recall errors is significantly different at delayed recall compared to that at immediate recall, for the first premise. Comparing the first and second premises shows that there is no effect of premise for immediate recall, but that the second premise has significantly more errors than the first for delayed recall ($F=0.95$, n.s.; $F=18.84$, $p<0.01$). This comparison also shows significant interactions between the premise and the quantifier type, as might be expected, as this implies that the pattern of errors for each premise is significantly different ($F=1.409$,

First premise		
	immediate	delayed
all	0.56	3.32
some	1.08	6.88
none	1.88	6.88
some not	1.96	7.16
Second premise		
	immediate	delayed
all	1.80	8.40
some	1.60	6.96
none	1.76	7.80
some not	1.04	5.88

Figure 3.1: Mean recall errors for the quantifiers

$p<0.05$; $F=21.36$, $p<0.01$).

An t test shows that the difference between "All" and "Some..not" in the first premise is significant at immediate recall and the difference between "All" and the other three quantifiers is significant at delayed recall. The difference between "All" and "Some..not" for the second premise delayed recall is also significant, but there is no overall difference between means for immediate recall of this premise.

Figures 3.2 and 3.3 give the mean errors for quantifiers as a bar chart, for each premise at immediate and delayed recall. There clearly is an effect of quantifier type - "All" has fewest errors in the first premise and "some not" in the second. "Some" has second fewest errors in the second premise and in the first premise at immediate recall. These differences are significant as has been shown. The pattern of errors is similar for delayed and immediate recall, though the differences between errors rates by quantifier type is less pronounced at delayed recall.

To further investigate the effect of quantifier type on recall, the choices subjects actually made when recalling a particular quantifier are shown in figure 3.4. For example, if the first premise quantifier of the syllogism being recalled was "some" was the subject

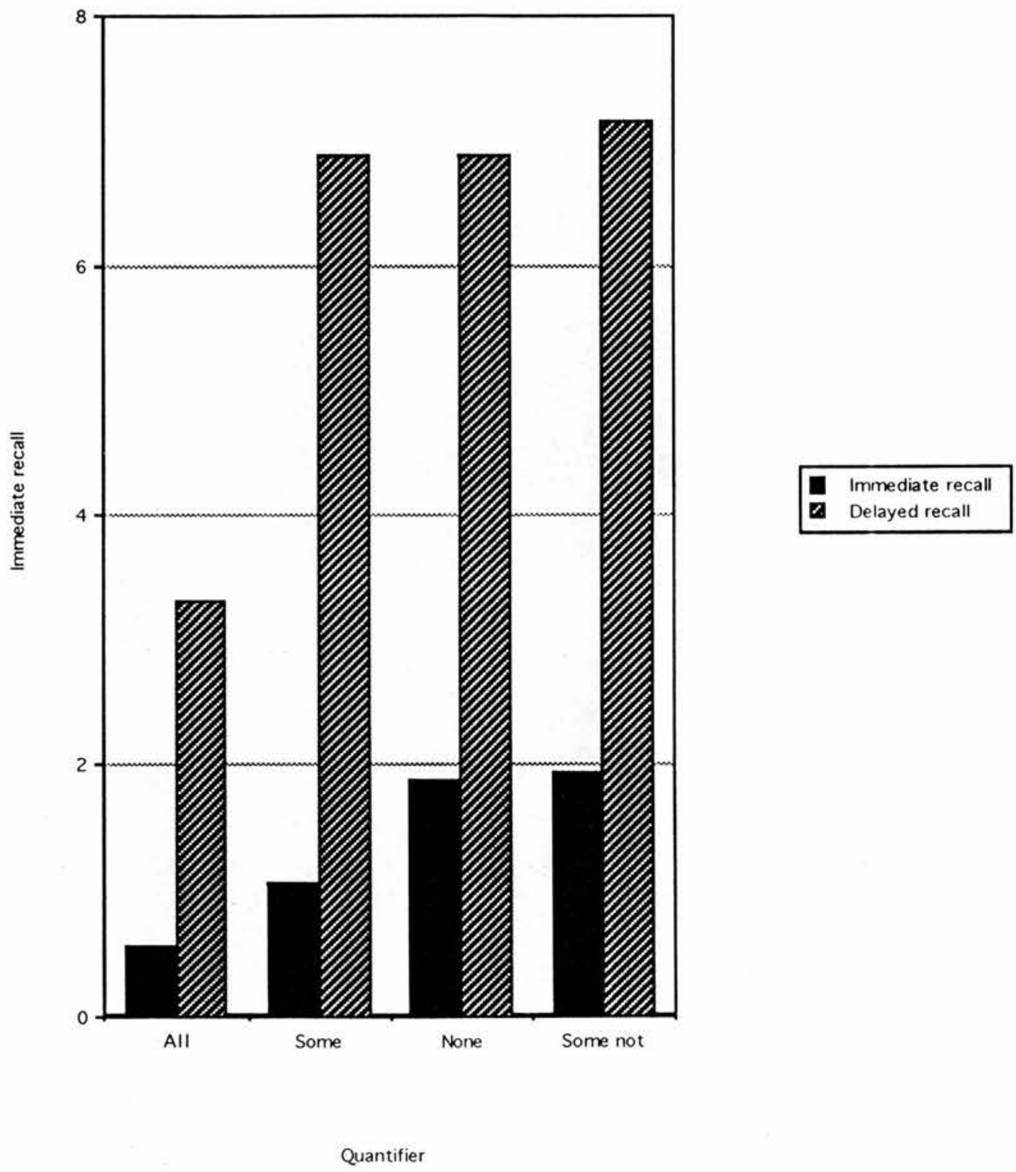


Figure 3.2: Mean errors for the first quantifier

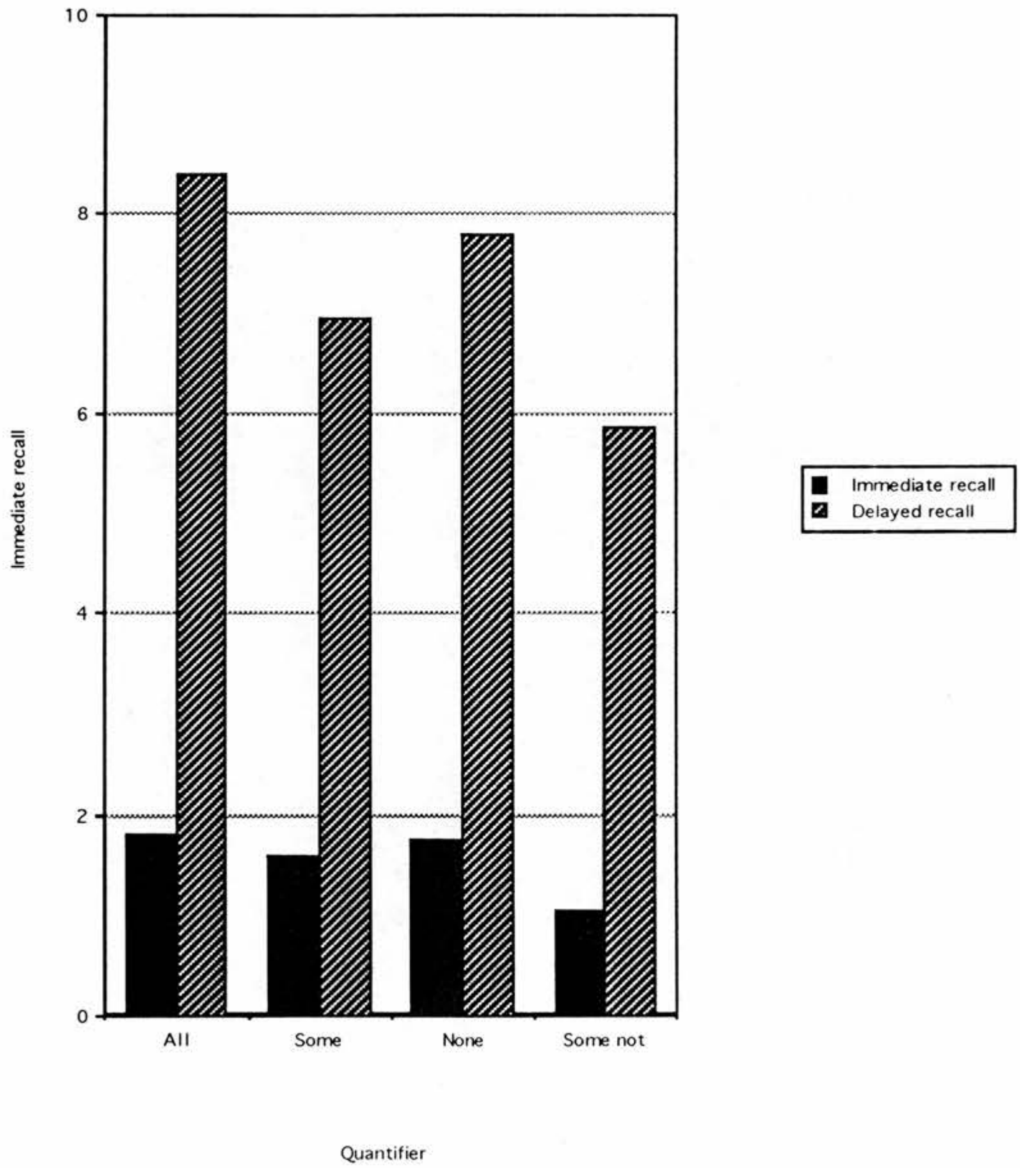


Figure 3.3: Mean errors for the second quantifier

equally likely to recall "all", "none" or "some not" when making an error.

Figure 3.4 demonstrates that "all" is a popular choice for the first premise quantifier. when an error is made and that "some..not" and "some" are preferred for the second premise. The patterns for immediate and delayed recall are roughly similar, but there is a larger proportion of errors for delayed recall. The recalled syllogism, even when incorrect tends to have a feature in common with the actual quantifier, thus "some" is recalled for "all" and "some..not", "all" and "some..not" for "some". Instances where the recalled quantifier has no features in common with the actual quantifier are uncommon, the only exception is when "none" was in the premise read, in this case "some" is the second most common quantifer recalled as an error.

3.1.1 Quantifier pairs by recall errors

Figure 3.5 shows the number of errors for each pair of quantifiers. The first premise quantifier is shown in the top row, the second quantifier is found in the leftmost column. The total errors for each pair are shown as is the number of times both quantifiers were wrong. Particularly high error rates are shown in bold type. Following the double errors are the number of times both quantifiers were wrong because they had been recalled in reverse order.

There is a cluster of high total error rates for quantifier pairs with "all" or "some" as the second quantifier and "none" or "some..not" as the first. The pairs "Some, None" and "Some..not, None" also attract many errors. The reason for some of these high error rates is shown by the number of double errors and reversals in the second table. Pairs with "all" as the second quantifier tend to be reversed so that "all" is the first quantifier. The pair "Some..not, Some" is also reversed to "Some, Some..not". The pair with high errors rates that is not reversed would be transformed by reversal to a pair that attracts a similarly high error rate. This is the case with "Some, None" and "None, Some". In the other cases reversals change a pair into one that would not attract many errors. These findings are in line with those above. Preferred pairs have "all" in the first premise and "some" "none" or "some..not" in the second. Pairs that violate this preference are liable to be reversed in memory so that they conform. The two pairs that are not covered by this explanation are "Some, None" and "None, Some". The same tables, for delayed recall, are shown in figure 3.6.

First Premise, Immediate Recall				
actual quantifier	chosen quantifier			
	all	some	none	some not
all	386	10	3	1
some	13	373	2	12
none	28	13	353	6
some not	4	45	0	351

Second premise, immediate recall				
actual quantifier	chosen quantifier			
	all	some	none	some not
all	355	17	18	10
some	8	360	1	31
none	7	14	356	23
some not	0	20	6	374

First premise, delayed recall				
actual quantifier	chosen quantifier			
	all	some	none	some not
all	317	42	17	24
some	82	228	24	66
none	69	51	228	52
some not	53	94	32	221

Second premise, delayed recall				
actual quantifier	chosen quantifier			
	all	some	none	some not
all	170	85	52	73
some	36	226	30	108
none	28	62	205	104
some not	12	109	26	253

Figure 3.4: Actual quantifier by chosen quantifier

Total errors	all	some	none	some not
all	2	12	23	19
some	6	6	21	22
none	11	17	9	23
some not	11	14	15	11

Both wrong	all	some	none	some not
all	0	9(8)	16(13)	7(4)
some	2(0)	0	5(0)	9(8)
none	3(2)	2(1)	2	4(0)
some not	0(0)	4(3)	4(0)	1

Figure 3.5: Errors for quantifier pairs, immediate recall

Total errors	all	some	none	some not
all	39	70	72	65
some	48	59	57	64
none	50	63	62	69
some not	40	69	54	65

Both wrong	all	some	none	some not
all	18	39(18)	40(14)	35(15)
some	13(4)	22	25(6)	34(15)
none	11(0)	27(4)	32	25(6)
some not	13(3)	21(11)	22(1)	19

Figure 3.6: Recall errors for quantifier pairs, delayed recall

The findings for delayed recall are similar to those for immediate recall, although fewer of the double errors can be explained as reversals. The latter are also not so strictly confined to pairs that are changed by reversal to pairs with low error rate. Perhaps at delay, memory for the order of the pair has deteriorated more than the memory for the actual items hence more pairs are recalled in incorrect order even though this may create a pair that does not accord with the general preferences for quantifier type in each premise outlined above.

Expressing quantifier errors in a different way can give further insight into the general trends underlying the individual mistakes that subjects make. Each of the four quantifiers can be expressed in terms of two features each of which have two values. One feature can have the value universal or particular the other the value positive or negative. The combination of these features that the quantifiers possess distinguishes it from the others. "All" is positive and universal, "some" positive and particular, "none" negative and universal and "some..not" negative and particular. This classification is useful because it helps to show how all the errors act together to change some features into others, including those where two features are changed. Figure 3.7 gives the absolute and percentage values for the first and second premises at immediate and delayed recall for each of the four possible ways the features can change when an error is made - universal goes to particular and vice versa, postive goes to negative and vice versa.

It can be seen from Figure 3.7 that for the first premise at immediate and delayed recall the strongest trend is for a negative feature to change to a positive, as would be expected with the tendency to choose 'all" for the first premise, but the next most frequent change is for a universal feature to change to a particular one at immediate recall and for this change and its opposite, from particular to universal to be almost equally frequent at delayed recall. This demonstrates that the preference for "all" is by no means overwhelming for the first premise. The second premise at immediate recall has two feature changes of almost equal frequency: universal to particular and positive to negative at delayed recall, changes from universal to particular have become a clear majority. Movement to no one particular quantifier can explain these changes, they demonstrate a trend that operates above the level of the individual items. These frequencies taken as a percentage of each group can be shown to become less pronounced from immediate to delayed recall, especially the change from negative to positive in the

	immediate recall		delayed recall	
	first premise	second premise	first premise	second premise
U to P	30	64	169	324
P to U	19	15	191	104
+ to -	18	60	131	263
- to +	90	41	267	211
U to P	19.1%	35.6%	22.3%	35.9%
P to U	12.1%	8.3%	25.2%	11.5%
+ to -	11.5%	33.3%	17.3%	29.2%
- to +	57.3%	22.8%	35.2%	23.4%

Figure 3.7: Feature change: absolute numbers and percentage

first premise.

3.1.2 Quantifier errors and the recall menu

It is apparent that recalling the quantifiers in syllogisms correctly or incorrectly does not purely depend on superficial characteristics of memory such as serial position. One example of this is the differential recall success for the quantifiers which depends on the type of the quantifier and the premise it appeared in. To summarise, it has been shown in the previous section that the quantifier "all" has fewer recall errors than the other three quantifers when it appears in the first premise and the quantifier "some..not" is similarly recalled more accurately than the other quantifiers when it appears in the second premise. In addition, subjects tend to choose these two quantifiers when they attempt to recall one of the others for the appropriate premise. There may be an obvious explanation for this effect. When the subject is asked to recall the syllogism, s/he are given a recall menu which contains firstly the four quantifiers, then the three nouns in the syllogism. The order of the nouns is randomised, because a particular feature of interest was whether the subjects would be able to correctly recall the order of the nouns in the syllogism. However, the order of the quantifiers in the recall menu was not randomised, but the quantifiers always presented in the order "all", "some", "none", "some..not". This has the effect that the quantifier "all' is always the highlighted item

First premise				
	all	some	none	some..not
Immediate recall	26.9%	27.6%	22.4%	23.1%
Delayed recall	32.6%	25.9%	18.8%	22.7%
Second premise				
	all	some	none	some..not
Immediate recall	23.1%	25.7%	23.8%	27.4%
Delayed recall	15.4%	30.1%	19.6%	33.6%

Figure 3.8: Percentage choices for quantifiers

when the subject is given the recall menu and must choose the first quantifier. When the identity of the actual quantifier is at all uncertain, the temptation must be, albeit unconsciously, to choose the first item on the list. This may contribute to the popularity of "all" as a choice for first quantifier.

Does this explanation completely explain the recall results for the first quantifier, however? For immediate recall the case that it does seems quite strong. "all" is generally the most popular choice of quantifier and where it is not "some" is instead and as this is the second quantifier in the menu, it seems reasonable to suppose that it will be second choice where "all" would be too different from the original - where "some..not" was the actual quantifier for example. If the same were to be true for delayed recall, one would expect the effect to be exaggerated: for as uncertainty about the identity of the quantifier increases, the likelihood of choosing one of the quantifiers from the beginning of the list should also increase. Figure 3.8 shows that the percentage of the total that each quantifier gets does increase in the favour of "all" at delayed recall, but that this is almost entirely at the expense of "none" and that "some" and "some..not" are not much affected in their relative popularity by delay. The arrangement of the menu cannot explain the fall in choices of "none" at delayed recall being greater than for "some..not" which is further from the beginning of the menu, nor the lack of an increase of choices of "some" which is next to "all" and might be expected to be chosen more frequently where there is uncertainty but 'all" would be too far from the original.

The second quantifier is also probably affected by the order of the quantifiers in the menu. When the subjects have completed their recall of the first premise, they find themselves positioned in the nouns composing the last three items in the menu. They must return to the top four items to choose the second quantifier and if they proceed back up the list from bottom to top, the first quantifier they will encounter will be "some..not". If, instead of proceeding up the list they choose to go down the list, the item at the top of the menu -"all"- will become highlighted. The fact that "some..not" is the most popular choice for the second quantifer may be due to many subjects proceeding back up the list as described and finding "some..not" the first quantifer they are able to choose. As hypothesised for the first quantifier, in cases of uncertainty, this may well be the quantifier they decide upon. However, continuing the explanation used for the first quantifier, "none" should then be the second most frequent choice of quantifier and this is not so. "Some" is the most popular choice of quantifier after "some..not" a finding that cannot be explained simply by the positions of the quantifiers in the menu. Order effects in the menu cannot be as strong an explanation for the choice of second quantifier as it is for the first, as the item chosen is not already highlighted when the subject comes to recall it, as is the case with the first quantifier. The subject must make a more active decision at this point in the recall to choose a particular item. Figure 3.8 shows an accentuation of the effect, but again the pattern of this effect cannot be explained by order effects in the menu except to note that the unpopularity of "all" in the second premise may be due to its overuse in the first - subjects may compensate for this by tending to ignore it when selecting the second premise. The accentuation of the effects of individual quantifiers upon recall should be placed alongside the finding that when the quantifier errors are expressed in terms of features, as in figure 3.7, the changes most pertinent to the quantifiers that might be affected by the menu order tend to diminish with delay, because of the cumulative effect of other types of changes.

When the choices for both premises are averaged and expressed as a percentage, as in figure 3.9 it can be seen that the effect of the second premise outweighs that of the first, producing an overall preference for the quantifiers "some" and "some..not" at both immediate and delayed recall. In cases of uncertainty, subjects are likely to want to make the most equivocal statements they can, the quantifiers "some" and "some..not"

	all	some	none	some..not
Immediate recall	25%	26.6%	23.1%	25.3%
Delayed recall	24%	28%	19.2%	28.2%

Figure 3.9: Percentage of overall choice

are the most likely to be true of the model formed from the syllogism and may be preferred for that reason.

3.1.3 Summary of the quantifier errors

It has been found that choices of quantifier are not uniform over the two premises. The first premise is significantly more likely to be recalled as containing the quantifier "all" and the second as having contained the quantifiers "some" or "some..not". When quantifier pairs occur that have these quantifiers but in the non-preferred order then there is a strong tendency to recall them as having occurred in the preferred order, that is in reverse order to the original. The fact that the recall menu always presented the quantifiers in a particular order suggests that this might be responsible for the trend and this is supported by the increase in certain choices at delayed recall. However, other features cannot be explained by the order of the recall menu, particularly the preference for "some" in the second premise and the overall unpopularity of "none"

3.2 Recall of Noun Order

3.2.1 Recall variables used

When analysing the recall of the order of the nouns in the syllogisms, a number of variables were used. The recall of the "rest of the syllogism", that is the nouns excluding the quantifiers, can be subdivided into two variables :- noun assignment, or whether the correct nouns were recalled as the major, middle and minor terms respectively, irrespective of whether the middle term was in the correct place and figure, or whether the middle term occurred in the correct position, irrespective of whether the nouns had been assigned correctly. For example the syllogism "All the French are butchers, some of the skiers are French" might be recalled as "All the skiers are French, some of the butchers are skiers" or "All the French are butchers, Some of the French are skiers".

Noun Assignment		
	immediate	delayed
ABBC	1.56	5.28
BACB	2.28	5.92
ABCB	1.64	5.72
BABC	2.52	6.24

Figure		
	immediate	delayed
ABBC	3.96	8.00
BACB	4.60	8.48
ABCB	3.76	8.80
BABC	3.40	9.04

Figure 3.10: Mean recall errors for noun order

In the first case the figure is correct (BACB) but the nouns are incorrectly assigned (middle term=skiers not French). In the second case the noun assignment is correct (middle term=French) but the figure is incorrect (BA–BC not BA–CB).

The mean errors in Figure 3.10 are out of a maximum of sixteen. Analysis of variance of the errors for noun assignment and figure shown in this figure confirm that delay significantly increases the number of errors for both variables ($F=90.04$, $p<0.01$ for noun assignment; $F=34.08$, $p<0.01$ for figure). The effect of figure type is significant for noun assignment ($F=2.84$, $p<0.05$) but not for figure. The interaction between delay and figure type is not significant for noun assignment but is for figure ($F=2.72$, $p<0.05$). One-way Analysis of variance of figure type by errors shows that figure recalled is significantly different by actual figure at immediate but not delayed recall ($F=3.97$, $p<0.05$). An t test shows that the difference in error rates for recalling figures BA–CB and BA–BC is significant at immediate recall.

Figures 3.11 and 3.12 show the relative success of recall of the roles of the nouns in and the figure of the syllogism for the four different figure types. For immediate and delayed recall figure type AB–BC has a slight advantage over the other types and BA–BC has the highest error rate for noun assignment for both immediate and delayed recall. At immediate recall, the figure BA–BC has fewest errors for recall of figure, at delayed recall it has most and AB–BC has least. It should be noted that the two figures that have a first premise of the form A–B have the lowest noun assignment errors at immediate recall and that at delayed recall the two figures with middle terms possessing different grammatical roles in each premise have fewer errors than those where the syllogism has the same grammatical role, in contrast to the error rates at immediate recall.

To further investigate the recall errors for the figure of the syllogism, the data has been adjusted to exclude those cases where noun assignment was incorrect. The effect of incorrectly recalling the noun assignment of the syllogism, even if the figure is correct, is often to produce a syllogism that has more in common with another figure than the original. For example the figure AB–BC might be recalled as BA–AC. The figure is correct, but the incorrect noun assignment has produced something closer to a syllogism with the figure BA–BC. The recall of the figure of the syllogism shows that the figure AB–BC tends to be the hardest to remember for immediate recall. This contrasts with delayed recall where the figures AB–BC and BA–CB have the lowest error rates and AB–CB the highest.

Figure 3.13 shows the kinds of errors the subjects made when recalling a particular figure. The figures have been adjusted to exclude all the cases where the noun assignment was incorrect. If this were not done then only the position of the nouns thought to be the major, middle and minor terms could be compared. In the cases considered, the terms A, B, C have been correctly assigned and the actual degree of similarity between the actual and recalled syllogism can be compared as A, B and C are the same noun in both the actual and recalled syllogism.

This figure demonstrates that the tendency at immediate recall is for the subject to choose a syllogism that has something in common with the one that was actually given. The preferred choice usually keeps the first two items in the syllogism the same, but keeping the second two the same is also much more popular than having no items the same. By the delayed recall stage the number of times the subject chooses a figure that has no items in common with the actual figure has risen a great deal, while the number of other errors has risen very little, in two cases not at all. The number of errors for figure overall has risen less than the errors for noun assignment, thus there are roughly

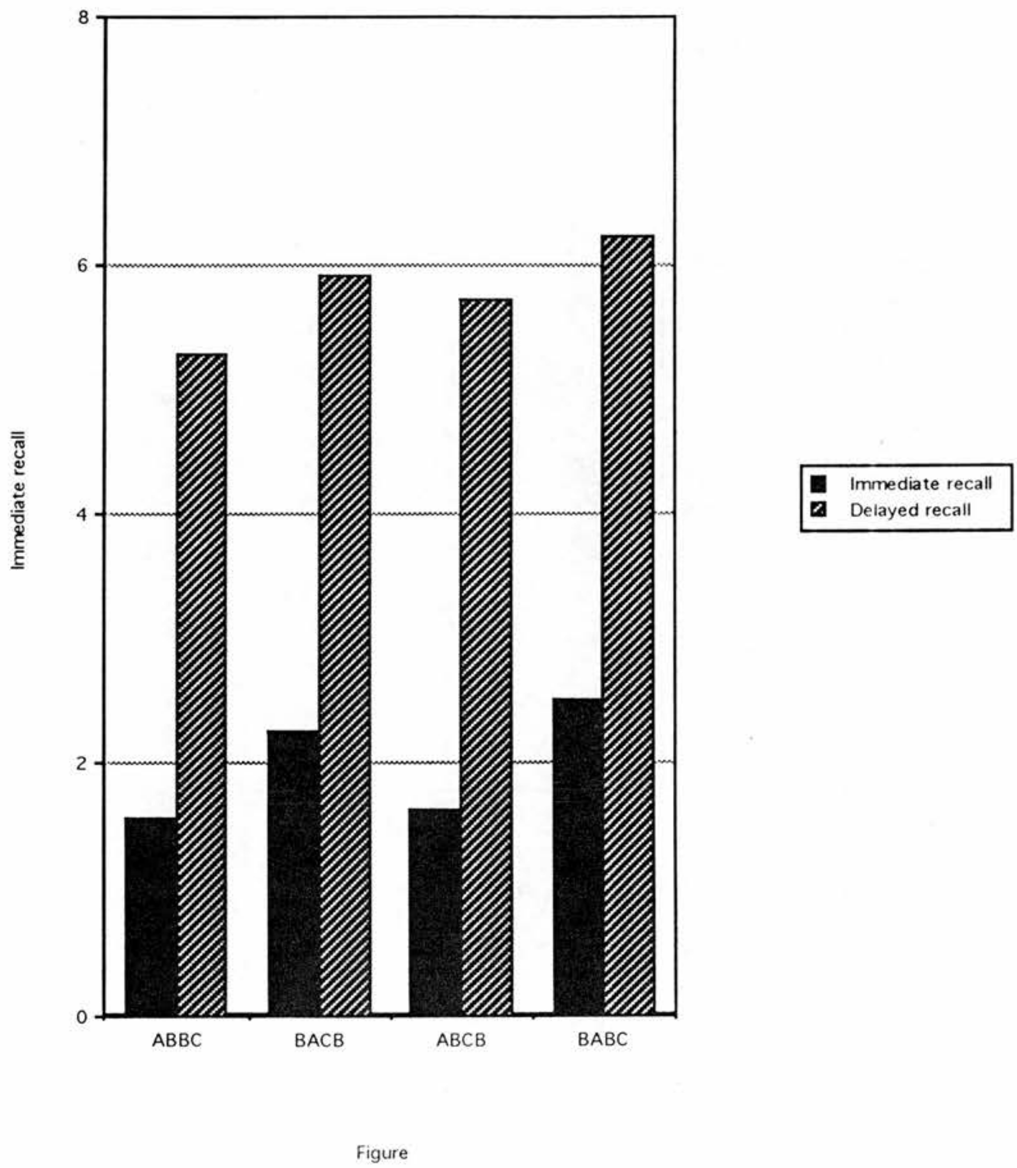


Figure 3.11: Mean error for noun assignment

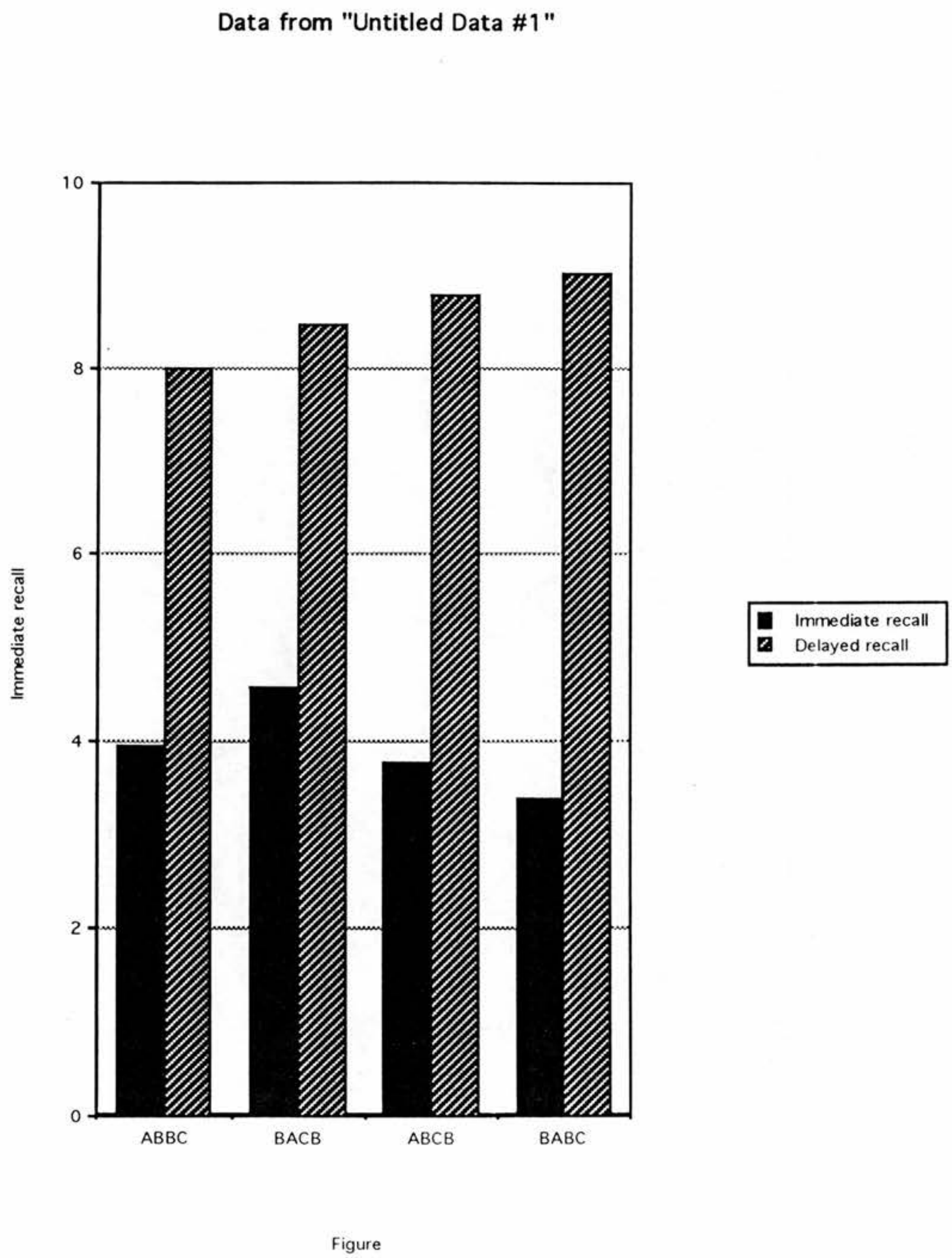


Figure 3.12: Mean error for figure

immediate recall					
actual figure	recalled figure				
	ABBC	BACB	ABCB	BABC	total
ABBC	280	8	25	32	180
BACB	7	259	21	38	169
ABCB	34	19	273	10	175
BABC	17	15	3	282	175
total	338	301	322	362	

delayed recall					
actual figure	recalled figure				
	ABBC	BACB	ABCB	BABC	total
ABBC	160	31	34	32	119
BACB	30	147	27	38	120
ABCB	44	37	140	29	120
BABC	40	49	11	132	115
total	276	264	212	231	

Figure 3.13: Actual figure by recalled figure

Class	Status
First Premise Nouns (A,B)	correct
Second Premise Nouns (B,C)	correct
First Premise Order (B,A)	correct
Second Premise Order (B,C)	correct

Table 3.1: Type 1 recall

the same number of texts at delayed recall that have an incorrect figure but correct noun assignment, as at immediate recall.

3.2.2 A preliminary model of recall of noun order

An examination of the frequencies of the various types of errors that subjects make when recalling the order of the nouns in the syllogisms, suggests an analysis that divides the process into a series of types.

Type 1

In these cases the subject has made no errors in the recall of the noun order of the syllogism. This is demonstrated for a syllogism of the form BA BC in table 3.1. At immediate recall 72.2% of syllogisms have made no errors (have a type 1 recall) at delayed recall this has dropped to 37.4%.

Type 2

Here, the identity of the nouns in each premise is correctly recalled, but their position in one or both of the premises is forgotten. This results in one of the premises having the correct nouns recalled in the wrong order, that is reversed. The commonest type of reversal error is for the first premise to be recalled correctly and the second to be reversed in the manner described. The next most common is for the first premise to be reversed and the second recalled correctly and the third for both premises to be reversed, that is only the content of the premises to be recalled, not the order of either premise.

Class	Status
First premise nouns (A,B)	correct
Second premise nouns (B,C)	correct
First premise order (B,A)	correct
Second premise order (C,B)	incorrect

Table 3.2: Type 2 error

	Immediate Recall	Delayed Recall
Second Premise Reversed	27.25%	17.38%
First Premise Reversed	20.80%	14.19%
Both premises reversed	6.5%	10.59%

Figure 3.14: Percentage of total error for Type 2 errors

The percentage of the total error that these error types accounts for is shown in Figure 3.14.

Type 3

Here, the subject can correctly recall the contents and order of one premise, usually the first, but cannot remember which item was repeated in the other premise. In such cases, the subject will not choose the original middle term in one premise to be repeated in the other, but will incorrectly choose the major or minor term to be repeated. In the majority of cases the position of the minor term is correctly recalled and the major term simply replaces the middle term in the recall. In other cases, the position of the minor term is also forgotten and it is recalled in the position the middle term originally occupied. The same error types are found for the case where the second premise has been correctly recalled and the first premise contains the error. An example of a type 3 error, using the classification for the syllogism of the form BA–BC shown in table 3.1 is given in table 3.3.

Figure 3.15 shows the percentage of total recall error that is accounted for by these error types. The commonest error type is for the first premise to be correct and the minor

Class	Status
First Premise Nouns (A,C)	incorrect
Second Premise Nouns (B,C)	correct
First Premise Order (C,A)	correct
Second Premise Order (B,C)	correct

Table 3.3: Type 3 error

C in correct position.		
	Immediate recall	Delayed recall
Major Term Swopped for Middle	11.26%	8.89%
Minor Term Swopped for Middle	3.38%	2.60%
C in incorrect position.		
	Immediate recall	Delayed recall
Major Term Swopped for Middle	6.08%	6.59%
Minor Term Swopped for Middle	1.35%	1.00%
Total	22.97%	19.08%

Figure 3.15: Percentage of total errors for Type 3 errors

term C to be in the correct position in the second premise. All that has occurred in this situation is for the major term to have been substituted into the second premise in the place of the middle term. The second most common error is for the first premise to be correct and the C term to have been moved to the other position in the second premise, with the A term occupying the original place of the C term. Clearly, information about the first premise is retained in preference to either noun assignment or order information in the second premise.

Type 4

The identity of the middle term is recalled at this stage, but its position in each premise has been at least partly forgotten. Most commonly the position of the middle term

Class	Status
First Premise Nouns (B,C)	incorrect
Second Premise Nouns (A,B)	incorrect
First Premise Nouns (B,C)	correct
Second Premise Nouns (A,B)	incorrect

Table 3.4: Type 4 error

	Immediate recall	Delayed recall
Middle Term Correct	10.81%	13.49%

Figure 3.16: Percentage of total errors that are Type 4

is correctly recalled, but the other two nouns are "swopped" that is the major term is recalled as having occurred in the second premise and the minor term as having occurred in the first and neither are repeated. in a smaller number of cases only the identity of the middle term has been recalled, not its correct position in each premise. An example is given in Table 3.4 following the syllogism described in Table 3.1.

This type of error involves recalling information that is forgotten in type 3 - namely the identity of the middle term. It is likely that the circumstances under which type 3 errors will be made and those under which type 4 errors will occur will differ therefore. Here, it will be noted that making a type 3 error requires a strategy that places most importance on recalling the exact contents of one premise and the use of the information that one term is repeated in both premises, however the wrong term is chosen to be repeated. type 4 errors occur when most importance is given to the identity of the middle term - this is correctly recalled, but the identities of the major and minor terms swopped . In this way each error type is complementary. The percentage of total error types that type 4 errors account for, shown in Figure 3.16, demonstrates that this type of error is less frequent overall than type 3 errors, although they become more common at delayed recall, while the overall percentage of type 3 errors drops from immediate to delayed recall.

	Types of Noun Order Error				
Information Preserved	Type 1	Type 2	Type 3	Type 4	Type 5
Figure	preserved	not preserved	not preserved	possibly preserved	possibly preserved
Noun Assignment	preserved	preserved	not preserved	not preserved	not preserved
Middle Term Identity	preserved	preserved	possibly preserved	preserved	not preserved
Major and Minor Term Identity	preserved	preserved	possibly preserved	not preserved	not preserved
At Least One Premise	preserved	preserved	possibly preserved	not preserved	not preserved

Figure 3.17: Information preserved or lost by noun order error types

Type 5

This category includes all the remaining errors that have been made. In these cases, it is possible that one of the noun roles has been correctly remembered and in a few cases the figure of the recalled syllogism may be the same as the original.

Figure 3.17 summarises the kind of information preserved and lost by each of these different types of error.

This demonstrates that there is a progression of loss of information moving from the top to the bottom of the table. There is some justification for thinking of these errors as representing poorer and poorer memory of the order of the nouns in the syllogism being remembered. It will be of interest to see if this type of progression can be linked to quantifier error and to features of the original syllogism and conclusion drawn.

3.3 Processing the Figure

In the preceding section, the error types committed by the subjects have been analysed into a number of categories. It has already been shown that the figures of the syllogisms are not uniformly represented in terms of different error types, of which these categories are composed. At immediate recall, the figures BA–BC and BA–CB both have higher levels of noun assignment errors (type 3 and type 4 errors capture most of these) than the other two figures. These findings might be explained by reference to the processing of the texts by the subjects. It is a feature of this experiment that the subject is obliged to read the first and second premises separately. This means that when the subject reads the first premise, s/he has no information about the nature of the second premise. This implies that the subject can have no certainty about the roles of the nouns in the first premise (which is the middle term, which is the major) or the figure of the syllogism until the second premise has been read. In some cases the whole of the second premise must be read before the whole structure of the syllogism is known, in others just the first noun in the second premise indicates the exact structure of the text. The subject is faced with a predicament when reading the first premise, therefore. Either s/he suspends the interpretation of the first premise until the second has been read and risks forgetting some of the information in it, or s/he makes an assumption about the identity of the nouns and figure and risks having to change this if it is incorrect. As the first premise has few recall errors it may be the case that the strategy is to make assumptions about the identity of the members of the first premise, so that it is well remembered, and risk some interference if this has to be changed.

When the first premise is read, the subject has a 50:50 chance of choosing the correct noun as being the repeated one. If s/he makes an incorrect choice about the identity of the middle term, there is an increased risk that at recall the term A will be remembered as being the repeated term because that will be the assumption that was made before the second premise was read. If the subjects have a bias towards choosing one or the other of the nouns in the first premise as being the repeated one, then this will affect the numbers of times this kind of recall error will happen by the form of the first premise. It was noted above that the two figures with B–A as the first premise are more prone to errors in the recall of the noun roles. It is hypothesised that this occurs because subjects tend to assume that the predicate of the premise is the repeated term. This preference will mean that subjects will be right about the identity of the middle term more often with AB–BC and AB–CB than with the other two figures, where one would expect to find higher numbers of errors where A has been incorrectly used as the middle

Immediate recall			
Noun Order	Both Right	Error	Reversed
Type 1 Error	91.0%	8.0%	1.0%
Type 2 Error	79.1%	19.4%	1.6%
Type 3 Error	72.7%	24.2%	3.1%
Type 4 Error	45.8%	12.5%	41.7%
Type 5 Error	56.3%	37.4%	6.3%
Delayed recall			
Noun order	Both Right	Error	Reversed
Type 1 Error	60.6%	37.6%	1.8%
Type 2 Error	42.4%	53.6%	4.0%
Type 3 Error	27.6%	64.7%	6.4%
Type 4 Error	14.1%	60.1%	25.9%
Type 5 Error	15.3%	73.5%	11.2%

Figure 3.18: Percentage noun order errors by quantifier error

term in the second premise. For the figures BA–CB and BA–BC 5% and 8% of the total incidence of each figure have this error at immediate recall, compared with 3.7% and 3.2% for figures AB–BC and AB–CB.

3.4 Interactions Between Recall Variables

Figure 3.18 gives the percentage of each type of noun order error found for quantifier errors across the first and second premise. Because the interesting interactions seem to involve differences in noun order error rather than quantifier error, the percentage of each type of quantifier error has been calculated for each category of order error. The changes for quantifier error by noun order error uniformly show a drift towards deeper levels of error for noun order as recall of the quantifiers deteriorates. This demonstrates a clear interaction between the two recall variables. Immediate recall is generally good and when there are no noun order errors (type 1) there tend to be few quantifier errors.

The relationship between quantifier and noun order recall is also demonstrated by the number of instances where neither quantifier was recalled correctly but not because

Delayed Recall		
	First Quantifier	Second Quantifier
All	39.2%	51.7%
Some	30.7%	34.2%
None	43.2%	39.9%
Some..not	38.0%	31.8%

Figure 3.19: Percentage of correctly recalled texts by recalled quantifiers

they were reversed. This type of error increases with type of noun order error, with the exception of type 4 errors. These are unusual in having much higher levels of cases where the quantifiers were correctly recalled but in reverse order. As this type of error consists of reversing the roles and therefore the order of the major and minor terms this implies that reversal of one kind of information is related to a reversal in another type of information.

The type of quantifier found in the first or second premise might affect the recall of the nouns in the premise. Certain quantifiers are preferred in each premise at recall and whether this has any effect on noun order error should be determined. This can be sub-divided into possible effects of the original quantifiers of the syllogism and the effect of recalling a particular quantifier, whether this was the same as the original or not.

As there are similar effects on accuracy of recall of nouns for both the original and recalled quantifiers, figure 3.19 shows the percentage of the occurrences of each type of quantifier at recall where there were no errors in the recall of the nouns. There is an interaction with both the first and second recalled quantifiers individually but not between quantifier pairs and correctly recalled nouns at delay. There is no interaction at immediate recall. For both the first and second quantifier "all" and "none" are chosen in circumstances where memory for other parts of the syllogism is good. "Some" and "some..not" do not have such high proportions of correctly recalled noun orders, particularly "some". The fact that the effects for the first and second premise are roughly similar demonstrates that the preference for "all" in the first premise can not simply be the result of guessing at the first item in the menu when memory is poor. There may be an element of this, hence the lower percentage of correct noun recalls when "all" is chosen for the first premise than when it is chosen for the second, but it cannot be the only cause of this trend. Choosing "some" is always associated with poorer memory and to a lesser extent "some..not". The summarised effect is that choosing universal quantifiers is related to good memory otherwise, choosing particulars when there is uncertainty about other parts of the text. The fact that the effect is also found for the original quantifiers allows one to go further and suggest that particular quantifiers do not allow as secure a memory for noun order as universal quantifiers. Similar effects are found both for type 2 and type 3 errors, so either encountering or recalling a universal quantifier seems to be related to fewer noun reversals within a premise and to better recall of noun roles. Part of this may be due to the logic of certain quantifiers only allowing one order of the nouns in its premise, if its logic is to be preserved. The influence of this will be discussed in a later chapter.

One other effect should be noted. When choosing a subject for recall of the premises, subjects have a tendency to prefer B as the subject term in both the first and second premises. This is also associated with a particular type of quantifier change. Referring back to the classification of the quantifier changes used above, when a quantifier in the first premise changes from negative to positive, there is a stronger tendency to recall the first term as having been B at both immediate and delayed recall. The other feature worthy of note is that the preference for B as subject is accentuated by correct recall of the quantifier of the premise in which it is to appear. These effects are summarised in figure 3.20

3.5 Summary

Some main effects have been found in the recall data. The quantifiers show large differences in recall success depending on premise and tend to reverse their order to suit an underlying preference for positive quantifiers to occur in the first premise and particular ones in the second. The recall errors bear out this preference. The possibility that the recall menu could be responsible for this trend is not supported. The recall of the noun order shows a tendency for BA–BC to be the preferred figure at immediate

Immediate Recall			
	B Subject	A Subject	C Subject
First Premise	49.3%	47.3%	3.4%
Second Premise	48.8%	5.5%	45.7%
No Quantifier	49.4%	48.2%	2.4%
Error	49.7%	4.5%	45.7%
- to + Change First Premise	51.1%	40.0%	8.9%
Delayed Recall			
	B Subject	A Subject	C Subject
First Premise	47.7%	42.7%	9.6%
Second Premise	40.3%	16.6%	43.0%
No Quantifier	50.7%	44.8%	4.5%
Error	45.3%	11.1%	43.5%
- to + Change First Premise	46.1%	38.6%	15.4%

Figure 3.20: Percentage of recalls with each noun as subject by total, recall without quantifier error and negative to positive changes

recall and AB–BC at delay. The recall of the noun roles (or noun assignment) varies according to whether the first premise of the figure was of the form B–A or A–B, errors being more prevalent in the former case. It is hypothesised that this is the result of subjects tending to assume that the predicate of the first premise is the repeated one and therefore mistakenly repeating A in the second premise, recalling this as the repeated term.

The noun order errors can be subdivided into various types and these related to recall of the quantifiers. It is found that there seems to be a progression of noun order error, with reversals of nouns within a premise type 2 errors occurring more frequently when there are no quantifier errors than type 3 errors, where the major or minor term is mistakenly recalled as the middle term. The finding that noun roles are better remembered on the whole than figure, suggests that forgetting this information represents quite a major disruption in memory. It is also found that reversal of quantifier order tends to occur with reversal of the major and minor terms (type 1 errors). The types of quantifier also play a part in the recall of the nouns. This is better (as represented by type 1 recall) when the recalled or original quantifiers are universal and there are more type 2 and 3 errors when the quantifier was particular suggesting that universal quantifiers lead to better consolidation of the rest of the premise they appear in.

The preferences for the quantifier will later be explained in terms of the ordering of negative and positive information, but the interaction between quantifier and noun order recall suggests that particular information is found in the second premise because it is less determinate than universal information and must therefore act as qualification rather than introductory material. The fact that only "all" is both positive and universal explains why this quantifier is so often found recalled in the first premise. Characterictic errors in the recall for the noun order will be explained in terms of the subjects' attempts to find a topic or focus for the argument the syllogism is about.

Chapter 4

The Reasoning Data

Experiment One allowed subjects to choose any form of conclusion they wished with the restriction that if a conclusion be drawn the two nouns in it be the subject and predicate terms. The answers given are summarised by syllogism in Appendix A. Experiment Two asked subjects to find valid individuals rather than conclusions. These results are given in Appendix B. A general discussion of the reasoning results found for each experiment is given and factors such as the number of models required to solve the syllogism and the figural effect are considered. Anomalies and findings peculiar to each experiment are also discussed and the implications of these for a formulation of the reasoning strategy of the subjects summarised.

4.1 General Findings: Experiment One

The percentage of answers that were correct overall was 63.9%. Johnson-Laird and Steedman (1978) found a correct response rate of 58% for the first trial and 68% for the second. The percentage of syllogisms with no valid conclusion that were correctly answered was 69.3%, so there seems to be a difference between subjects ability to give the response "no valid conclusion" correctly and to report a correct conclusion.

After Johnson-Laird the syllogisms are divided into those requiring one, two or three models to be given a correct valid conclusion and two and three models to be found to have no valid conclusion and the percentage correct responses shown in Figure 4.1.

The claim that the number of models affects performance is upheld by this data ($F=39.69$, $p<0.01$) although there is clearly not a linear relationship between number of models and percentage correct answers. In fact the difference between two and three

	Valid Conclusion	No Valid Conclusion
One Model	87.6%	–
Two Model	41.6%	75.3%
Three Model	37%	47.5%

Figure 4.1: Percentage correct conclusions by model: Experiment One

model syllogisms with a valid conclusion is not significant (t=0.83 n.s.). Again, performance is improved on the syllogisms without a valid conclusion. Although Johnson-Laird claims models are constructed for these syllogisms as for those with a conclusion, the difference in subjects' success at correctly answering "no valid conclusion" and drawing a correct conclusion tends to suggest that the correspondence between these two types of syllogisms is not as close as he suggests. Testing shows that two model syllogisms with a valid conclusion are answered with significantly less accuracy than two model syllogisms without a valid conclusion ($t=7.67$, $p<0.00001$). The same is also true of three model syllogisms with and without valid conclusions ($t=2.95$, $p<0.003$). Given that the explanation for differences in error rates lies with the number of models required to solve the syllogism, in that the greater the number of models the greater the likelihood that memory will fail, it is difficult to see how such a hypothesis can account for these differences. The lower number of errors for syllogisms without a valid conclusion suggests that subjects are able to detect the lack of a valid conclusion without cycling through all the models required and are therefore avoiding the major source of error.

4.1.1 Syllogisms with no valid conclusion

A closer examination of two model syllogisms with no valid conclusion reveals differences in error rates within the group. Syllogisms with the quantifiers "none" and "some..not" in whichever order are sometimes prone to higher error rates (see figure 4.2).

The type of quantifiers affects the ability of the subjects to correctly solve the syllogism ($t = 5.79$, $p<0.00001$). A similar effect is found for Johnson-Laird and Steedman's (1978) data. For example in the first experiment, subjects achieved the percentage correct answers shown in Figure 4.3.

Type of Syllogism	% Correct
None/Some..not	61.5%
Other Two-model Syllogisms	79.8%

Figure 4.2: Correct responses for syllogisms with no valid conclusion

Type of Syllogism	% Correct
None/Some..not	65.0%
Other Two-model Syllogisms	72.9%

Figure 4.3: Johnson-Laird and Steedman's data

As all these syllogisms require two models for a correct answer to be given, there is no explanation for this effect in these terms, as the probability of finding the correct response should be the same. Another explanation for this effect must then be found. Examining the syllogisms with the quantifiers "none" and "some..not" it can be seen that the error rates within this group are not uniformly affected. Figure 4.4 demonstrates this.

As shown, these asymmetries are reflected in the number of correct responses for the syllogisms with the quantifiers "none" and "some", but in the opposite direction. This tendency is particularly strong for the two figures AB–BC and BA–CB and the asymmetry for the "none, some" pairs is related to the figural effect. The cases with high correct responses are those where a figural conclusion exists and those with low

Figure	None/Some..not	Some..not/None	None/Some	Some/None
AB–BC	19	13	6	17
BA–CB	13	21	15	5
AB–CB	13	20	10	8
BA–BC	14	10	14	12

Figure 4.4: Number of correct responses for None/Some..not syllogisms

rates, where an antifigural conclusion only is correct. In the latter case, high levels of incorrectly given NVC[1] reponses are found and this provides the probable answer to the asymmetries in the "none, some..not" pairs. It has already been mentioned that "some" and "some..not" are easily confused. Recall data is used above and will be used later to demonstrate this point. If this is taken to be the case, then these conclusion rates can be explained. If "some" is substituted for "some..not" then in the figures AB–BC and BA–CB there will be a strong temptation to report an erroneous conclusion in cases where a figural conclusion exists for the "none, some" pair. In cases where transforming "some..not" into "some" results in a syllogism that has an antifigural conclusion, there will be a tendency to report "no valid conclusion" as happens for the "none, some" pair and there will be an apparently higher correct answer rate. The asymmetries for the two other figures are harder to explain. In the case of BA–BC the asymmetry is small and as would be expected the correct response rates are low for the "none, some..not" syllogisms as the responses to the "none/some" cases are not affected by figure. The asymmetry for AB–CB may be related to the very faint asymmetry found for the "none, some" pairs in this figure, or perhaps to the rather larger asymmetry in incorrectly given NVC responses for these syllogisms - "some, none", to which "some..not, none" is supposed to be equivalent, has 15 NVC responses, while "none, some", to which "none, some..not" is supposed to be equivalent has only 9. This reflects the asymmetry of the equivalent "none, some..not" pairs, but the explanation for this has yet to be found.

4.2 General Effects: Experiment Two

Rather than perform the standard[2] task subject were here asked to try to find a valid individual for the syllogism. Validity in this case is defined in the same way as for a valid conclusion: the individual found must describe a state of affairs that necessarily exists, whatever the interpretation of the syllogism. All three nouns in the syllogism must be used, but the individual can be asserted to have or not to have any combination of these properties. There are eight distinct types of individual the subject can choose from plus six orders in which the nouns can appear, so for example if a subject chose

[1]Here, as in other parts of the text, the response "no valid conclusion" is referred to by its initals, for brevity's sake

[2]The paradigm where subjects are asked to search for and report valid conclusions linking the terms A and C will be referred to as the"standard" task.

the individual +A-B+C the nouns could appear in this order, or in the order -B+A+C or +C+A-B and so on. The full list of types of individuals is given in figure 4.5. The subjects also have available the response "no valid individual"[3] if they are unable to find an individual that meets the criteria for validity. It can be seen, of course that individuals with one or two terms negated are the most common, in practice there are seven syllogisms that have an individual with all positive terms, twenty where the valid individual has one negative, eight where two negative terms are needed and none where all three terms can be negative. The latter response is chosen occasionally, but it is very rare.

Dividing the syllogisms by model as was done for the data in Experiment One shows a rather different pattern from that found above for the standard task. Syllogisms with no valid individual cannot be divided into two and three models in quite the same way as before because the groups do not necessarily contain the same members[4] and this would make a comparison meaningless. All no valid individual syllogisms are given as one group therefore. The syllogisms with valid conclusions all have valid individuals also, so a direct comparision is possible. In figure 4.6 a difference can be seen between syllogisms of only one model and those with two or three models, but the progression of difficulty between two and three model syllogisms is not found, in fact two model syllogisms prove to be a little harder than their three model counterparts. Syllogisms without a valid conclusion also show the advantage found in Experiment One having a rather higher success rate than two and three model syllogisms. The differences within the groups as a whole is significant ($F=8.7$ $p<0.001$) and testing between groups shows that the one model syllogisms fare significantly better than all other types of syllogism and that those with NVI better than two and three model syllogisms that do have a valid individual.

4.2.1 Individuals where no valid conclusion exists

When Stenning's modified Euler's circles notation is adopted, it has been shown that certain syllogisms that do not normally have a conclusion do have a valid individual as mentioned above, revealed by the presence of double hatching. The EE syllogisms always

[3] This will be occasionally referred to as NVI for short

[4] This is due to the fact that some syllogisms that have no valid conclusion do have a valid individual and therefore these syllogisms have been excluded from this analysis

No Negative Terms	One Negative Term	Two Negatives Terms	All Negative Terms
hline +A+B+C	+A+B-C	+A-B-C	-A-B-C
	+A-B+C	-A+B-C	
	-A+B+C	-A-B+C	

Figure 4.5: Types of individual for Experiment Two

Number of Models	% Correct
One Model	92.1%
Two Model	73.7%
Three Model	74.1%
NVI	81.3%

Figure 4.6: Percentage correct responses by number of models: Experiment Two

have individuals of this type, so do three of the EO and three of the OE syllogisms. From these syllogisms, examples of which are shown in figure 4.7 can be drawn the fact that an individual of the form +B-A-C exists. The data from Experiment Two has been analysed to see if subjects are capable of finding these conclusions. A comparison of the percentage of correct responses is given in figure 4.8 for the syllogisms that do yield an individual and the EO, OE and OO syllogisms that do not. From this it can be seen that subjects do find the individuals, but that they are less accurate than for similar syllogisms that do not have a valid individual, or making the comparison to figure 4.6, than syllogisms in the standard task that have two or three models. The presence of these individuals that cannot be expressed in terms of conclusions, does not affect the standard task. In fact syllogisms with no valid conclusion that are double hatched are easier not harder to solve. Therefore the presence of one of these individuals does not confuse the subject into believing that a conclusion also exists.

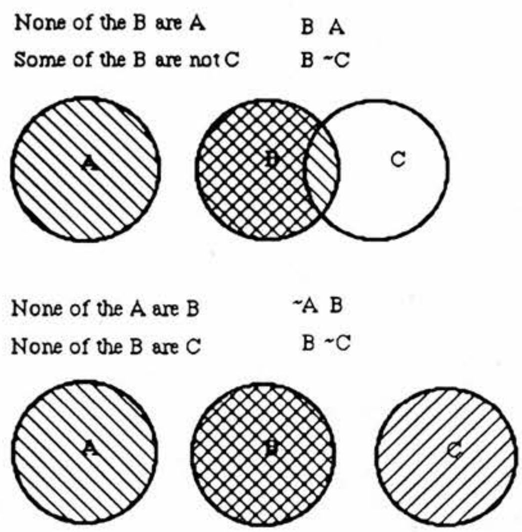


Figure 4.7: Syllogisms with no valid conclusion but a valid individual

Syllogism Type	% Correct
Individual Exists	63.8%
No Individual Exists	84.2%

Figure 4.8: Percentage correct responses for syllogisms with no valid conclusion, but where in some cases an individual does exist

	Overall Success Rate	A–C Conclusions	C–A Conclusions
AB–BC	63.9%	78%	22%
BA–CB	62.75%	24.4%	75.6%
AB–CB	64.75%	50.6%	49.4%
BA–BC	64.5%	58.1%	41.9%

Figure 4.9: Choice of conclusion by figure: Experiment One

Figure	Overall Success Rate	A first	B first	C first
AB–BC	78.3%	60.5%	35.9%	3.6%
BA–CB	74.8%	4.7%	41.9%	53.4%
AB–CB	74.3%	35.4%	20.0%	44.5%
BA–BC	88.2%	6.0%	91.2%	2.8%

Figure 4.10: Percentage of each noun in first position by figure: Experiment Two

4.3 The Figural Effect

There is no significant difference in the success rate for correct conclusions between figures in Experiment One but a strong figural effect is found in Figure 4.9.

Overall there are slightly more A–C type conclusions (52.56%) than C–A (47.44%) possibly reflecting a tendency to choose the first premise noun for the subject of the conclusion, where other factors such as the figural effect do not operate ($t=1.511$, $p<0.05$). There is a significant effect of figure on the frequencies of different types of conclusions ($F=6.75$, $p<0.01$).

A figural effect can also be found for Experiment Two by comparing the numbers of times each noun appeared in first position when an individual was reported. Comparing the percentage of the responses for each noun and figure demonstrates a strong figural effect that is significant ($F=16.85$ $p<0.00001$). The overall success rate is significantly different for the four figures ($F=7.7$ $p<0.001$), with the difference between BA–BC and the other figures being significant. Thus, subjects find this figure easier to solve than the other three.

Type of Syllogism	% of Incorrect Conclusions	% of Total Responses
NVC	92.3%	not applicable
One Model	78.9%	44.1%
Two Models	84.1%	49.3%
Three Models	88.6%	36.3%

Figure 4.11: Atmosphere response as a percentage of incorrect responses

4.4 The Atmosphere Effect

Strong evidence has been presented that subjects do use an agglomerative representation followed by cycles of testing to solve syllogisms. Discrepancies in the results for syllogisms with no valid conclusion have been investigated, further discussion of this follows in a later section. As previously stated however, there are claims that subjects do not necessarily draw conclusions by the mental models method but are, consistently or on occasion, using the "atmosphere" of the syllogism to generate conclusions. Many correct conclusions are in line with atmosphere anyway, particularly those requiring only one model to be solved. Two and three model syllogisms often have conclusions not in line with atmosphere, syllogisms with no valid conclusion never do, of course. Instances where an incorrect conclusion was given, either because none existed, or because a different conclusion was correct are examined to see what proportion are in accordance with the atmosphere effect. Figure 4.11 gives the results.

Clearly, when an incorrect conclusion is drawn, it is overwhelmingly likely that it will be in accordance with atmosphere. This is particularly the case for syllogisms that have no valid conclusion and the tendency increases steadily with the number of models. However it is also not true that all conclusions are obtained by this method, otherwise "no valid conclusion" would never be reported. Politzer(1989) suggests that in cases where the capacity of the subjects is outreached, in this case one presumes memory capacity, then conclusions will be given according to atmosphere. One would therefore expect increasing numbers of atmospheric conclusions as the numbers of models needed to solve the syllogisms increases, both in terms of the proportion of incorrect conclusions given and the proportion of incorrect answers given. As already stated the former is found, as the number of models increases, so the percentage of incorrect conclusions that are atmospheric increases. The same is not true however of the proportion of the total incorrect responses that are atmospheric. This increases between syllogisms with one or two models, but declines sharply for three model syllogisms. The reason for this, also shown in figure 4.11 is that the number of NVC responses increases sharply at this point. This suggests that rather than draw conclusions in line with atmosphere when their capacity to solve a syllogism rationally is exhausted, subjects will give the response "no valid conclusion". The atmosphere effect then appears more to be an underlying trend, perhaps as the result of a group of subjects failing to combine and test the premises and using a representation of the form Politzer suggests.

4.4.1 Matching

Gilhooley and Wetherick claim to have found a group of subjects who do not interpret the syllogism in such a way that conclusions can be rationally drawn. As already discussed they identify a group of subjects whose characteristic performance consists of finding a conclusion whose quantifier matches that of the original syllogism.

This group was found by using a set of six syllogisms that are particularly tempting for subjects who might be using this strategy, as both the quantifiers are the same and the syllogism actually possesses no valid conclusion, making matching responses easy to identify. Here, the syllogisms that can be used to identify matching responses have been extended to cover all the syllogisms where both quantifiers are the same and no valid conclusion exists, to combat sparsity of data.

To show that matching is a strategy explicitly adopted by certain subjects, the number of correct conclusions can be correlated with the number of syllogisms from the subgroup that were given a "matching" conclusion. If some subjects are not treating the syllogisms as logical problems and are generating solutions for them by choosing a quantifier from the syllogism, then their overall success rate will be lower than it will be for subjects who approach the syllogisms logically, as matching cannot generate a correct conclusion in every case. Correlating number of correct conclusions with number of matches gives $r=-0.502$ $p<0.005$. In other words, the higher the number of correct conclusions, the lower the number of matches on the identifying subgroup.

Matchers might also be expected to use the response "no valid conclusion" rather less frequently than those using a logical strategy, as matching cannot generate this response. A correlation between the number of "no valid conclusion" responses, excluding the diagnostic syllogisms, and the number of matches gives $r=-0.469$, $p<0.01$, showing that the number of matches is inversely related to the number of times a subject responds that there is no valid conclusion.

It is also the case that the number of matches a subject produces correlates significantly with the number of A–C conclusions s/he has produced. The higher the number of matches, the higher the number of A–C conclusions ($r=+0.481$, $p<0.05$). There is a small negative correlation between the number of matches and the number of C–A conclusions, which is not significant ($r=-0.12$, n.s.). "Matching" subjects prefer A–C conclusions, as do the subjects as a whole, but the more likely subjects are to choose non-logical responses, the more likely they are to chose A–C conclusions. The preference for A–C conclusions appears to be related to a non-logical approach to syllogisms.

A sub-group of subjects can thus be identified that are generating conclusions without considering the logic of the syllogisms whatsoever. They are obviously regarding the task as a question of interpreting the likely outcome of a state of affairs and are basing their judgements on a commonsense evaluation of the information they have been given. In the case of syllogisms where both quantifiers are the same it is manifestly sensible to assume that the same relationship holds between the two elements of the conclusion. It is most helpful to think of matching as the part of the atmosphere effect that relates to a particular group of syllogisms. It is to be expected then that the group of matchers found, will go on to use the other principles of the atmosphere effect to generate conclusions for syllogisms where the two quantifiers are different.

4.4.2 The relevance of the atmosphere effect

It might be argued that the existence of a group of subjects that are not using a rational approach to solving syllogisms is of little relevance to the mental models paradigm. If the latter is taken as the normative way subjects who comprehend the logic of the syllogism set about finding a correct solution to it, it seems of minor interest to note that some subjects do not grasp the logical requirements of the task and find other ways to provide a response. As already argued however, it is important to establish that such subjects exist in order to be able to answer the question "Do all subjects, irrespective of strategy use the same representation for the syllogism before seeking a response?". From Johnson-Laird's point of view it is important to claim that at some level all representations are the same and are mental models, because this conception of the representation is crucial to his explanation of the figural effect (see below). Figure 4.12 demonstrates that the figural effect is common to all subjects, even when those subjects using matching responses are considered separately from those that are not. An Anova demonstrates that the effect of figure is significant ($F=7.48$ $p<0.0001$) and that there is an effect of being assigned to the group "matching" or "non-matching" ($F=5.12$ $p<0.02$). This relates to the percentage of A–C and C–A conclusion given by the two groups. For the non-matchers the proportions of the two response types are nearly equal(50.1% for A–C, 49.9% for C–A) while for the matchers there is tendency to draw more A–C conclusions (57.4% versus 42.6%).

The figural effect does not depend therefore on the strategy used to solve the syllogism, although it is affected by it. One candidate theory is that all subjects use a common strategy or representation that predisposes the choice of certain conclusions following certain syllogisms and then deductive or non-deductive strategies are used. Johnson-Laird's tableaux are certainly a candidate for this. The alternative is to suppose that the figural effect has two causes: for subjects not using a deductive approach it is the result of, for example, a representation such as the one Politzer proposes, discussed in chapter 1. For subjects who do grasp the logical requirements of the syllogism, tableaux are used. One's preference is naturally for the former argument, it stretches credulity a little to believe that an effect as strong and consistent as the figural effect is in reality the result of two quite different mechanisms, which may change in the course of one experiment.

The question is to determine if all subjects do in fact use a representation such as the one that Johnson-Laird proposes and that some just fail to use this representation in any deductive way. One interesting comparison to make is between the performance of subjects making greater than average matching responses and those making less than average on syllogisms that require only one model for a correct valid conclusion to be drawn. In a case such as this, the conclusion requires no testing to be validated, it simply presents itself as the only possible answer to the syllogism. In many cases it also matches

Non-matchers		
Figure	A–C conclusions	C–A conclusions
AB–BC	81.7%	18.3%
BA–CB	13.7%	86.3%
AB–CB	48.8%	51.2%
BA–BC	54.1%	45.9%
Matchers		
Figure	A–C conclusions	C–A conclusions
AB–BC	76.1%	23.9%
BA–CB	36.4%	63.6%
AB–CB	52.6%	47.4%
BA–BC	45.9%	48.2%

Figure 4.12: The figural effect by matching and non-matching subjects

Syllogism Type	Non-matchers	Matchers
No Valid Conclusion	19.7%	42.6%
One Model	5.4%	22.5%
Two Model	49.2%	71.7%
Three Model	48.1%	81.9%

Figure 4.13: Error rate for conclusions by matcher and non-matchers

one of the quantifiers in the syllogism. If all subjects use a tableau representation to agglomerate the premises of the syllogism, one would expect all subjects, no matter their strategy in other circumstances to be equally able to find the valid conclusion for one model syllogisms. Figure 4.13 shows that error rates for one model syllogisms are different depending on the performance of the subject on the diagnostic "matching" syllogisms. This leads to the conclusion that subjects prone to make matching responses do not have access to an agglomerated representation such as a mental model.

The implication of this, is that the figural effect is not the result of using a particular kind of representation, but that its mechanism lies in some feature of the syllogism commonly accessible to all types of strategy. Recall data will be used in the next chapter to further illustrate this point. With respect only to the recall data, asymmetries in conclusion choice can be found for the parallel figures that seem to be explicable on the basis of the quantifiers found in the syllogism.

4.4.3 Conclusion types and quantifier order

The distribution of types of conclusion is not uniform for the two figures not supposed to show a figural effect. For the figures AB–CB and BA–BC the frequency of conclusion types varies for different syllogisms. This is best demonstrated with the one-model syllogisms where either an A–C or C–A conclusion is valid. Figure 4.14 shows the frequencies for each type of syllogism considered.

It is evident from the frequencies that only one of the four syllogisms shown gives an equal number of A–C and C–A conclusions. An inspection of the quantifiers of the syllogisms shows that the A–C preference arises when the syllogism begins with the quantifier "all". This preference does not arise when the syllogism has another quantifier

Syllogism	Conclusion
All A are B. No C are B.	No A are C (17) No C are A (7)
No A are B. All C are B.	No A are C (9) No C are A (13)
Some B are A. All B are C.	Some A are C (12) Some C are A (12)
All B are A. Some B are C.	Some A are C (16) Some C are C (8)

Figure 4.14: Frequencies of conclusion type

in the first premise. A chi-square test finds that the type of conclusion is related to the type of syllogism in the case of the figure AB–CB (X=3.19 p<0.1) but not for figure BA–BC (X=1.14 n.s.). There are two plausible mechanisms by which such an asymmetry could arise. The first might be to suppose that A–C and C–A are equally likely to be chosen for these two figures, but that certain conditions favour the choice of A–C conclusions or suppress the choice of C–A conclusions, much in the same way as the figural effect operates. The second would be to assume a slight preference for A–C conclusions overall, as already shown, and suppose that a second preference is supplementing the first under certain conditions and conflicting with it in others. Thus for some syllogisms there will be a marked increase in the number of A–C conclusions and in others, the slight disadvantage of C–A conclusions will be overcome, producing nearly equal numbers of conclusion types. The second hypothesis better explains the data.

One of the conditions for the second preference seems to be the presence of the quantifier "all" in the first premise. It has already been shown in the recall data that "all" is the preferred quantifier when the first premise is being recalled. Subjects tend to make fewer recall errors when the quantifier in the first premise was "all" and they also tend to choose "all" when making an error in recalling another quantifier for the first premise. They will also tend to reverse the quantifiers in a syllogism that has "all" in the second premise so that it is recalled as having occurred in the first premise. Subjects

Response Type	"all" first	"all" second
A–C	27.6%	26.6%
C–A	24.9%	29.9%
"NVC"	11.8%	10.1%

Figure 4.15: Percentage of responses by quantifier order

appear to have, therefore, a preference for quantifiers to be in a certain order and it is hypothesised that this may be the result of a tendency to state positive information before negative. When searching for a conclusion to syllogisms that conform to this preference subjects might be inclined to choose A as a topic for the conclusion in line with the preference. In the case where "all" occurs in the second premise, some subjects might seek a conclusion that gives subject position to the noun in this premise, in this case the C term.

By dividing the syllogisms into those that have "all" as the first quantifier and those that have it as the second (but excluding the syllogisms where both quantifiers are "all", the effect of quantifier order on the syllogisms as a whole can be examined. In figure 4.15 the percentages of each type of conclusion are given for these two groups of syllogisms. A test of significance finds a just significant result ($F=2.7$ $p<0.1$) demonstrating that a syllogism with the quantifier "all" in the first premise is more likely to be given an A–C conclusion than C–A and that the opposite is true for syllogisms where "all" appears in the second premise.

4.4.4 The figural effect - a summary

The figural effect has been shown to operate and an overall preference for A–C conclusions found. The finding of a group of subjects who make "matching" responses, inexplicable in a purely deductive framework that also show a figural effect throws doubt on explanations for this effect that depend on the existence of an agglomerated representation from which simple deductive responses can be made. Rather than propose a two mechanism model for the figural effect, it is argued that the effect is the result of processing or strategies that occur independently of deductive reasoning. One such strategy has been found to be the order of the quantifiers in the syllogism, with the end-term associated with a universal quantifier tending to take subject position in the conclusion of the parallel[5] figures.

An alternative explanation for the figural effect, one that does not depend on an agglomerative representation is therefore needed. One such, suggested by Gilhooley and Wetherick and mentioned in Chapter 1 is that the relevant subject term is taken as the subject of the conclusion. Thus for the figures AB–BC and BA–CB, A and C will be chosen as the subject terms respectively. The other two figures cannot be explained so simply, because in these cases no one suitable term possesses subject position. Noticing that A–C conclusion are preferred overall suggests that subject might simply choose the first mentioned end-term as the subject for the conclusion where grammatical roles are not unambiguous. In the case of BA–BC where neither candidate is in the subject, this strategy should and does appear to influence the overall incidence of A–C conclusions. The factor relating to quantifier order also plays a part, and influences the choice towards C–A conclusions under certain circumstances. The figure AB–CB has two subject terms that could be chosen, one would expect and finds that almost equal numbers of each conclusion appear, with the quantifier order biassing the choice for particular syllogisms.

Thus an alternative explanation for the figural effect can be found, which, while not unitary, can account for any subjects displaying the effect, whatever their logical competence.

4.5 Strategies for Syllogisms Without a Conclusion

As was noted above, there are discrepancies between the response success for syllogisms that require two or three models to be solved, that have or do not have a valid conclusion. Syllogisms of this type that have no valid conclusion are answered with greater success than syllogisms that do. This effect is not explicable using the mental models paradigm and it seems clear that subjects have access to another strategy that allows them to make a decision about the presence or absence of a conclusion without needing to cycle through candidate conclusions and run the risk of erroneously accepting a non-valid conclusion because of a failure of memory capacity. Stenning(1990) has proposed that

[5]Figures where the middle term has the same grammatical role in both premises can be described as "parallel figures (BABC and ABCB). The other two figures can be termed "diagonal" figures

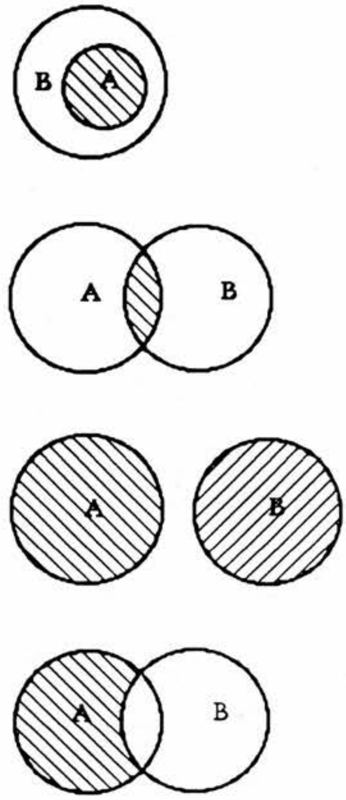


Figure 4.16: Stenning's version of Euler's circles

subjects, once they have begun to use an agglomerative representation, notice that almost all syllogisms with a valid conclusion have something in common. This can best be illustrated by the use of his adaptation of Euler's circles, using a hatching notation. Figure 4.16 shows that hatching the areas where an individual is definitely asserted to exist can in some cases lead to one area being double hatched. This indicates, in some cases that a valid conclusion to the syllogism exists. In only two syllogisms is there a valid conclusion where the corresponding diagram has no double hatching (figure 4.17). There are some syllogisms that have double hatching but no valid conclusion and a greater number that have no double hatching and no valid conclusion. Double hatching is closely related to the phenomenen of cancellation, which allows the B term to be removed from the syllogism because it is either positive or negative in both premises. Some subjects therefore could have access to this strategy without needing to form an agglomerated representation of the syllogism. To investigate the existence of such a strategy, data from both the experiments will be considered.

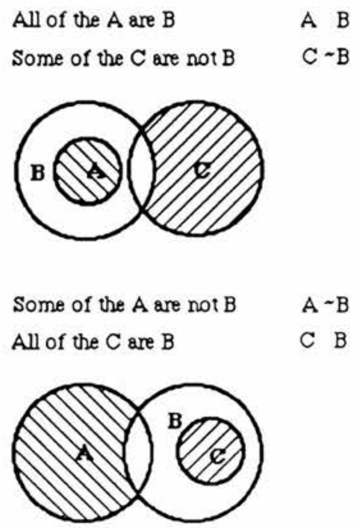


Figure 4.17: Syllogisms with a valid conclusion but no double hatching

4.5.1 The individuals experiment

It has been shown in an earlier section that for this experiment syllogisms without a valid individual are not answered more accurately overall than those that do have a valid individual, in contrast to the findings for the standard task. The next question to be considered is whether the presence or absence of double hatching or cancellation will affect the subjects ability to correctly answer the syllogism. Syllogisms that did not cancel were answered with greater accuracy than those that did cancel (see figure 4.18) A break down of accuracy of response by whether the syllogism cancelled and whether a valid individual existed or not is shown in figure 4.19. Here it can be seen that the presence or absence of cancellation strongly affects the ability of the subjects to find that no valid individual exists. It is also the case that syllogisms that do not cancel but have a valid individual are answered with much less accuracy than those that do cancel and have a valid individual. If a syllogism does cancel then finding a valid individual or concluding that no individual exists is done with nearly equal accuracy, so the strategy for solving syllogisms with cancellation does not favour one type of response. The differences between these groups are all significant (F(cancellation)=8.26 $p<0.004$, F(individual)=25.4 $p<0.00001$, F(interaction)=28.74 $p<0.00001$). Clearly cancellation is acting as a cue to subjects about the presence or absence of an individual, lack of cancellation facilitates the response "no valid individual" and makes it more difficult to

Cancels	Does Not Cancel
18.8%	21.8%

Figure 4.18: Percentage of incorrect responses by cancellation

Individual	Cancels	Does Not Cancel
No Individual	22.6%	13.6%
Individual	21.4%	52.6%

Figure 4.19: Percentage of incorrect responses by cancellation and presence of a valid individual

find a valid individual where one exists.

4.5.2 The standard task

Addressing the same issues for the standard task is taken in two parts. First one must ask the question whether the two syllogisms that do not cancel but have a valid conclusion are answered with greater or lesser accuracy than those that do cancel. The question must also be asked as to whether syllogisms with no valid conclusion are answered more or less readily when they cancel or are double hatched than those that are neither. The first question is addressed by figure 4.20. Here AO and OA[6] syllogisms of figure AB–CB (which are not double hatched but which have a valid conclusion) are compared with AO and OA figure BA–BC (which are double hatched and have a valid conclusion). Both types of syllogism require two models to be solved correctly.

The difference between the numbers of correct answers for these two types of syllogism is significant at the $p<0.1$ level ($t=1.45$) and the number of "no valid conclusion" responses given significant at $p<0.004$ ($t=2.96$). Therefore the presence or absence of double hatching appears to have a small effect on the ability of the subjects to find the correct valid conclusion and a much larger effect on their tendency to report (incorrectly)

[6]The four quantifiers can be referred to by means of four letters, derived from the latin words affirmo and nego. Using the first two vowels in each word, "all" becomes "A", "some" becomes "I", "none" becomes "E" and "some..not" becomes "O". At points throughout the text this convention will be used to refer to quantifier pairs thus "all, some..not" becomes "AO" and "some..not, all" becomes "OA".

	Not Double-Hatched Valid Conclusion	Double-Hatched Valid Conclusion
Percentage correct responses	36%	48%
Percentage NVC responses	42%	16%

Figure 4.20: Frequencies of response types for selected syllogisms

Not Double-Hatched	Double-Hatched
32.4%	28.0%

Figure 4.21: Percentage incorrect responses by double hatching for syllogisms with no valid conclusion

that there is no valid conclusion.

For the syllogisms that have no valid conclusion, cancellation and double hatching are not equivalent as they are for the AO and OA syllogisms studied above. It is not relevant to study double hatching for the individuals task as the presence of this always denotes the existence of an individual. However, for the standard task, double hatching may or may not imply the existence of a valid conclusion and it must be studied separately from cancellation therefore. Figure 4.21 shows that syllogisms with no double hatching and no valid conclusion are more prone to error than those with double hatching. This is significant ($t=1.67$ $p<0.09$). Therefore for the standard task lack of double hatching is not used as a cue that no valid conclusion exists, but instead makes answering these syllogisms more difficult. Examining only those syllogisms that cancel exaggerates the effect: non-cancelling syllogisms are even more prone to error with respect to cancelling syllogisms than is the case with non-double hatched versus double hatched ($t=2.1$ $p<0.03$).

Thus although the two tasks are similar in finding that syllogisms with no double hatching or cancellation are more difficult to solve correctly if they possess a valid conclusion, they differ in responses to syllogisms without a valid conclusion. The use of cancellation to give a quick solution to syllogisms with no valid conclusion seems to be specific to the individuals task. In the standard task, this feature and that of having no double hatching makes the syllogism harder to solve. That these syllogisms require more processing is backed up by analysis of the reading time data. Figure 4.22 shows that there are longer reading times for the second premise of cases that either have no double-hatching or that do not cancel. The overall difficulty in processing the group of syllogisms that do not have cancellation or double-hatching explains why they tend not to be used to discriminate response types and also suggests something about the way the syllogisms are represented as this factor seems to play quite a large part in determining the ease with which the syllogism can be understood.

This makes sense when viewed in terms of the requirements of the task. When isolating an individual, the subject has to use all three terms of the syllogism. A non-cancelling B term will immediately alert her/him to the fact that no consistent individual can exist. The standard task requires a search for a relation between the A and C terms. For this to be achieved B must be removed from the representation and presumably this will not be facilitated if this noun is negative in one premise and positive in the other. Thus the syllogism that does not cancel is harder to process and will be prone to more errors.

The greater accuracy for syllogisms with no valid conclusion responses for the standard task must still be explained however. The simplest hypothesis for the effect must be that subjects notice, after a little experience that certain pairs of quantifiers will never yield a valid conclusion. They might formulate some rules of logic, resembling Aristotle's maxims, perhaps. These might include the observation that syllogisms where both quantifiers are the same (and not "all") will never have a valid conclusion and that pairs of quantifiers that are both negative or both particular also never have a valid conclusion. Once these observations have been made, the subject will be able to conclude, without requiring a representation, that no valid conclusion exists.

While double-hatching and cancellation have not provided a solution to the "no valid conclusion" question, the results have indicated that syllogisms that do not have double-hatching or do not cancel are more difficult to solve (hence the observed disparity

	Cancellation	Double-Hatching
Present	2213.4	2394.6
Not Present	2621.3	2537.0

Figure 4.22: Reading times of the second premise by type of syllogism

between the AO and OA syllogisms observed above). That these syllogisms might be more difficult to process is supported by analysis of the reading times for the second premise, a rough indicator of the time taken to process the syllogism as a whole.

4.6 Summary

It has been found that both in the standard and the individuals task, the number of models needed to solve the syllogism correctly has an effect, if the syllogism possesses a valid conclusion. The fact that no significant differences exist between syllogisms with two and three models for either task (and in fact is in the direction opposite to that predicted for the individuals task) suggests the relationship between numbers of models and processing capacity is not strictly linear as has been suggested. These results point much more to a mechanism that can agglomerate premises without a great deal of difficulty and read off a conclusion, but that is immediately compromised by any amount of manipulation and that greater amounts of manipulation do not significantly add to the burden on memory. It may also be the case that the division into two and three model syllogisms is artificial and that subjects use ways to solve the syllogism that only ever demand one cycle of testing.

For both experiments it was also found that syllogisms that do not possess a valid conclusion, or in the case of Experiment Two a valid individual, are more easily solved than their counterparts with valid conclusions or individuals. In the case of Experiment Two this is found to be dependent on whether the syllogism in question cancelled or not, suggesting that when performing this task subjects will use cancellation to guide their initial decision about whether an individual exists or not. The situation with the standard task is different. Here there is also a strong advantage for syllogisms that have no valid conclusion, which suggests that subject have found some way of solving these syllogisms that does not require the use of cycles of testing. However, the presence or absence of cancellation does not affect accuracy of response in the way it does for the individuals task, in fact the opposite effect is found: non-cancelling syllogisms are harder to solve. The same is true for double-hatched syllogisms that do possess an individual but not a valid conclusion. Rather than be confused by this, subjects find these syllogisms easier to respond to. It is proposed therefore that the two tasks invite subjects to use different strategies. In the case where all three nouns are to be linked in the syllogism, the subject can see from a non-cancelling syllogism that no relationship exists, in other words, they learn to assume that B must cancel for a relationship between the nouns to be possible. The two exceptions to this amongst the AO and OA syllogisms are rather difficult to solve therefore and the conclusion missed because some of the syllogisms do not cancel. When looking for conclusions however, the same factors operate but this time subjects seem not to spot the shortcut cancellation offers. Non-cancelling syllogisms seem to be more difficult to solve, possibly because if viewed from the point of view of someone attempting to link A and C they constitute an indeterminate problem. The costs of attempting to interpret such a problem are high and it is therefore not surprising that subjects are often led into error. However, to explain their good performance on other syllogisms with no valid conclusion, one must assume that subjects are able to use their experience with syllogisms to derive logical principles which can be used to determine the presence or absence of a conclusion and thereby obviate the need for cycles of testing.

The figural effect is also found for both experiments and it is also found for a group of subjects who appear, according to the criteria laid down by Gilhooley and Wetherick to be drawing conclusions to the syllogism by a process of "matching". These subjects have a high error rate, as might be expected, but this extends to one model syllogisms as well as other types. If these subjects are not able to draw conclusions that can be derived from the representation without testing, one must raise the possibility that they are not using an agglomerated representation at all. If this is the case, then the fact that they display a figural effect brings into question whether this effect is in fact the result of forming the representation as Johnson-Laird claims, as this is exactly what this group of subjects appears not to be able to do.

It was also found that incorrect conclusions tend to be drawn in line with the atmosphere effect, leading to the possibility that Politzer's claim that subjects will use atmosphere to draw conclusions when their capacity to solve the syllogism by other means has been exceeded. However, the proportion of atmospheric conclusions does not increase with increasing difficulty of the syllogism, but the number of NVC responses does, suggesting that the latter will be the subject's answer when the process of validating a conclusion fails and that atmospheric (and incorrect) responses are given by a subgroup of subjects who do not use an agglomerative approach at all. This explanation would make most sense, because it avoids the need to argue that subjects have an agglomerated representation, but resort to one of the form Politzer suggests when the syllogism is too difficult. It is easier to envisage that some subjects do not grasp the logic of the syllogisms immediately and use other means to draw conclusions until they learn to agglomerate the premises. It may be that the group who produce atmospheric responses are the same as the group who match, but this has not been empirically tested.

Chapter 5

Testing Hypotheses

An important use of the data gathered in these experiments is to examine various theories about syllogisms in the light of this new information. With respect to the recall data, certain theories make very specific claims about the operation of memory in syllogistic reasoning. As already discussed, Johnson-Laird uses the idea of working memory to explain two effects: the differences in difficulty between syllogisms and the figural effect. These are two obvious first candidates for investigation. Other processes that might be expected to leave traces in the recall profiles are the misinterpretation of quantifiers, including the effect of Gricean implicatures and conversions.

5.1 General Recall Effects

There are many well documented effects relating to the interaction between memory and language. Primary amongst these is the depth of processing effect, first documented by Craik and Lockhart (1972) and which forms the theorectical basis for theories of language processing such as Kintsch and van Dijk's work. The difficulty of a task subjects are required to perform after reading a piece of text has an inverse relation to errors - the harder the task the fewer the errors[1], but this is affected by the relevance of the task and a ceiling effect is often observed, even a worsening of memory with very demanding tasks. Depth of processing has a clear relation to questions such as the role of non-logical processing of the syllogisms and also to the number of models needed to solve the syllogism. Difficulty also encompasses the latter however and if differences in error rates for memory are observed between groups divided on the basis of models, it will be difficult to assign the changes unequivocally to one or the other.

[1] see Einstein and McDaniel(1990) for a review.

	Immediate Recall	Delayed Recall
First Premise	14.9%	41.6%
Second Premise	20.6%	51.2%

Figure 5.1: Percentage error by first and second premise

One would in any event expect syllogisms with two or three models to have increased depth of processing when answered correctly and to constitute a more difficult task and therefore be better recalled than one model syllogisms.

One way to measure depth of processing is by the time taken to read the premises. The first premise is read for a much shorter time than the second but the error rates between the two premises do not reflect the division of labour. In figure 5.1 it can be seen that the second premise has many more instances of an error occurring than the first premise and that this effect is accentuated by delay. The extra time spent reading the second premise does not improve its chances of being recalled correctly therefore. However the results found do suggest that primacy rather than recency is a main factor in improving recall. This would be expected at delay, it is interesting to find that drawing a conclusion has such a strong effect on recency at immediate recall, as it could be argued that the process of reasoning constitutes a form of rehearsal.

The effect of models on recall will be studied solely with regard to correctly answered syllogisms, because it would be futile to expect depth of processing or difficulty to be represented by model number when subjects might not have treated the syllogism in question as a reasoning problem at all. Figure 5.2 also excludes syllogisms with no valid conclusion as the differences in correct response rates for these syllogisms found in chapter 4 suggests that their processing may not relate to model in any simple way. At immediate recall number of models and recall are not linearly related, although three model syllogisms have more errors than one or two model syllogisms, but at delayed recall the number of models seems to increase the syllogism in question's propensity to error. This is exactly the opposite effect to that predicted: that an increase in the number of models required to correctly solve the syllogism increases both depth of processing and difficulty of task, thus leading one to suppose improved recall with increasing number of models. One explanation might be that cycles of testing improve only information directly relevant to the task in hand, that is using rational means to solve the syllogism. Information such as the subject and predicate status of the nouns in the syllogism might be lost as the syllogism becomes more integrated and the material less relevant. The logical properties of the quantifiers however, would be much less readily lost as this determines the logical status of the syllogism and one would expect their order to be retained except in circumstances where it is irrelevant to logic. The role each noun plays in the syllogism is also important to the logic.

Syllogisms of the form AE or EA have, in four out of eight cases one model solutions and in the other four cases three model solutions. Studying these syllogisms separately, by number of models will show whether three model cases, correctly answered retain more logical information than one model examples. Figure 5.3 demonstrates that the information regarding the identity of the quantifiers is no better retained for the three model group and in fact at delayed recall is more readily forgotten. Noun assignment is better remembered however, again especially at delayed recall and recall of the order of the nouns in each premise is unaffected by numbers of models. This can only suggest that logical information in the form of the quantifiers is not better consolidated by cycles of testing, but it is interesting to note that noun assignment is better retained. It seems then that forming more than one model is detrimental to most types of information in the syllogism, whether important to logic or not. This may be due to the worsening of memory noted when the task following a text becomes too difficult. One additional factor affecting the quantifier recall is the identity of the conclusion itself. In all one model syllogisms and all but one two model syllogisms (where valid conclusions exist) the conclusion contains a quantifier also found in the original syllogism. In all three model syllogisms, the conclusion quantifier is different from both the quantifiers in the original. It is possible that arriving at a situation where this new relation is represented and presented as the correct answer causes confusion as to the identity of the original quantifiers when the representation is used for the purposes of recall. If this is so, and it has not been tested, then the observation by Lea et al (1990) that inferences made from a text are not confused in memory with the text itself, does not apply to these syllogisms.

The results shown here give very little reason to suppose that previously observed

Number of Models	Immediate Recall	Delayed Recall
One Model	28.2%	67.1%
Two Models	24.0%	72.0%
Three Models	32.7%	82.2%

Figure 5.2: Percentage of recall errors by models for correctly answered syllogisms

Type of Error	One Model	Three Models
Quantifiers Immediate Recall	14.9%	15.2%
Quantifiers Delayed Recall	39.1%	54.4%
Noun Assignment Immediate Recall	4.6%	4.3%
Noun Assignment Delayed Recall	11.5%	8.7%

Figure 5.3: Percentage of recall errors by models for correctly answered syllogisms of AE and EA type

effect such as depth of processing or difficulty of task have any relevance to syllogisms. Studying answers by type and including incorrect responses does show that answers of the form "no valid conclusion" are more often followed by error-free recall than syllogisms given a conclusion (18.5% errors versus 37.8% at immediate recall). This only applies at immediate recall and is to be expected, given the subjects have the opportunity when giving this response to go straight to recall without passing through a menu. However at delayed recall, the advantage is all but lost (61.2% versus 65.4%). It may be that here is to be found a depth of processing effect. The shorter reading time of the second premise for syllogisms given "no valid conclusion" responses has already been noted and the superior performance on syllogisms without a valid conclusion attributed to responses being given without the need for cycles of testing. The reduced processing presumably devoted to these syllogisms may well be the reason for their performance at delay, where they do not have the same advantage as at immediate recall.

5.2 The Figural Effect

A valid conclusion for a syllogism must relate the "end terms" of the syllogism[2]. Once a quantifier has been chosen for the conclusion, these two nouns must be used to complete it. The nouns can be arranged in the order A–C or C–A and in some cases, syllogisms with a valid conclusion can have either order and the conclusion will still be correct. In a number of cases, the conclusion must have a particular order of nouns to be right. The figural effect has already been described, to recap it consists of the observation that the figure AB–BC tends to be followed by conclusions of the form A–C and the figure BA–CB by conclusions of the form C–A

5.2.1 The Mental Models explanation of the figural effect.

To explain this effect, Johnson-Laird (1983)[3] appeals to the properties of working memory and the mental representation of the syllogism. When processing the figure AB–BC, subjects are able to form a model of the first premise, A–B and then immediately substitute C into that model using its relationship to B. Thus the noun A enters working memory before C. When the figure BA–CB is encountered, the information regarding C

[2]As has been the case so far, the end terms are referred to as A and C for the noun in the first and second premise respectively. The middle term in the syllogism is referred to as B.

[3]see also Johnson-Laird and Bara (1982).

cannot be immediately substituted into a model of B–A. If the premises of the syllogism are interpreted in reverse order as CB–BA then substitution is possible. In this case C enters working memory before A. Assuming that working memory operates on a "first in-first out" principle then when constructing the conclusion for an AB–BC syllogism A–C conclusions will be preferred and for BA–CB syllogisms C–A conclusions. For the other two figures AB–CB and BA–BC, Johnson-Laird hypothesises that reversals within a premise, either to create the figure AB–BC or BA–CB, are used and then the figures are processed in the order described above. In Figure 5.4 the processes required to represent each of the four figures are laid out.

There are two possible interpretations of Johnson-Laird's theory, however. It is clear that Johnson-Laird believes in the case of the figures BA–BC and AB–CB that a different kind of representation is constructed for each of the two types of conclusion. Which conclusion is chosen will depend on the way in which the subject manipulated the premises to obtain a transitive arrangement of the nouns in the representation. For figures AB–BC and BA–CB it is not obvious whether Johnson-Laird is claiming that a different representation underlies the non-preferred conclusion, or whether only one kind of representation is ever constructed for these figures and the non-preferred conclusion arises by chance, or as a result of especially careful checking by certain subjects. Both possible interpretations of the hypothesis need to be considered when using the recall data to assess its plausibility.

One interpretation will be that different conclusions require different type of representation for all the figures and this will be called version 1 of the mental models explanation. The other interpretation is that different representations underlie only the conclusions for the parallel figures and that the diagonal figures only ever give rise to one type of representation from which the preferred conclusion generally arises. This will be called version 2 of the mental models explanation. In Figure 5.4, the operations needed for version 1 of the theory are shown for the non-preferred conclusions for figures AB–BC and BA–CB; version 2 differs in that these operations would not occur.

5.2.2 Predictions for recall

A number of general predictions can be considered regarding Johnson-Laird's "figural theory", concerning the relative numbers and types of errors for certain figures. As

Figure	AB–BC		BA–CB	
Conclusion	A–C	C–A	A–C	C–A
First premise processed	A–B	C–B	A–B	C–B
Second premise processed	B–C	B–A	B–C	B–A
first premise	—	reverse	reverse	—
second premise	—	reverse	reverse	—
both premises	—	reverse	—	reverse

Figure	AB–CB		BA–BC	
Conclusion	A–C	C–A	A–C	C–A
First premise processed	A–B	C–B	A–B	C–B
Second premise processed	B–C	B–A	B–C	B–A
first premise	—	reverse	reverse	—
second premise	reverse	—	—	reverse
both premises	—	reverse	—	reverse

Figure 5.4: Mental models processing for the figural effect

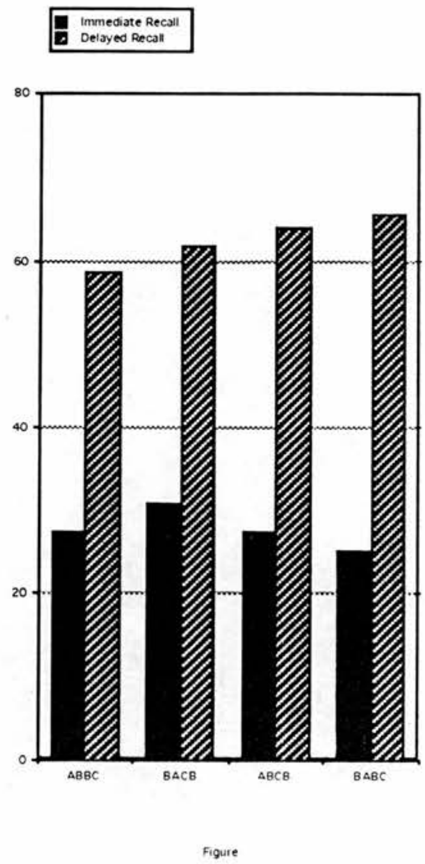


Figure 5.5: Percentage of noun errors by figure

well as evaluating these predictions, more detailed analysis of possible errors by conclusion types and figure will allow a comparison of the mental models approach with the grammatical roles theory of the figural effect, described below.

In general, if the figural effect is to have an impact on recall, one would expect this to be mainly on memory of the nouns of the syllogism as it is these around which the figural effect operates. Quantifier errors would not be expected to vary with the predictions of the figural effect. The following predictions can be made.

1. When the figure AB–BC is encountered of the mental models hypothesis, the subsequent recall of the nouns should contain fewer errors than recall following the other figures as the former requires no reordering before the premises can be combined. All the other figures by contrast need some kind of reordering before any

Figure	Percentage chosen
AB–BC	30.8%
BA–CB	28.5%
AB–CB	22.7%
BA–BC	18.4%

Figure 5.6: Percentage of each figure type chosen when an error in the figure was made at delayed recall

kind of conclusion can be drawn and this should be reflected in greater numbers of recall errors for the nouns of the syllogism. Figure 5.5 shows this to be the case for delayed recall, but not immediate recall.

In addition, at delayed recall also, the figure AB–BC is found to be more likely to be the figure chosen if one of the other figures is remembered incorrectly. This suggests that a transitive arrangement of the nouns recalled after a delay is preferred, if the actual figure has been forgotten (see Figure 5.6).

2. The processing times for the figures should vary as a function of the transformations needed to create a representation. The only measure of processing time available is represented by the reading time for the second premise and although this may not correspond exactly to the time taken to solve the syllogism it seems reasonable to suppose that the transformations needed to create the transitive representation will take place before the second premise is removed from the screen. The mental models account of the figural effect would predict that the figure AB–BC would have the shortest reading time for the second premise, the two figures AB–CB and BA–BC would have similar reading times as they require similar amounts of processing and that the figure BA–CB will have a longer processing time than AB–BC. Figure 5.7 shows that at least one of the predicted differences does occur, as BA–CB has a longer processing time than AB–BC. However, the shortest time is found for BA–BC, which cannot be explained by the mental models approach. A test of the differences in reading time for figure (F=2.96, n.s.) leads to the conclusion that there is no significant difference in processing time

Figure	Reading time /ms
AB–BC	2445
BA–CB	2615
AB–CB	2477
BA–BC	2260

Figure 5.7: Processing time by figure

	AB–BC	BA–CB
Immediate recall	2.6%	4.9%
Delayed recall	9.7%	6.9%

Figure 5.8: Percentage premise reversals

for the four figures.

3. A high number of cases with the nouns within a premise recalled in reverse order might be expected with figures AB–CB and BA–BC, which are supposed to have one premise reordered to form AB–BC or BA–CB. However this type of error is found no more frequently for these figures, than it is for the other two figures, which do not require this type of reordering according to Johnson-Laird (see Figure 5.9).

4. It is the central point of this theory that the premises for the BA–CB are interpreted in reverse order as CB–BA and those for AB–BC in order of presentation. If this is the case, then many more instances of recall profiles where the premises have been reversed would be expected for the figure BA–CB than AB–BC. In figure 5.8 the percentage of recall profiles where a premise reversal occurred are given for A–C conclusions following AB–BC and C–A conclusions following BA–CB. There are slightly more premise reversals for BA–CB at immediate recall, but at delayed recall AB–BC has more. Either processing the premises in reverse order does not play a part in the figural effect or the results of this processing are not reflected in memory.

Figure 5.8 also shows that the proportion of this type of error is very low. In addition, although the percentage of reversed premises is higher for BA–CB at immediate recall, at delayed recall it is higher for AB–BC. If premise reversal is a feature of processing these figures then at both immediate and delayed recall, BA–CB would have many more instances of this type of error than AB–BC. Taken as a whole the recall data shows that there is really no difference between the two figures for this type of error.

The predictions from Johnson-Laird's theory are not supported by the data here and in some cases are contradicted by it. Either one must assume that memory has no relationship to the processing of the syllogism, even though the features of memory play a crucial part in this explanation for this effect, or this explanation must be rejected and another found.

5.2.3 Other explanations for the figural effect.

As already argued, some researchers, such as Gilhooley and Wetherick (1989) propose that the figural effect is simply the result of the subject maintaining the grammatical roles of the noun in the syllogism when constructing a conclusion. This is taken to be a consequence of the subject viewing the syllogism as an argument about something. Many studies in linguistics have noted that language comprehension is characterised by the recipient of language attempting to identify the object or person that a communication is about. Kieras (1980,1981) has found that there are a number of factors affecting the choice of focus. The subject of the discourse is one such, but the first-mentioned item and the most frequently mentioned item are also taken to be topic. When A is the only candidate for the conclusion in subject position, the syllogism solver will take this as the subject of the conclusion. When C is the only candidate, it will be chosen.

The parallel figures have no unequivocal choice for the conclusion, thus other factors will come into play. In chapter 4, some of the asymmetries resulting from this are discussed. However, some indirect support is found for the view that it is the grammatical role of the nouns in the syllogism that influences the form of the conclusion (see Figure 5.10) by considering the parallel figures and their recall profiles alone. For each type of conclusion the number of times A or C appeared in subject term in the recall (in either premise) is given as a percentage of the number of times that conclusion occurred.

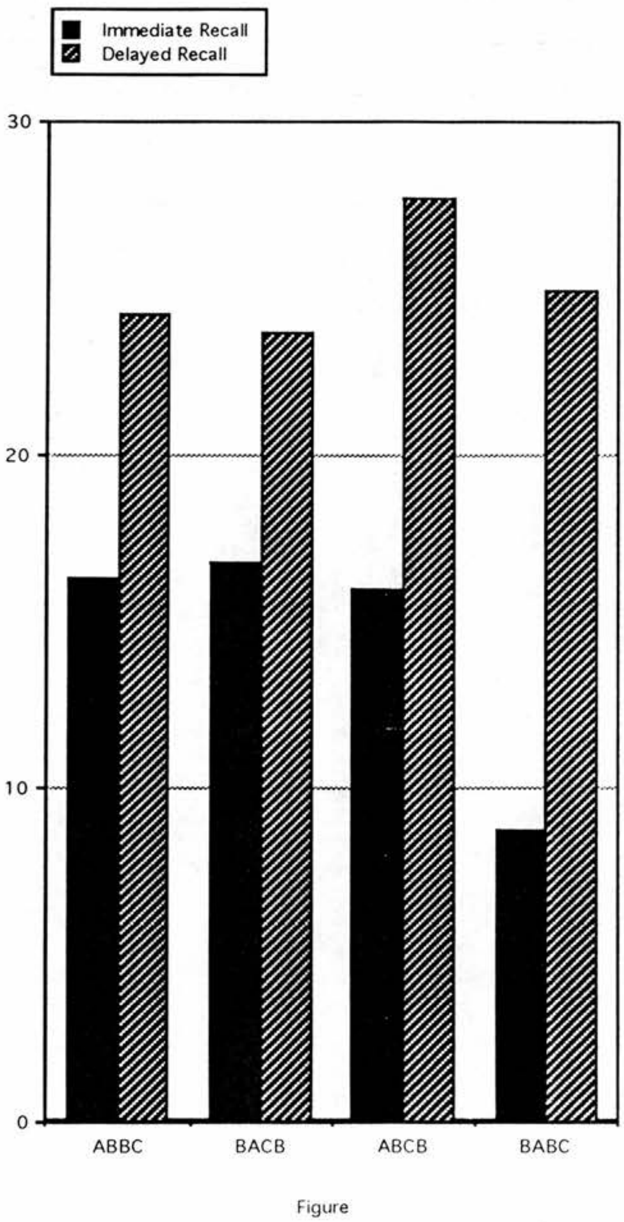


Figure 5.9: Percentage of recalls where the noun order in one or both premises was reversed by figure

to remove bias due to differences in numbers of conclusions.

It is found that at immediate recall the conclusion A–C is more likely to be followed by recall profiles where A is the subject than where C is the subject of the recalled premises. Similarly, the conclusion C–A is more likely to be followed by recall profiles where C is the subject than A. There is a significant interaction between the conclusion type and the noun recalled at immediate recall ($F = 9.7$ $p<0.01$) and there is a significant effect of noun type ($F = 10.6$ $p<0.01$), so A is recalled in subject position more frequently than C. At delayed recall, there are no significant differences. This data might indicate two things: that the conclusion interferes with the recall of the syllogism so that some nouns are recalled with the grammatical role they had in the conclusion rather than that in the original syllogism, or that when unable to recall the role of the noun in the syllogism, the subject uses the grammatical role of the nouns in the conclusion as a guide or substitute. The actual mechanism for this effect is probably a combination of the two. Both proposals show that the grammatical roles of the nouns in the conclusion exert an effect on the form of the recall. In fact it could be argued that the subjects expect the nouns of the syllogism to have the same grammatical role as those of the conclusion, as this is the main basis for choosing the topic for the conclusion and are therefore influenced by this to change the recalled syllogism to resemble the conclusion. Given that memory for order of the nouns, that is grammatical role, is lost relatively quickly, the explanation that subjects will use the conclusion to reconstruct the syllogism seems the most likely one. This only serves to emphasize that the subjects' expectations about the syllogism depend upon the form of the conclusion, which in turn strongly suggests that the form of the conclusion is derived from the same features of the syllogism.

Thus it can be argued that if the grammatical roles of the nouns in the conclusions affect the recall of the syllogism, then grammatical role has some salience for the subject at least regarding memory. If this is so, then it is possible to extend the argument to the choice of the subject and the predicate of the conclusion. If grammatical role is important for recall, in that subjects make errors that reflect the grammatical role of the conclusion, then it may also be important for the construction of the conclusion from the original syllogism. As Johnson-Laird's explanation of the figural effect is unsupported by this data, then more weight can be given to any evidence, even if indirect, that supports the grammatical roles theory, the only other good explanation for this effect.

	A–C	C–A	A–C	C–A
A subject	24.39%	22.86%	20.54%	20.28%
C subject	14.37%	22.96%	20.08%	25.98%

Figure 5.10: Percentage recall by conclusion type

5.2.4 Comparing the mental models and grammatical roles explanations

Some indirect evidence has been gathered to support the grammatical roles explanation of the figural effect, the mental models explanation has not been supported by other data. However, these findings are not conclusive for a number of reasons.

1. The processes described by Johnson-Laird might not cause errors that directly correspond to the operations required. To reverse the order of the nouns in a premise when constructing a representation might not necessarily cause the premise to be recalled with reversed nouns. The quantifier in the premise affects the possibility of this occurring, as does the serial position effect.

2. The type of conclusion affects the type and complexity of the processing. Specific predictions for certain conclusions must be evaluated to exclude the possibility of trends being obscured in too general a picture.

3. The evidence for the grammatical roles theory is indirect and the possibility that these findings are a result of mental models is not excluded.

By examining trends by conclusion and figure, the predictions of version 1 of the mental models theory (different represenations are used to create different conclusions). version 2 of this theory (diagonal figures always have the same representation, whatever the conclusion given) and the grammatical roles theory can be compared to the obtained pattern of errors. The numbers of instances where an unspecified error occurred will be considered, to avoid the problem raised in 1 above, that the error types might not correspond to types of process.

Figure 5.11 gives the relative error rates, again this would be for recall of the nouns of the syllogism, predicted by the three theories. No two theories give exactly the same

Figure	mental models version 1	mental models version 2	grammatical roles
AB–BC A–C	low errors	low errors	low errors
C–A	low errors	very high errors	high errors
BA–CB A–C	medium errors	medium-high errors	high errors
C–A	medium errors	medium errors	low errors
AB–CB A–C	low-medium errors	low-medium errors	medium errors
C–A	high errors	high errors	medium errors
BA–BC A–C	low-medium errors	low-medium errors	medium errors
C–A	high errors	high errors	medium-high errors

Figure 5.11: Predictions for mental models and grammatical roles accounts

pattern of errors, so they can be distinguished. Figure 5.12 gives the obtained error rates, expressed as a percentage of the number of conclusions of this type found for each figure. Thus the effect of numbers of conclusions is eliminated.

At immediate recall, the pattern of errors most resembles the predictions of the grammatical roles theory, with version 2 of the mental models theory being ruled out by the high numbers of errors for the non-preferred conclusions for figures AB–BC and BA–CB. Version 1 of the mental models theory would predict a difference in errors for the other two figures, with conclusion C–A creating more errors than A–C. The grammatical roles theory would not predict a difference for conclusion types here and this is the result found.

5.2.5 Other strategies affecting the choice of conclusion.

The analysis of the recall data in chapter 3 suggests that there are non-logical principles that affect the subjects choice of quantifiers and how accurately they can recall the different types of quantifier. It was argued in chapter 4, on the basis of the conclusions drawn by the subjects, that there may be other strategies used when seeking a topic or focus for the argument in the syllogism. One such strategy seemed to be to choose the end term in the first premise. This preference might be derived from a general principle, established by Kieras and mentioned above that topics of arguments tend to be introduced early on in discourse. Another factor that appeared to affect the type

Immediate recall		
	A−C	C−A
AB−BC	23.3%	61.0%
BA−CB	54.3%	31.2%
AB−CB	45.9%	40.0%
BA−BC	33.1%	35.4%

Delayed recall		
	A−C	C−A
AB−BC	53.5%	68.3%
BA−CB	73.9%	56.8%
AB−CB	61.0%	67.5%
BA−BC	64.6%	76.3%

Figure 5.12: Percentage recall errors for the nouns of the syllogism by conclusion type and figure

of conclusion was the order of the quantifiers in the syllogism. There is a preference, revealed in the recall data, to argue from the positive to negative, that is to introduce the argument with the quantifier "all" or "some" and qualify the situation described with the quantifiers "some..not". This expresses itself when the conclusion is drawn as a preference for nouns associated with the positive quantifiers to be the subject of the conclusion. This preference might cause error in the recall of other parts of the syllogism, such as the order of the nouns. For example, when a syllogism which has a positive quantifier in the second premise and some other quantifier in the first premise is recalled, there is a strong tendency to recall the quantifiers in reverse order, so that the positive appears first. When this occurs, there might be a related tendency to recall the nouns from the second premise in the first. This might not simply result in reversed premises, but also in the appearance of the C term in the first premise.

The following predictions regarding recall of the syllogism can be tested.

1. If the A term is generally preferred for the subject of the conclusion then it should be preferred as the subject of the premises at recall. Figure 5.13 shows that A

	A subject	C subject
Immediate recall	33.12	30.72
Delayed recall	32.9	31.2

Figure 5.13: Mean choice of role for nouns

is the subject of one or other of the premises most frequently at immediate and delayed recall. This finding supports the hypothesis that in a general sense A takes priority as a subject for the syllogism at the stage when the syllogism is being solved.

There is a significant interaction between delay and the number of each noun recalled in subject position (Fint=13.5 $p<0.01$) but no effect of either delay or noun alone (Fdelay=0.18 n.s., Fnoun=0.2 n.s.). A t-test shows that the difference between the nouns for immediate recall is significant – that there are more instances of the A noun being recalled in subject position at immediate recall than C.

2. Syllogisms that conform to the preference for the form of the argument should show fewer errors in the recall of the nouns than syllogisms that do not. In particular syllogisms that do not conform should have a higher number of instances where C is recalled as having occurred in the first premise than syllogisms that conform. This would be because in the cases where a positive quantifier occurs in the second premise (non-conforming syllogisms) there is a tendency to recall this quantifier as having occurred in the first premise. The C term will also tend to be recalled in the first premise therefore. This is supported by analysis of the recall errors for the figure and noun assignment of the syllogisms. When the syllogism conformed, there were fewer errors at both immediate and delayed recall for both variables then when the syllogism did not conform.

   Inferential analysis of these errors demonstrates that the differences shown in Figure 5.14 do not simply imply that conforming syllogisms are less likely to attract recall errors in general. The errors for figure are not significantly higher for non-conforming syllogisms (Fdelay=51.93 $p<0.01$, Fsyll=1.81 n.s., Fint=0.64 n.s.), but for noun assignment there is a significant change (Fdelay=61.63 $p<0.01$,

Immediate recall		
Syllogism type	Noun assg.	Figure
conforming	7.7%	22%
non-conforming	14%	28%
Delayed recall		
Syllogism type	Noun assg.	Figure
conforming	30.3%	50.3%
non-conforming	38.7%	52.3%

Figure 5.14: Mean errors for conforming and non-conforming syllogisms

Fsyll=4.73 $p<0.01$, Fint=0.21 n.s.). This suggests that the effect of having a nonpreferred order for the quantifiers in the syllogisms is to specifically disrupt the memory of the roles the nouns played in the syllogism. This is supported by the finding that there are significantly more instances of c being recalled in the first premise for syllogisms that do not conform (Fdelay=40.5 $p<0.01$, Fcterm=9.52 $p<0.01$, Fint=0.22 n.s.). Thus the form of the memory disruption is to lead subjects to recall the minor term as having occurred in the first premise, that is to give it the role of major or middle term.

5.3 Misinterpretation of the Quantifiers

Work was cited in the first chapter that investigates the role of an incorrect understanding of the quantifiers as leading to incorrect reasoning. One effect has already been found in the previous chapter that seems to be the result of a confusion between "some" and "some..not", but closer examination of subjects' understanding of what the quantifiers imply logically can be carried out by use of their errors when recalling the quantifiers themselves and the premises they appear in.

5.3.1 Conversions

A common mistake when recalling the nouns of the syllogism is to reverse their order. The tendency to reverse the order of nouns within a premise when it is not logically

	Immediate	Delayed
All	14.08%	31.49%
Some	22.63%	43.51%
None	16.71%	33.39%
Some..not	20.14%	38.28%

Figure 5.15: Percentage of type 2 errors for each quantifier

possible to do so has been used as an explanation for reasoning errors by Chapman and Chapman (1957) and Revlis (1975) who have called the phenomemon "conversion". Figure 5.15 demonstrates that the type of quantifier does have an effect on the number of instances where the nouns are recalled in reverse order[4].

A two-way Anova shows that delay has a significant effect on the number of conversions (Fdelay=71.94 p<0.01) as does the type of quantifier (Fquant=1.81 p<0.05) but that there is no interaction between the variables (Fint=0.32 n.s.). Although the type of quantifier makes a difference to the number of times a subject will recall the nouns of the premise in reverse order, clearly this does not follow exactly from the logic of the quantifiers. While "all" has significantly fewer conversions than the other quantifiers "some...not" does not and logically, "some..not" will not allow a conversion and logic to be preserved unlike "some" and "none" which do imply their converse. This may be because "some..not" is often misinterpreted as being equivalent to "some", so making this error will also allow the error of assuming that "some..not" logically implies its converse. The relationship between these recall results and the question of whether subjects do make errors in reasoning through illict conversions is hard to determine. What these results do indicate is that in cases where the subject perceives that the order of the nouns in the premise is relevant to meaning, the order will tend to be preserved. Thus in cases where the order of the nouns in the premise is unimportant to meaning it will be lost. This does not imply that the errors found in recalling the order of the nouns for the quantifier "all" are an indication that in these cases conversion at the reasoning level has occurred, rather that in general the logic of "all" is appreciated and that conversions are rather likely not to be a source of error, at least as far as this

[4] a type 2 error after chapter 3

Immediate recall		
	Logicians	Non-logicians
All	5.59%	23.26%
Some	14.79%	31.12%
None	11.98%	21.84%
Some..not	16.00%	24.63%

Delayed recall		
	Logicians	Non-logicians
All	23.58%	40.07%
Some	37.75%	49.76%
None	29.82%	37.26%
Some..not	32.67%	44.36%

Figure 5.16: Percentage conversions for each quantifier

quantifier is concerned. What is indicated and pursued further below, is that subjects do not appreciate the logic of the quantifier "some..not" but rather tend to assume it means the same as "some". The lower number of type 2 errors for "none" is also interesting, because the logic of this quantifier does not demand that the order of the nouns in its premise be retained. This taken with the lower type 2 errors for "all" suggests that universal quantifiers allow the grammatical structure of the premise they are in to be better retained. It has already been shown in chapter 3 that particular premises are associated with higher levels of a number of noun order errors and this may be another example of this phenomenon.

As the incidence of conversions appears to be related to the logic of the quantifiers, the extent to which each of these two groups of subjects understand the logic of the quantifiers can be measured by studying the numbers of conversions for each quantifier type for each group in Figure 5.16

Logicians make fewer of this type of error than non-logicians. The pattern of errors for logicians indicates that they grasp the full meaning of the quantifier "all" but not of "some..not", suggesting that they are prone to the misinterpretation of "some..not" as being equivalent to "some" and thus the order of the nouns following it being irrelevant to meaning. Non-logicians made most conversions following the quantifier "some", but did not make fewest for "all", thus implying that they do not have a full grasp of its meaning.

5.3.2 Other types of misinterpretation

It has been shown, in other studies[5] that some subjects do not grasp the full logical implications of the quantifiers used in syllogisms, but interpret them according to heuristics used in general language processing. Gricean implicatures, for example, state that people will not interpret "some" as meaning "some and possibly all" but as meaning "some but not all" and similarly "some..not" is interpreted as meaning "some..not but not none" rather than "some..not and possibly none". If this is true, similar findings may be revealed by the recall data, on the principle that if subjects make an error at the level of reasoning that error is likely to reveal itself at recall.

1. If subjects interpet "some" as meaning "some but not all" then recalling "some" as "all" should be rare.

2. If subjects interpret "some..not" as meaning "some..not but not none", then recalling "some..not" as "none" should also be rare.

3. If subjects misinterpret "some..not" as implying "some" then these two quantifiers should be commonly confused.

The direction of the results shown in Figure 5.17 confirm the predictions and a two-way Anova shows that both delay and the interaction between delay and type of confusions are significant ($F=210.7$ $p<0.01$, $F=4.7$ $p<0.01$). Type of confusion alone is also significant ($F=13.2$, $p<0.01$). The effect of Gricean implicatures have most effect in preventing "some..not" from being recalled as "none". The smaller differences between "some" being recalled as "all" and "all" being recalled as "some" are probably due to the preference for "all" in the first premise, which will increase the incidence of all quantifiers changing to "all". There is a slightly greater tendency to recall "some" when the actual quantifier was "some..not" than vice versa. Subjects often reported difficulty in understanding the quantifier "some..not" and this may account for the preference for

[5]see especially Newstead and Griggs(1983) and Newstead(1988).

	Immediate recall	Delayed recall
"all" — "some"	3.3%	15.8%
"some"— "all"	2.6%	14.8%
"none"— "some..not"	3.6%	19.5%
"some..not"— "none"	0.7%	7.2%
"some"— "some..not"	5.3%	21.7%
"some..not"— "some"	8.1%	25.3%

Figure 5.17: Percentage of confusions for quantifiers

"some". It is certainly the case that these quantifiers are much more likely to be confused with each other than with their universal counterparts. This is further evidence for the existence of Gricean implicatures in interpreting the quantifiers. This also suggests that the quantifiers "some" and "some..not" share a uniform representation, as they are so easily confused.

As in the previous section, the extent to which using a deductive reasoning strategy implies a full understanding of the quantifiers is investigated. The effect of Gricean implicatures might only be found for subjects treating the syllogism as a piece of communication and nothing more: subjects approaching the syllogisms as reasoning problems might be immune to textual effexts such as these. This is not the case. The same effects that have been found for the group as a whole are present in each of these subgroups showing that the type of confusion between quantifiers has a significant effect for both logicians ($F=4.6$, $p<0.01$) and non-logicians ($F=6.0$, $p<0.01$).

5.4 Recall and Features of the Representation

It was found in the previous chapter that there was evidence that cancellation and double-hatching played a part in reasoning strategies for the individuals task. the same findings were not found for the standard task, but there was evidence that cancellation and double-hatching were playing a part in the reasoning process, by making syllogisms that did not cancel or have double-hatching more difficult to solve. The recall data will now be used to investigate the proposition that this increased difficulty is related to

	Cancels	Doesn't Cancel
Error in Recall (Immediate)	27.1%	30.3%
Error in Recall (Delayed)	62.3%	63.4%
Reading Time(ms)	2343.8	2829.7

Figure 5.18: Recall errors and reading times for the second premise by the presence or absence of cancellation

	Double-Hatched	Not Double-Hatched
Error in Recall (Immediate)	27.5%	28.2%
Error in Recall (Delayed)	62.1%	63.3%
Reading Time(ms)	2394.6	2537.0

Figure 5.19: Recall errors and reading times for the second premise by the presence or absence of double-hatching

indeterminacy in the text. It was demonstrated by Stenning (1986) that descriptive texts take longer to read and are more poorly recalled if indeterminate than determinate ones. The reading time and memory data will be used to see if these two variables relate to the presence or absence of cancellation and double-hatching in such a way as to suggest that the absence of these factors constitutes indeterminacy.

Figures 5.18 and 5.19 give the overall percentage of texts that had any kind of error at immediate and delayed recall for texts with and without cancellation or double-hatching and the reading times for the second premise for the same groups. The differences between recall errors at immediate recall for cancelling and non-cancelling texts is significant ($t=2.05$, $p<0.04$) as is the difference in reading times ($t=3.92$, $p<0.0001$). The numbers of errors for delayed recall is not significantly different between texts, nor are any of the differences for the double-hatched versus non-double-hatched texts.

Cancelling texts therefore are found to have a lower error rate for immediate recall and less time is taken to read their second premise. As the second premise reading time constitutes some measure of the time it takes for the subject to feel confident of a representation, if not to have decided on a response, it seems safe to say that cancelling texts are easier to understand and possibly easier to represent. The presence or absence of double-hatching does not appear to make a difference.

It will be shown later that the type of conclusion drawn for the syllogism has an effect on the ability of the subject to recall the syllogism. Bearing this in mind, the analysis above is repeated, but only for syllogisms that do not possess a valid conclusion. These syllogisms can be divided into three groups: those that cancel and have double-hatching, those that cancel and have no double-hatching and those that do not cancel and have no double-hatching. Figure 5.20 gives an example for each category. There are no example of syllogisms that do not cancel and yet have double-hatching. There are significant differences between these three groups for all variables but delayed recall errors ($F(\text{imm rec})=3.44$, $p<0.03$; $F(\text{del rec})=1.05$, n.s; $F(\text{resp})=6.58$, $p<0.001$ $F(\text{rt})=2.61$, $p<0.07$) and the percentages given in figure 5.21 indicate that syllogisms that cancel and have no double-hatching are recalled with fewest errors, followed by those that cancel and do have double-hatching and those that neither cancel or have double-hatching having the poorest recall. This is true at both stages of recall (but only significant at immediate recall), also of the accuracy of the reasoning responses given and the reading time for the second premise. Testing between groups finds that only the difference in recall errors between syllogisms with no double-hatching but which cancel or do not cancel are significant ($t=-2.83$, $p<0.004$). The reading time for syllogisms that cancel and have no double-hatching is significantly shorter than the reading times for cancelling syllogisms with double-hatching and non-cancelling syllogisms without ($t=3.81$, $p<0.0002$ for doesn't cancel, no double-hatching versus cancels, no double hatching, $t=2.47$, $p<0.01$ for cancels, double-hatching versus cancels, no double-hatching). The accuracy of the responses is significantly lower for syllogisms which do not cancel and do not have double-hatching than the accuracy for the other two groups ($t=2.3$, $p<0.02$, $t=2.36$, $p<0.01$) but the presence or absence of double-hatching for the two groups which cancel does not make a difference in the subjects' ability to correctly solve the syllogism.

The main effect appears to be that syllogisms whose middle terms cancel are read more quickly, recalled more accurately and have more correct reasoning responses than those that do not cancel, when the analysis is confined to those syllogisms that do not have a valid conclusion. Rather than cancellation being used as a tool to detect syllogisms without a valid conclusion, its absence appears to make syllogisms harder to process and to solve, when a conclusion is searched for. It has already been suggested that the reason for this may lie in subject's attempts to relate the end-terms and being prevented from doing so by the fact that the middle term cannot be removed from the "equation" simply. Figure 5.22 shows why this might be so. Double-hatching does not seem to make syllogisms with no valid conclusion harder to answer, but its presence does increase the processing time for the second premise and the number of recall errors at immediate recall.

5.5 Summary

In this chapter the recall data has been used to address various issues relating to theories about syllogistic reasoning. It has been found that the mental models explanation of the figural effect is not supported by an analysis of the data, either for predictions generated by that model alone or in terms of a comparison of predicted levels of recall error for this explanation and the grammatical roles explanation. The latter hypothesis is supported though, from indirect evidence showing that the grammatical roles (or focus) of the conclusion is preserved in the recall, suggesting that the factors that lead to its being preserved are the same that led to its being chosen as the topic of the conclusion in the first place. The predictions of the grammatical roles theory are also borne out by the relative levels of noun recall error for various combinations of figure and conclusion type.

Recall errors are also found that reveal misinterpretations of the logic of the quantifiers. There appears to be an appreciation of the logic of "all" as the nouns in the premise containing this quantifier are recalled in reverse order rather less frequently than the other nouns, however, the fact that "none" also has reversed nouns less often in the premises it heads at recall tends to suggest along with findings made earlier that universally quantified premises are less prone to all types of error. The frequency with which certain quantifiers are changed to others reveals the operation of Gricean

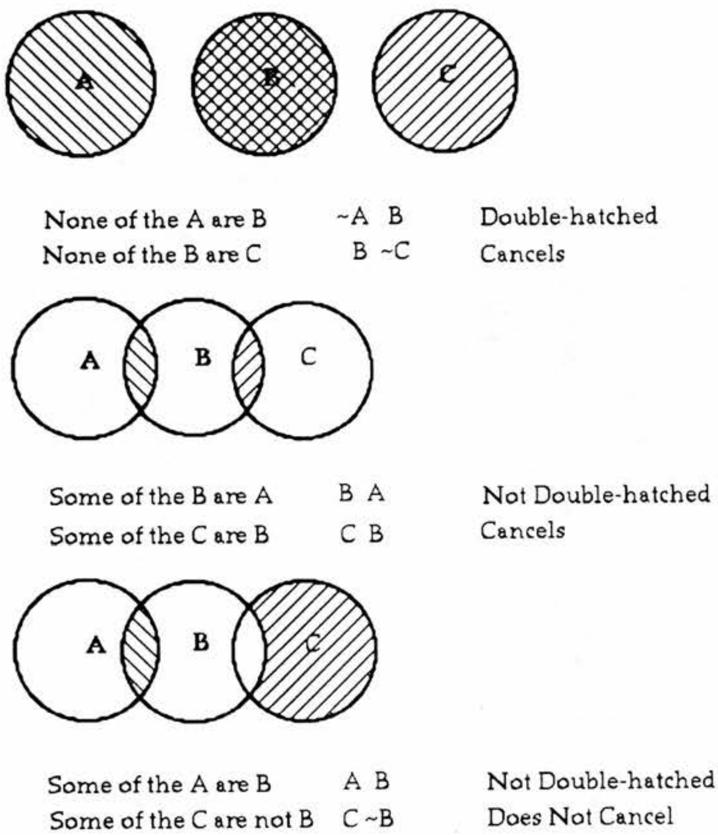


Figure 5.20: Examples of syllogisms with no valid conclusion with combinations of cancelling and double-hatching

	Cancels Not Double-Hatched	Cancels Double-Hatched	Does Not Cancel Not Double-Hatched
Error in Recall (Immediate)	28.8%	35.1%	40.0%
Error in Recall (Delayed)	78.6%	80.3%	81.0%
Incorrect Responses	27.7%	28.0%	36.7%
Reading Time(ms)	2163.1	2538.9	2797.0

Figure 5.21: Recall errors, Incorrect responses and reading times for the second premise by the presence or absence of double-hatching and cancellation for syllogisms with no valid conclusion

All the A are B

Some of the C are not B

becomes

A B

C ~B

from which the B term cannot be removed because it is positive in one premise and negative in the other

Figure 5.22: An example of a non-cancelling syllogism

implicatures. Changes from universal to particular are more common in general than vice versa, which is in line with the observation that subjects are less prone to observe the full logical meaning of particulars, which implies the universal statement also to be possibly true. The most common types of quantifier error however involve changes between the particular quantifiers, suggesting that the interpretation of these two may not be distinct in many cases.

Finally, recall data is used to examine the proposition that cancelling syllogisms are more determinate than non-cancelling ones. This is upheld, particularly with reference to the group of syllogisms that have no valid conclusion. Double-hatching, which does not seem to play much part in processing overall, is also found to influence recall and processing, but in such a way as to worsen or lengthen these factors: the opposite of cancellation. In the previous chapter it was found that the absence of double hatching made syllogisms with valid conclusions more difficult to answer and while the presence of this feature does not affect correct response rates, its impact on the recall suggests that double-hatching may be used by subjects as a cue to the presence of a conclusion.

Chapter 6

Implications for Syllogistic Reasoning

In the previous chapters basic recall effects have been demonstrated and this data along with the reasoning data from both experiments has been used to look critically at various theories pertaining to syllogistic reasoning and memory phenomena in general. It now seems appropriate to use the information gained from these experiments to formulate an impression of the priorities and strategies used by subjects when solving these problems. Of particular interest is information pertaining to the type of representation actually used by subjects when reasoning. Must this be based on individuals as Johnson-Laird claims, or is there evidence that set-based representations might also be used? Does the evidence suggest that this is a useful distinction to make, or are these theories only distinguishing between learned strategies, themselves arising from homogenous processes at another level?

6.1 Reasoning and Recall

It has been established that there exists a group of subjects who in the standard task, at least initially, do not appreciate the logic of syllogisms and who generate conclusions from them on the basis of judgements about the plausibility of certain relations existing on the basis of the situation described. There also exists a rather larger group of subjects who appear to be using rational means to solve the syllogism, of the general form: combine the premises and use cycles of testing to identify the presence of a valid conclusion. The relationship that these different processes bear to recall has already been examined in rather specific ways in the preceding chapter, but a broader look at this

Type of Conclusion	Immediate Recall	Delayed Recall
Correct NVC	21.2%	63.4%
Correct VC	25.2%	57.4%
Incorrect NVC	15.9%	59.9%
Incorrect VC (should be NVC)	44.0%	66.2%
Incorrect VC (should be VC)	44.1%	67.6%

Figure 6.1: Percentage error by types of response

connection is required. The responses to the syllogisms in Experiment 1 can be grouped into five categories: those that were correct and a valid conclusion drawn, those that were correct and no valid conclusion drawn, those that were incorrect because a conclusion was given where none exists, those that were incorrect because no conclusion was given and one does exist and those that were incorrect because the wrong conclusion was given although a valid conclusion does exist. It has already been recognised that answering "no valid conclusion" will, at immediate recall cause less disruption of memory than drawing a conclusion. However it has also been shown that this advantage disappears at delayed recall, presumably because of poorer consolidation of syllogisms where no valid conclusion is given. It is of interest to see whether this effect is consistent over all types of response, for example, giving "no valid conclusion" to a syllogism with a conclusion may arise from exitting cycles of testing prematurely, whereas in other cases giving this response may be accompanied by a superficial representation of the syllogism because logical principles have been used to solve it rather than an agglomerated representation. When conclusions are drawn distinguishing between those that are correct and those that are not is important because some will be the result of matching and therefore are thought to be the result of a rather different reasoning strategy.

Figure 6.1 shows that these distinctions are well motivated. There is a very much higher percentage of errors for syllogisms where the response was one of the incorrectly drawn conclusions at immediate recall and a rather less dramatic difference at delay.

Incorrectly stated "no valid conclusion" responses have the lowest percentage of recall errors at immediate recall and the second lowest at delay. Correctly given valid conclusions also have low percentage error rates at both immediate and delayed recall, correctly given "no valid conclusion" is well remembered at immediate recall, as generally found, but does less well at delay.

These findings can be explained in the following way. The evidence gathered so far suggests that subjects acting non-logically tend to draw more, not fewer valid conclusions. The response "no valid conclusion" is rather underused by these subjects suggesting that it is not the response used as a fall-back or guess. When it appears incorrectly one must assume that it is the result of cycles of testing finishing prematurely or, in the case of the figural effect, a valid conclusion being missed because it is antifigural. In this case then subjects may well have combined the premises and performed some testing, but do not go through the process of drawing a conclusion and interfering with the recall immediately after reading the syllogism in this way. Valid conclusions will be arrived at after a similar process, but on average, more testing. This does not confer an advantage on memory, as shown in the previous chapter however, and the act of drawing the conclusion seems to affect memory adversely at immediate recall. Correctly given "no valid conclusion" will also often result from testing and therefore the recall of this and incorrect "no valid conclusion" should be similar. As noted in chapter 4 however, some of these conclusions do not result from testing, but from the use of heuristics or logical principles. The lesser amount of consolidation will affect recall at delay as is found. Incorrect "no valid conclusion" should suffer less at delay because they are more likely to have been integrated and tested. Correct valid conclusions will be better retained at delay, as they result from a well-integrated structure. Conclusions given where none exists will always have poor recall, they result from matching in a high number of cases and will be based on a superficial understanding of the syllogism and suffer from the effect of drawing a conclusion. Incorrectly drawn conclusions where one does exist might be assumed to be the result of the same process and these also are found to have poor recall.

Thus, recall in a general sense appears to be affected by the strategy underlying the interpretation of the syllogism, if not the amount of processing. Combining the premises seems to be a major determining factor, although this does not relate to depth of processing in any simple way as shown by the analysis of numbers of models and recall in chapter 5. The strong "evening out" effect at delayed recall suggests that the information that improves immediate recall for valid conclusions for example is not robust enough to withstand thirty-two texts delay in a large number of cases. Given the assumption that subjects drawing incorrect conclusion are doing so because of a more superficial representation, can evidence be gathered about the nature of this representation from investigation of types of recall errors and types of response? Analysis shows not. There are no interesting variations in the types of errors found for syllogisms with different types of response, rather there is a global effect of response type, which increases the tendency for some part of the syllogism to be forgotten without affecting one type of information more than any other.

6.2 Ordering of Information in the Syllogism

The hypothesis that the preference for B as subject term is related to the subject's attribution of topic to the middle term of the syllogism conflicts with the requirement of the standard task that one of the end-terms be taken as the subject of the conclusion (and thus act as the focus or topic of the argument). However, in the case of the second experiment, the subjects is under no such restraint: choosing an individual gives him/her freedom to choose any of the three nouns in the syllogism as primary. It is of interest therefore to examine the orders preferred by the subjects in this experiment and relate these preferences to recall choices from the standard task.

Subjects were asked to decide if a valid individual followed from the premises of the syllogism. If they concluded that one did they were asked to write down the properties, no order was suggested, and make a note of whether the individual did or did not possess the property in question. Studying the results of this leads to a number of interesting observations. The first position noun is much more likely to be B than A or C (see figure 6.2). The relative frequencies of A and C reflect rather closely those of A–C and C–A conclusions for the standard task. This suggests that B does indeed take topic position when circumstances allow it to do so. A figural effect has already been demonstrated for Experiment 2 in chapter 4, the effect of the four figures on the frequency of B in first position has not been studied. The figural effect can again be observed, but in addition one can see that B collects a sizeable proportion of the choices

Noun in First Position	Percentage Occurence
A	25.0%
B	53.2%
C	21.8%

Figure 6.2: First position noun for the individuals task

Noun in First Position	AB–BC	BA–CB	AB–CB	BA–BC
A	62.3%	4.1%	74.5%	7.4%
B	35.9%	41.9%	5.4%	91.2%
C	1.8%	54.0%	20.0%	1.4%

Figure 6.3: First position noun by figure for the individuals task

for these two figures, particularly for BA–CB. This noun is a great deal less popular for AB–CB and is overwhelmingly the favourite for BA–BC.

It is difficult to imagine that the factors creating these preferences are not the same ones responsible for the figural effect. The type of task the subjects are asked to perform is certainly different in some important respects, but it is also similar in others, particularly in that to correctly solve the individuals problem the subjects must agglomerate the premises in the same way as for the standard task, even if the way they manipulate them to search for an individual is different from cycles of testing. If the figural effect results from the process of agglomeration, one would expect the choice of first noun for the individuals task to follow in the same way. It should be obvious that any explanation of this effect based on transitivity of a mental model will have difficulty accounting for these results. According to this hypothesis, B must always enter working memory between the end terms, because only in this way can a coherent model be built. Because A and C are never found in the same premise, the subject has at the time of combining the premises no information that would allow these two nouns to be linked, which would have to be the case if B were to enter the model first. Because of this, the occasions where individuals beginning with B are chosen cannot be explained and as these constitute more than chance occurences, it must be assumed that some other mechanism is responsible for the effect.

In chapters 4 and 5, an alternative account of the figural effect was examined and found to fit the recall and conclusion data rather better than Johnson-Laird's explanation. This consists of the observation that the end term which is subject is taken to be the subject of the conclusion and in the case where there are two candidates or no candidates for the subject, factors such as which noun was first mentioned play a greater or lesser part. In the individuals task, there is always a choice of two subjects for the conclusion. In only one figure are they both the same noun - BA–BC, thus the preference for individuals with B as the first noun. In one case B is never in subject term, for AB–CB and the incidence of B-first choices is therefore small. For the figural cases the end-term in subject position is most often chosen, but in the case of BA–CB the middle term is also a popular choice. Taking into account the other proposed factor - that the noun in the first premise is also preferred for the subject position, one can see that in the case of BA–CB there will be a conflict between B and C in first position, leading to nearly equal choices while for AB–BC A has both factors counting for it and therefore obtains a larger share of the choices. This also explains the disparity between A and C choices for AB–CB - A is chosen for first position more often because it is in the first premise.

The grammatical roles account of the figural effect can be applied to the results of this new task and can explain findings which the mental models formulation can not. It has already been shown that choosing an individual predisposes the subjects to different strategies, suggesting that to an extent the representation used by the subject is determined by the task they are requested to perform. The standard task, for example requires the subject to remove from the equation one of the nouns - the middle term - which is exactly the noun that appears to assume referential importance for the individuals task. It may be that the elimination of B for the standard task takes place at a stage after the consolidation of the premises. The consolidated representation may then reflect the preferences for noun order found in the individuals task, while the reasoning component of the process will be dependent on the type of task involved, thus the differential effects of cancellation and double hatching.

One must therefore look for evidence that underlying the recall for the standard task are structures that reflect the preferences found for the individuals task. Certain syllogisms lend themselves well to an analysis of the ways in which these structures might affect recall. Figure 6.4 shows the most popular noun orders for the EA, AE, EI and IE syllogisms from the individuals task. In no case was any other individual more popular, in all but one case the displayed choice accounted for 50% or more of the total and in that one case the most popular choice was "no valid individual". The AE and IE pairs tend to be given in an order that conforms to the order of the nouns of the syllogism, that is ABC for figures AB–BC and AB–CB and BAC for figures BA–CB and BA–BC. The EA and EI pairs tend to have A as the final term of the individual and often C as the first term. The effect this might have on memory is difficult to predict exactly as there are many ways that subjects might use a representation such as this to inform recall about the position of the nouns in the syllogism, assuming that this in fact happens at all. One simple mechanism that might be operated however, is for the subject to retain the individual, choose one of the first two nouns as the middle term and repeat that middle term somewhere in the individual retained, to end up with four nouns which then represent the noun order of the syllogism. Figure 6.5 shows that so long as the correct middle term is picked, the individuals commonly drawn for the AE and AI pairs will often lead to a correct recall and always to one where the noun roles of the syllogism are preserved and that for the EA/EI pairs, even if the correct middle term is picked correct recall never follows and in particular in all cases will be placed in the second premise. In half the cases included, C will appear in the first premise. In addition one would expect the first mentioned noun in the individual be the first mentioned term in the recalled noun order. These observations indicate some predictions to be examined:

- The EA and EI pairs will tend to have greater numbers of cases where A is recalled as having occurred in the second premise than the AE and IE pairs.
- The EA and EI pairs will tend to have greater numbers of cases where C is recalled as having occurred in the first premise than the AE and IE pairs.
- The first mentioned term in the individuals drawn from AE and IE syllogisms is also the first mentioned term in the original syllogism, so little of interest can be

Type of Syllogism	AB–BC	BA–CB	AB–CB	BA–BC
AE/IE	ABC	BAC	ABC	BAC
EA/EI	BCA	CBA	CBA	BCA

Figure 6.4: Most popular types of individuals for AE/IE and EA/EI syllogisms by figure

Type of Individual	Possible Noun Orders (with B as middle term)
ABC	ABBC BABC ABCB
BAC	BABC BACB
BCA	BCAB BCBA
CBA	BCBA CBAB

Figure 6.5: Possible noun orders by individuals

predicted for these cases in terms of which noun will be recalled in first position. However, in two of the EA and IE cases the first position noun would be predicted to be B when in only one case is this the actual first noun and in the other two cases, C would be the first noun.

These predictions are compared to the recall in figure 6.6, figure 6.7 and figure 6.8. The predictions for the recall of A and C are borne out. There are significant differences between all the comparisons across the two groups with the EA/EI pairs always having a greater percentage of error ($t=1.46$ $p<0.1$, $t=2.01$ $p<0.05$, for A in the second premise at immediate and delayed recall respectively, ($t=2.11$ $p<0.03$, $t=1.54$ $p<0.08$ for C in the first premise at immediate and delayed recall). Looking within the EA and EI pairs does show that instances where B was chosen as the first property in the individual are also instances where B was recalled as the first noun. This difference is significant at immediate recall ($t=1.85$ $p<0.06$), but not at delay ($t=1.15$ n.s.).

By isolating this particular group of syllogisms and comparing the individuals found to the recall profiles obtained in the first experiment it can be seen that there appears to

Type of Syllogism	Immediate Recall	Delayed Recall
AE/IE	7.5%	27.0%
EA/EI	10.5%	32.5%

Figure 6.6: Percentage of recalls with A in the second premise

Type of Syllogism	Immediate Recall	Delayed Recall
AE/IE	7.0%	13.1%
EA/EI	17.0%	10.4%

Figure 6.7: Percentage of recalls with C in the first premise

Figure	Immediate Recall	Delayed Recall
AB–BC and BA–BC	49.0%	48.0%
BA–CB and AB–CB	45.0%	47.0%

Figure 6.8: Percentage of recalls with B in first position for the EA/EI syllogisms

be recall phenomena that reflect the existence of individuals of this type. The question of why these particular individuals should be chosen in the first place has not yet been addressed and in the next section it will be shown that some, if not all of these phenomena can be traced to a common cause.

6.3 Positive and Negative Information

At points throughout this work it has been noted that information in the syllogism must be ordered in a particular way according to whether it is positive or negative. Research has already established that negative information is more difficult to process and tends to be understood as a conversion to affirmative statements (Wason (1959,1961,1963,1965), Donaldson (1959)). Wason makes the observation that positive statements are usually used to introduce information and negative ones to qualify those statements. There is found an intuitive observation that this experiment has shown empirically to be the case. The strong ordering effects of the quantifiers give dramatic evidence of this process at work and also shows a tendency to prefer universally quantified statements to be followed by particular ones.

6.3.1 Negative information and the recall data

Interactions between this and the recall of the nouns have already been observed. It can also be shown that these effects exist independently of the process of solving the syllogism. Two groups of syllogisms can be isolated, the IA and EA syllogisms and the AI and AE syllogisms of figures AB–BC and BA–CB. These two groups are not logically equivalent within themselves but each member of a group has a logically equivalent member in the other group. The IA and EA syllogisms have the non-preferred order of quantifiers, however and the AI and AE group the preferred order. The only factor that can affect the recall of these two groups is the order of the quantifiers, as logical factors have been controlled for. When the percentage of texts that had any kind of error at all is observed, it can be seen (figure 6.9) that a very large difference exists between the percentage of errors for the former group as opposed to the latter. Examining every category of recall error produces the same result, with the exception of type 2 noun order errors (reversal of noun within a premise) which are equally common for both groups (t=0.43 n.s., t=0.08 n.s. at immediate and delayed recall respectively). The

Type of Error	Preferred Order	Non-preferred Order	Significance
Any Error			
Immediate Recall	25.0%	38.0%	t=2.32 p<0.01
Delayed Recall	61.0%	84.0%	t=2.75 p<0.008
First Quantifier			
Immediate Recall	5.0%	16.0%	t=2.57 p<0.01
Delayed Recall	21.0%	50.0%	t=4.47 p<0.00001
Second Quantifier			
Immediate Recall	3.0%	17.0%	t=3.38 p<0.001
Delayed Recall	36.0%	61.0%	t=3.17 p<0.002
Reversal of Nouns			
Immediate Recall	14.0%	13.0%	t=0.43 n.s.
Delayed Recall	23.0%	22.0%	t=0.08 n.s.
A in Second Premise			
Immediate Recall	5.0%	16.0%	t=2.57 p<0.01
Delayed Recall	18.0%	36.0%	t=2.91 p<0.001
C in First Premise			
Immediate Recall	3.0%	13.0%	t=2.64 p<0.01
Delayed Recall	13.0%	27.0%	t=2.5 p<0.01

Figure 6.9: Errors in recall by preferred and non-preferred quantifier order

quantifiers are particularly affected as would be expected as are the numbers of recall errors where A or C appear in the inappropriate premise.

6.3.2 Negative information and choosing individuals

It was found in chapter 4 that this order preference for positive and negative information causes asymmetries in the type of conclusion chosen, both for particular syllogisms not generally thought to show these types of order effects and over the entire set. Preference for order of nouns in individuals is also affected by negative and positive information. Subjects in Experiment 2 were asked to find individuals which must exist under every circumstance, but who could either have (positive information) or not have (negative information) the attributes in the syllogism. There are, as already stated, eight possible individuals, if the order the nouns are given in the response is ignored. One of these types of individuals should never occur, others are found rarely. Figure 6.10 lists the individuals and the percentage of each noun order that were found for each one. One individual is excluded - -A-B-C - because it is very rarely given (only five cases appear overall) and can never be correct. The percentage of each noun order is given for the responses as a whole. The type of individual given and the noun order do not interact randomly. Inspection reveals that very rarely is a noun that conveys negative information in the individual found in first position. When only one noun is positive, then there is a strong tendency for it to be found in first position, when two nouns are positive then B tends to be the first choice, followed by A and finally C. The percentage of noun order for the total responses reflects this preference and also shows that noun orders where B is given last in the order are rather rare. This is due to the fact that noun orders where B is negative are very rarely given. These findings are summarised in figure 6.11. From this we can see that when a given noun is negative, it will take first position in only around 3% of cases, with one of the other two nouns (usually B) in first position. Looking at the nouns in last position it can be seen that negative A and C take last position in 87.1% and 79.2% of cases respectively. The exception is B. Even when this noun is negative, it is not to be found in last position. A is the preferred choice for this.

6.3.3 The overall effect of negative information

The previous two sections have demonstrated that, irrespective of the type of task subjects are asked to perform, the order of the information conveyed by the quantifiers has a profound effect on the types of responses subjects give and the number and type of recall errors they make. Subjects prefer to take in negative information after positive information, to reiterate Wason's comments: affirmative statements are generally used to introduce information and negative ones to qualify that which has been introduced. Subjects will find syllogisms that contain information in the opposite order difficult to process, it seems likely from the data that they reverse the order of the premises in memory, thus tending to recall both the quantifiers and the nouns in reverse order and also tend to structure conclusions in the standard task around the reversed, not the

	Order of Nouns					
Type of Individual	ABC	ACB	BAC	BCA	CAB	CBA
+A+B+C	31.1%	2.4%	28.1%	18.0%	2.4%	18.0%
-A+B+C	3.7%	0.0	9.2%	42.9%	0.0	44.2%
+A-B+C	75.0%	0.0	25.0%	0.0	0.0	0.0
-A-B+C	4.2%	0 0	0.0	0.0	20.8%	75.0%
+A+B-C	46.7%	0.0	46.1%	5.9%	0.0	1.3%
-A+B-C	2.8%	1.9%	57.0%	33.6%	0.9%	3.7%
+A-B-C	84.2%	0.0	0.0	0.0	5.3%	10.5%
Total	12.6%	0.6%	16.1%	11.9%	0.9%	10.6%

Figure 6.10: Percentage of type of noun order for each individual

	First Position Noun		
Negative Term	A	B	C
A	3.7%	52.1%	44.1%
B	14.2%	3.6%	82.1%
C	33.9%	62.9%	3.9%
	Last Position Noun		
	A	B	C
A	87.1%	0.0	12.9%
B	64.3%	17.9%	17.9%
C	19.1%	1.8%	79.2%

Figure 6.11: Position of nouns by negative terms

original order.

This explains at least in part the interactions found between the types of individuals drawn in Experiment 2 for certain syllogisms and recall errors for those same syllogisms in Experiment 1. The reason that certain terms are always found as the last noun of the individual is because they are negatively quantified. These same syllogisms when recalled will have the negative information appearing first. It has already been noted that this tends to cause a reversal of both quantifiers and nouns to achieve a preferred order. The kinds of recall errors found reflect this kind of reversal. One need not suppose however that these two phenomena result from two separate processes. The reordering noticed in the memory for these syllogisms may result in a representation from which an individual of the type and order found can be drawn. The fact that within this group preferences for order within the individual also affect to a lesser degree the recall results suggest that the relationship between two tasks operates at the level of a common representation.

6.4 Processing the Information in the Syllogism

Also in the data for Exeriment 2 has been found a preference for B as the first term in the individual. B is very rarely given a negative value and when it is it tends not to be placed at the end of the syllogism as are A and C. The figure BA–BC has an overwhelmingly high number of cases where B is the first term, an effect reminiscent of the figural effect and an explanation for this and other observations based on the grammatical roles theory has been given above. Relating this to the recall data produces some interesting hypotheses about the strategies subjects apply to understanding the syllogism. It is noted in chapter 3 that of the four figures, BA–BC is the quickest to be processed during the standard task and tends to be the preferred choice when recalling the syllogism immediately after reasoning and attracts fewest errors overall, particularly type 2 errors (reversal of the nouns within a premise). At delayed recall BA–BC enjoys no immunity from error and is no longer the most popular choice of figure, but it remains more popular than its counterpart parallel[1] figure AB–CB which fares badly in this respect at both immediate and delayed recall. To this can be added the observation that B is

[1]Here as in other parts of the text, the figures will be referred to as being "parallel" or "diagonal". This groups the figures into those that have the middle term with the same grammatical role and those in which it has different grammatical roles in each premise

the preferred subject term for the first premise overall and that it is associated with negative to positive changes in the first quantifier when it does take this position.

It seems that the arrangement of the nouns in syllogisms of the figure BA–BC has some property that makes them easy to process and preferred for recall. Returning once more to Kieras' work on the factors that subjects use to identify the topic of a set of sentences, he found that the first mentioned term, the subject term and the most frequently mentioned term were all more likely to be taken as the topic around which the discourse was organised. Recognising that these factors do not always coincide, he went on to discover that poorly structured texts, where topic is ambiguous, can be resolved but the processing times are higher and the choice of topic more heterogenous between subjects. The implications of some of these points have already been discussed in relation to the figural effect, the second experiment now shows how these factors might be influencing the processing of syllogisms in a broader sense. The four figures all possess nouns that on the basis of these factors identified by Kieras are candidates for the topic of the syllogism. However, what differs between figures is the extent to which these factors coincide in indicating the same topic. On the basis of subject position, A, B or C could be topic, depending on the figure. The first mentioned noun could be either A or B and the most frequently mentioned item is B for all figures. How these factors interact is shown in figure 6.13 from which it is clear that BA–BC possesses the structure that most unambiguously identifies a topic - B, the two diagonal figures are split between A and B in the case of AB–BC and C and B in the case of BA–CB. AB–CB has the most conflicting arrangement, any noun could be the topic, A perhaps enjoying a slight advantage because of it being both a subject and the first mentioned item. In terms of processing then, BA–BC should be the easiest to process, then AB–BC and BA–CB and finally AB–CB.

These observations provide a good fit to the noun order frequencies by figure collected from Experiment 2 as already noted and for the reading time and recall accuracy and preferences observed at immediate recall (see figure 6.12 for reading times). The change in recall pattern at delayed recall remains to be explained, however. Firstly it is pertinent to note that many of the effects relating to noun order recall errors disappear at delay. Waiting for thirty-two texts before recalling the text again increases the frequency of all types of error and tends to reduce rather than increase the differences

Figure	Reading time /ms
AB–BC	2445
BA–CB	2615
AB–CB	2477
BA–BC	2260

Figure 6.12: Processing time by figure

in recall success between a number of groups. The disappearance of BA–BC's advantage is probably part of that tendency. Other preferences for these particular variables do appear however. The diagonal figures are more often chosen at delayed recall than the parallel ones, the opposite being true at immediate recall. One needs to consider the whole process of deriving a conclusion to see why this might be so.

It has been proposed that BA–BC is the easiest figure to process because it unambiguously indicates a topic for the argument. However, the topic it indicates happens to be the one noun that cannot play a part in the conclusion, should one exist. This implies that once the syllogism has been interpreted and the premises combined, in the case of BA–BC the subject must choose another topic for the syllogism which directly conflicts with all the information contained in the text itself. In the cases of the diagonal figures, the topic is ambiguous anyway and the choice the subject makes has a 50% chance of being one of the end terms, although the choice of topic will depend on the weighting given to the three factors. In a significant proportion of cases, the choice of topic for the conclusion will be the same as that chosen for the text as a whole and no interference will result. For AB–CB the choice of topic for the text will hardly ever interfere with the choice of conclusion topic as the indicators are that the former will rarely be B. It might be expected that this interference will only really make itself felt at delayed recall, when the subjects will have to rely on the least superficial aspects of the syllogism to inform their memory and as has already been noticed, the internal grammar of the premises, upon which the choice of topic is largely dependent, is fragile information. What the subject will retain at delay is instructions about the topic of the syllogism which will conflict to a greater or lesser extent. To demonstrate that this is so, in figure 6.14 the

	Factors Affecting Choice of Topic		
Figure	First-mentioned Term	Subject of Premise	Most Frequently Mentioned
AB–BC	A	A/B	B
BA–CB	B	C/B	B
AB–CB	A	A/C	B
BA–BC	B	B/B	B

Figure 6.13: Choice of topic for each figure according to the factors identified by Keiras

percentage of errors in the noun order at immediate and delayed recall for each figure, divided by whether a conclusion was drawn or not are considered. The final distinction is important because although giving the response "no valid conclusion" must in some cases involve the searching out of a topic for putative (but finally rejected) conclusions, in a large number of cases, as already established, no such process is attempted when this type of answer is given. A conclusion, valid or not, must always have a topic and will always invoke the interference under discussion.

The data shows that the differences are as predicted: at immediate recall the variation in percentage of correct noun recalls is increased for BA–BC. Responding "no valid conclusion" seems to create particularly little interference for the parallel figures. At delayed recall, the stage of real interest, it can be seen that BA–BC has similar proportions of errors for cases where no conclusion was drawn as other figures. It is only in the cases where a conclusion is given that the percentage of correct noun recalls drops sharply, exactly as predicted. It is interesting to note that drawing a conclusion seems to improve the recall of AB–CB, the other parallel figure. One might suppose that this is because, as pointed out above that drawing a conclusion will not interfere with the topic chosen and this gives it a relative advantage compared to the other figures which are all likely to suffer a degree of interference.

6.4.1 Recall of the Figures and Types of Conclusion

These processing factors can also be of use in explaining another recall phenomenon, also related to figure and to the type of conclusion drawn. It has been found that at immediate recall there is a tendency to recall parallel rather than diagonal figures

Immediate Recall		
Figure	Conclusion Given	No Conclusion Given
AB–BC	69.3%	75.9%
BA–CB	63.8%	74.1%
AB–CB	57.2%	83.8%
BA–BC	66.5%	85.5%
Delayed Recall		
Figure	Conclusion Given	No Conclusion Given
AB–BC	42.9%	39.5%
BA–CB	40.7%	35.8%
AB–CB	57.9%	35.5%
BA–BC	31.3%	38.7%

Figure 6.14: Percentage of correctly recalled noun orders by figure and conclusion given

Immediate Recall			
Type of Figure Chosen	No Conclusion Given	Conclusion Given	Percentage of Total
Parallel	54.8%	47.9%	51.3%
Diagonal	45.2%	52.1%	48.7%
Delayed Recall			
Type of Figure Chosen	No Conclusion Given	Conclusion Given	Percentage of Total
Parallel	46.4%	41.4%	43.9%
Diagonal	53.6%	58.6%	56.1%

Figure 6.15: Percentage of type of figure chosen by response given to syllogism

(figre 6.15). A breakdown by type of conclusion shows that this tendency is confined to cases where no conclusion was given. Syllogisms that were given a conclusion, correctly or incorrectly, tend to be recalled as having a diagonal figure. At delayed recall on the other hand, it can be seen that the overall tendency is for diagonal figures to be chosen and this is especially true of syllogisms for which a conclusion was given. Syllogisms where no conclusion was given also tend to be recalled as diagonal, but this trend is less strong.

It has already been shown that for one figure at least, syllogisms given no conclusion are less likely to suffer interference as a result of conflict between topics. This is because, it is argued, in many cases no topic for possible conclusions has been derived and therefore there is no conflict. In these cases, the likely choice for the recall of the figure will be that which is easiest to process - BA–BC, as there is no other coding of the syllogism other than the initial interpretation. This preference is strong enough to create a preference for parallel figures overall. Syllogisms for whom a conclusion has been drawn will have two codings to choose from: the initial one and the one for the conclusion. In some cases there will be no conflict, in others the choices of topic will be different, but the most recent coding will be that for the conclusion. The preference will be then, to choose a figure that preserves the subject-predicate structure of the conclusion - a diagonal figure. At delayed recall, this tendency is exaggerated and in addition the syllogisms for whom there was given no valid conclusion are also mostly recalled as diagonal. The cases from this group where coding was given for the conclusion are now also using this information at recall, because, as has already been shown, surface information in the syllogism is fragile and often superceded by the coding used for the conclusion[2].

6.5 The Nature of the Representation

In this section the implications of these finding for the representation underlying syllogistic reasoning will be discussed. Of particular interest here is the form of representation used by logically experienced subjects - the nature of the agglomerated representation. The issue of interest is the status of such a representation - is it task specific, or is there any evidence that all subjects tend to prefer one type of notation over another?

The evidence studied in chapters 4 and 5 tended to show that subjects were using a strategy that would reveal the presence of individuals, not normally evident in the standard task. There was also evidence that the presence or absence of cancellation helped this group of subjects to make decisions about the presence or absence of valid individuals. In the standard task facilitation of this kind was found, but cancellation could not have been responsible for it. It was decided therefore that subject had abstracted logical principles and used those to detect the likely absence of a valid conclusion. Double-hatching was also not found to hinder subjects in making this type of decision. That is not to say, however that cancellation and double-hatching were entirely without effect - in fact these factors were found to have the opposite effect to that predicted for syllogisms without valid conclusions and worsen response accuracy. It was then suggested that in the standard task, where a relationship between A and C is the object of interest, a syllogism that either fails to allow the B term to be eliminated or that does not yield a representation where a relationship between A and C can provide a starting point for testing will be hard to interpret at all, much in the same way that indeterminate texts are hard to interpret. To back this hypothesis, the recall data, grouped according to cancellation and double-hatching shows that non-cancelling

[2]Another example of this can be found in chapter 3, where subject/predicate roles at recall are changed to resemble those of the conclusion

and non-double-hatched texts are more difficult to recall.

The safest position to take might be to state that exact forms of representation are task- and individual-specific. There is good evidence that subjects attempting the individuals task use a representation that includes cancellation and double-hatching, the absence of such results for the reasoning data of the first experiment could simply be because the task did not invite such a notation. It has also been one of the major points of this work to demonstrate that subjects do not bring to this type of reasoning problem a ready-made type of strategy. Depending on their previous experience and general ability for problems of this kind, subjects learn through the course of the task to find a means of representing the syllogism that will allow them to draw and assess conclusions. What all subjects do appear to bring to the task, is not anything like a mental model, but a series of preferences based on language comprehension, that underlie the figural effect and that reveal themselves in various recall phenomena. These language effects have not been shown to significantly affect subjects' ability to reason successfully, apart from influencing subjects away from considering certain starting points for conclusions, as has been shown to be the case with the figural effect. These factors are not lost once the syllogism has been processed, however, or they would not appear at recall. In some way, they must form part of the representation, therefore.

To say, however, that subjects do include information about cancellation and double-hatching, as the data seem to suggest, may not be the same as making a strong claim about the nature of their representation. Just as Stenning has shown that mental models and Euler's circles can be made to be equivalent by the addition of notation, so the additional phenomena observed in these augmented representations can be expressed in a mental models terminology. Double-hatching might be expressible in terms of "straight-through links" and cancellation, it must be obvious, is not observable from the representation, but from a simple, unagglomerated formulation of the two premises. Although at some level the two representations are indistinguishable it should be noted that there are two respects in which an Euler's circles based notation can be shown to have an advantage over mental models.

1. It has been shown at points throughout this work that the quantifiers "some" and "some..not" are often confused. There is evidence both from the reasoning data of

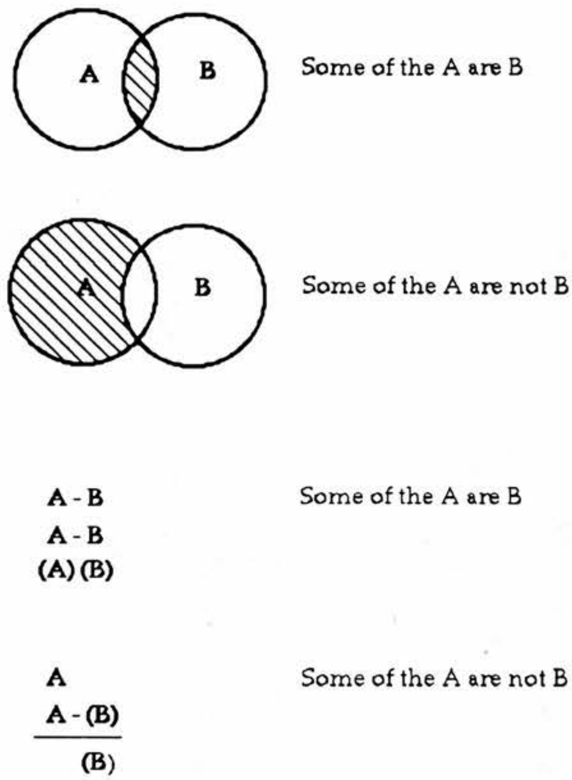


Figure 6.16: A comparison of Euler's circles and mental models representations for the quantifiers "some" and "some..not"

Experiment One and the recall data from the same experiment that supports this proposal and there is evidence from earlier studies that gives the same result. As has been already argued this type of confusion follows naturally from a conception of the representation as being Euler's circles based. Figure 6.16 compares this and the mental models formulation to demonstrate that all that is required to confuse the two quantifiers in the former notation is misplaced shading, or asserting an individual to exist where this is not certain. The mental models interpretations however, are hard to confuse. The impression has been gained that this has been considered one of mental model's strengths, however the data suggests it is, in fact, a weakness.

2. The evidence collected and examined fails to support the mental models explanation of the figural effect, but rather suggests a mechanism based on linguistic features aimed at choosing topic. The position can be taken that mental models

offers simply a description of the figural effect and that Euler's circles, suitably amended could offer the same, rendering the two notations equal. However, the finding that B is the favoured choice of topic for the figure BA–BC is damaging to the position of mental models, because this noun cannot be represented as having primary position in this formulation in the way used so far (making the primary noun the left-most in the model), without asserting links between nouns that do not exist. Euler's circles suffer no such problems. The only way to save mental models from this problem is to also use additional notation to indicate topic status leading to two contradictory notations within the same representation.

6.6 Summary

It has been shown that while the numbers of models required to solve a syllogism do not affect recall in the way depth of processing accounts would suggest, the nature of the response given to the syllogism in Experiment One does affect recall. These differences do seem to relate to the probable amount of processing the syllogism received before the response was given, but this cannot be related to different types of recall errors occurring for different types of processing - it seems to be a global effect on the tendency to commit an error at recall.

The ordering of the information in the syllogism, in the form of figure is found to relate to strong preferences for certain orders of nouns to be chosen in the individuals task (Experiment 2). This in turn is found to be explicable by reference to linguistic factors affecting the attribution of topic in the text. In the recall data this can be seen as causing preferences for certain patterns of recall immediately after giving the reasoning response, which decay by delayed recall because at this stage what is used for recall are features of the agglomerated representation which conflicts with the choice of topic made at the intial stages of interpretation.

It is also found that certain syllogisms in Experiment Two tended to attract certain orders of nouns for the individuals determined to exist for these cases. The recall data from the first experiment tends also to show evidence of these orderings. These two phenomena can be traced to the already found preference for positive information to be followed by negative information (where negative information has to be used at all). This tends to force a certain order of nouns on both the choice of individual and the recall, and as shown in an earlier chapter affects the type of conclusion chosen in the standard task. The effect of this preference for the ordering of information has profound effects on recall, in that it increases recall errors of every kind, not simply of the quantifiers. Grouping the syllogisms to create cadres that are logically symmetrical but that differ in that one group has the preferred order of information and the other group the non-preferred order shows that this effect operates independently of the logic of the syllogism.

Discussion of the status of the candidate representations for agglomerated premises leads to the conclusion that at some level they are indistinguishable empirically and in any event are the result of learnt strategies, rather than making claim to a universal code for cognitive operations. There are, however, one or two findings that suggest that the Euler's circles approach may be more descriptively adequate. There may be ways in which this claim could be cemented by further experimentation, but this is without the scope of this work.

Chapter 7

Conclusions

7.1 Reasoning effects

In chapter 1, a number of explanations are given for subjects' characteristic performance for syllogistic reasoning. Some of these claims have been investigated by using the reasoning data collected from two experiments and the recall data from experiment 1. The evidence from this suggests that explaining subject's choices will not be possible by choosing to support any one of these theories alone.

While it is certainly true that subject's observed rationality must be explained, it is also the case that effects of quantifier misinterpretation and evidence of matching and atmospheric strategies have been found. These effects do not seem to appear for all subjects in circumstances where the syllogism has become too difficult for them to solve, but appear to be produced by a certain sub-group of subjects. Therefore some subjects are not using rationality to solve syllogisms and it should be noted that misinterpreting the quantifiers does not appear to be restricted to this particular group of subjects, so imperfect understanding of the logical information represented by the syllogism appear to be a fairly widespread trait. It is also the case that subjects do not adopt a uniform strategy for all syllogisms. They appear to recognise at some point during both the standard and the individuals task that certain syllogisms will never have a valid conclusion. How they make this decision seems to depend on the task they are being asked to perform, for the standard task they appear to have internalised logical maxims for example of the form "a syllogism with two particular quantifiers never has a valid conclusion". In this way they are able to answer certain syllogisms correctly without recourse to cycles of testing, thus improving their performance on this group.

For the individuals task they seem to use the presence or absence of cancellation in the original syllogism to decide whether there is a conclusion or not.

One wants a theory of syllogistic reasoning that will allow all these effects to take place. It is important that subjects should be able to represent the syllogism in such a way that they will fail to be able to draw correct valid conclusions even to those syllogisms that require no testing. Their representation of the syllogisms must contain the information that influences them to produce conclusions in accordance with the figural effect. Other subjects are behaving rationally and it seems to be the case that they do this very much in the way that Johnson-Laird describes: by agglomerating the premises and using cycles of testing to isolate correct conclusions. They must also be able to abstract from these with experience certain logical principles that will enable them to answer syllogisms correctly without going through this process and they too will show a figural effect.

It seems reasonable to suppose that all subjects possess a grammatical analysis of the syllogism which will conform to a greater or lesser extent with their expectations about the topic or focus of the syllogism. This will influence the ease with which they will process certain figures and it will create a preference for the terms of the syllogism to appear in a certain order, both in the individuals task and the standard task.

7.2 The Figural Effect

It has already been mentioned that subjects must possess a representation that will influence them to choose conclusions in accordance with the figural effect even in situations where they can be shown not to have created an agglomerated representation. There is further evidence, already summarised that points to the fact that the figural effect cannot be due to a mental models type representation combined with working memory effects, but is much more likely to be the result of a coding of the syllogism in terms of topic. The two main points of evidence are:

1. The recall data from the first experiment does not show many of the effects that would be expected if the manipulations of the nouns in the syllogism, necessary to create a transitive mental model actually take place. A comparison of the predictions of this theory, in two versions and a grammatical roles theory relative

to error rates for recall of the noun order find that the grammatical roles theory is supported.

2. The results from the individuals task show that subjects often prefer orders beginning with the noun B. Making the (not unwarranted) assumption that this effect springs from the same source as the figural effect, one finds that the mental models explanation cannot account for this effect, as it could only be produced by a non-transitive model, which is inadmissable for this notation. Given that the transitive nature of the mental model only provides a description of the figural effect and that it is the operation of memory on this model that creates the preference, the lack of recall data supporting this explanation is particularly damaging.

The grammatical roles explanation is accepted as the cause of the figural effect. This implies that the early stages of interpretation of a syllogism are characterised by the search for the topic or focus of the argument. This process is heavily influenced by linguistic factors such as which terms are in subject and predicate position (hence the use of the name "grammatical roles" for the explanation) and also which term is mentioned first and which most frequently mentioned. These factors combine in some cases to lead the subject to choose B as the topic for the argument, which causes no problems for the individuals task, but of course will create difficulties for the standard task where the topic of the conclusion must be A or C. This conflict leads to increased errors for certain figures at delayed recall and the overall profiles suggest that the information about topic derived from the conclusion has a strong effect on what is chosen for recall at delay. This information must form an integral and indistinguishable part of the representation of the syllogism and demonstrates that in this case subjects are not able to differentiate between information created as part of the process of inference and information actually occurring in the syllogism. Whether they can distinguish inferred material from the original in terms of the nature of the conclusion is more difficult to determine. Adequate analysis of this point has not been carried out, one reason for this being that attempts to find instances where the conclusion quantifier differs adequately from the originals and isolate particular factors quickly results in a data set that is too small. The effect of drawing negative conclusions to positive premises and vice versa is ruled out because of this and the same is true for particular and universal conclusions.

7.3 The move to positive information followed by negative information

The finding that people prefer to assert positive information before making a negative statement is not a new one. A similar effect is found in this data in the form of a very strong preference for quantifier order between the premises. This can be summarised as recall errors resulting in a shift from negative features to positive ones in the first premise and the opposite for the second premise. There is also a shift from universal statements to particular ones in the second premise. This in turn affects the recall of the nouns in the syllogism, some cases of which have already been noted above. The preference for B in subject position when the first quantifier changes from being negative to positive may simply indicate that this change disrupts other parts of the memory less than other quantifier errors, as in this case there are fewer instances of the noun C being wrongly recalled as having occurred in the first premise. When the second quantifier changes from positive to negative, there is a higher number of errors where A is recalled as having been in the second premise, when it can only ever actually be found in the first premise. There are fewer reversals of nouns within a premise when this type of quantifier error happens, so its effect is not just to cause greater amounts of superficial disruption of memory. The possibility that A is being carried into the second premise by some kind of feature swopping of the quantifiers is yet to be investigated. It is certainly that case that when two quantifiers are reversed in order the premises are often recalled in reverse order also, so there is a tendency to keep the information in premises together. In many cases where A is found in the second premise it will have replaced B as the repeated term: quantifier change may specifically disrupt information about which noun was repeated. This may also be coupled with the fact that first premises of the form BA are more prone to this type of error, presumably because the subject assumes that the second-mentioned noun in the first sentence is going to be the repeated term, (the second premise is not available at the point the first is read) and then repeats the error at recall. The change in quantifier may accentuate this trend.

A similar effect of negative information is found for the individuals experiment. Here, negative terms are very often found last in the individual and if there is only one positive term, it is usually found first.

The representation of the syllogism must therefore either lose the order information pertaining to all the quantifiers, which seems unlikely as subjects are usually rather unlikely to reverse the order of the quantifiers, or it must reorder information that does not conform to the subject's preference.

7.4 The Nature of the Representation

The process of understanding a syllogism begins with an initial linguistic analysis which assigns a value to the topic of the syllogism. At this stage also, possibly, the subject must resolve indeterminacy created by non-cancellation, or will use information about cancellation to decide on the presence of an individual, if this is the task being performed. If the standard task is being attempted, experienced subjects may perceive from the quantifiers alone that there is no conclusion and never attempt an agglomeration of the premises. In some cases, subjects will fail to realise that agglomeration is necessary and will draw a conclusion on the basis of the linguistic analysis alone. The information in the quantifiers will be reordered at this stage so that negative information follows positive and particular information universal. This will usually match the quantifiers in the syllogism or be drawn on the basis of atmosphere. Subjects who agglomerate will do so and then search for a conclusion with the topic term as its subject. In cases where B has been chosen a new topic will be needed. Once a conclusion has been derived, it is often tested. In cases where testing fails the response "no valid conclusion" tends to be given (because the original has been invalidated). Subjects who go on to test another possibility and find the correct solution remember these syllogisms no better, possibly because the conclusion differs from the original and they are no longer able to tell the two types of information apart. This and other recall findings suggest that the inferred material is incorporated into the syllogism and that it is often used at delay as a guide to recall. Grammatical information in the original syllogism is often lost as the subject is influenced by other considerations such as ease of processing or the inferred material at the recall stage.

Reasoning with a syllogism is not simply a question of using rational processes therefore. It is part of language processing and tends to be treated as such. The representations that subjects use when they grasp the logic of the syllogism are at some level all equivalent - they are all agglomerative, all can be used to test candidate conclusions. It seems unlikely that any one of these can lay claim to capturing the true nature of the coding in which reasoning is carried out. What is more likely is that all these representations are learned means for subjects to make explicit a process that remains hidden from view.

7.5 Implications for Future Research

The findings of this research allow a number of predictions to be made which could be the subject of future research.

7.5.1 The Figural Effect

The major claim here is that the figural effect is not the result of "first-in first-out" memory structures which are themselves dependent on a particular form of representation, but rather the consequence of the strategies adopted for successful language processing, based on awareness of simple grammatical roles. As a result one might expect this to be a phenomenen that emerges with language development and it should be possible to link the phenomenen with the appearance of certain types of linguistic competence or performance. The difficulty here would be to create syllogistic paradigms within the capabilities of children of the right age, but there have already been many successful attempts to key into children's reasoning performance and demonstrate the parallel key linguistic factors that are related to (or indeed underpin) the emergence of types of logical performance.

A similar set of possibilities exists for linking the operation of the figural effect with the linguistic competence of adult subjects. One would not expect, for example to find the effect varying very much with variations in literacy to the level of average or just below average functioning as the effect seems to be the result of a type of language comprehension strategy basic to everyday successful functioning. It should be noted however that one of the reasons for selecting subjects for this experiment in the way that was done was in order to collect data that was not distorted by relative inexperience with these types of problems and in particular the concept of validity versus truth. It is possible that using groups of subjects with differing levels of literacy will demonstrate different types of reasoning performance, but if any effect is noted for the figural effect this would have to be shown to be the result of a demonstrable difference in linguistic strategy at the level that has been described.

Those who are language impaired may well exhibit a different phenomena and again this would be worthy of research, should it be possible to find a suitable reasoning task in which the figural effect could be demonstrated to operate. Establishing the constancy of the figural effect under conditions where subjects are encouraged to use representations other than mental models or demonstrating that explicitly teaching the mental models technique does not exaggerate the effect is evidence that at least at the top level, representation is not a factor in the causation of the effect.

7.5.2 Positive and Negative Information

The strong effect of positive and negative information both on memory and reasoning results is worthy of further exploration. Again here is a linguistic strategy "invading" or rather determining a cognitive task and the same predictions about its appearance during the course of language development, its robustness in adults who use and understand language without impairment and its likely disppearance or distortion in those who do not have a normal language function are all relevant here.

There are also a number of specific assumptions made about the way in which the information in the quantifiers interacts with the nouns in the two sentences and it would be of value to examine more closely exactly how the mechanism operates. For example the hypothesis that moving a quantifier from one sentence to another in memory also causes one of the nouns to be moved in memory should be established experimentally.It would also be useful to investigate whether subjects do make assumptions about which noun is to be the repeated term and hence the topic of the syllogism. The set of criteria established by Chafe as interacting to determine "subjecthood" could be used to generate statistical predictions about the likelihood of errors occuring, both in reasoning and recall and used to model error rates. One might also like to look in more detail at the question of how inferred material becomes incorporated into the memory for the syllogism and presumably the representation so that it cannot be distinguished from the information originally given. Some thought would have to be given to ways of creating an experimental design where the original material differed significantly from the inferred material so that the emergence of the latter in a memory task could be isolated.

7.5.3 Matchers Versus Non-matchers

Finally, it should be possible to further explore the question of the use of non-logical strategies to answer syllogisms. Here finding a group of subjects that are relatively inexperienced in questions of logic would be desirable in order to look at reasoning results where the likelihood of matching is much greater and also to then examine recall data in comparison with a more experienced group to see if the linkage between the type of answer given to a syllogism and its correctness still persists and is linked with similar qualitative memory phenomena.

Bartlett, F.C. (1932) *Remembering.* CUP.

Begg, I. and Denny, J.P. (1969) Empirical reconciliation of atmosphere and conversion interpretations of syllogistic reasoning. *Journal of Experimental Psychology.* **81.** p351-354.

Begg, I. and Harris, P. (1982) On the interpretation of syllogisms. *Journal of Verbal Learning and Verbal Behaviour*, **21**, p595-620.

Bransford, J. D. and Franks, J. J. (1971) The abstraction of linguistic ideas. *Cognitive Psychology*, **2**, 331-350.

Chafer, W. (1976) Giveness, Contrastiveness, subjects and topics in W. Chafer (ed.) *Subject and Topic.* New York Academic Press.

Chapman, I.J. and Chapman, J.P. (1959) Atmosphere effect re-examined. *Journal of Experimental Psychology*, **58**, p220-226.

Craik, F.I.M. and Lockhart, R.S. (1972) Levels of processing: a framework for memory research. *Journal of Verbal Learning and Verbal Behaviour*. **11, p671-684.**

Dickstein, L.S.(1978) The effect of figure on syllogistic reasoning. *Memory and cognition*, **6, p76-83.**

Donaldson, M. (1959) Positive and Negative Information in Matching Problems. *British Journal of Psychology*, **50, p233-262.**

Ebbinghaus, H. (1964) *Memory: A contribution to experimental psychology.* **New York: Dover. Originally published 1885.**

Einstein, G.O. and McDaniel, M.A. (1990) Encoding and recall of texts: the impotance of material appropriate processing. *Journal of Memory and Language*, **29, p566-581.**

Erickson, J.R. (1974) A set analysis theory of behavior in formal syllogistic reasoning tasks. In R. Solso (ed.) *Loyola Symposium on Cognition. Vol. 2.* **Hillsdale,New Jersey: Erlbaum.**

Erickson, J.R. (1978) Research on syllogistic reasoning. in Revlin and Mayer (eds.) *Human Reasoning.* **Wahington D.C.: Winston.**

Evans, J.S.B.T. (1973) Matching bias in the selection task *British Journal of Psychology*, **64, p391-397.**

Fisher, D.L. (1981) A three-factor model of syllogistic reasoning: the study of isolable stages. *Memory and Cognition*, **9, p496-514.**

Frase, L.T.(1968) Associative factors in syllogistic reasoning. *Journal of Experimental Psychology*, 76, p407-412.

Gilhooley, K.J. and Wetherick, N.E. (1989) Syllogistic reasoning: the premise generation task. Paper presented at British Psychological Society Cognitive Section Meeting, Cambridge.

Grice, H.P. (1975) Logic and Conversation in P.Cole and J.P.Morgan (eds) *Syntax and Semantics vol. 3: Speech Acts*, New York: Seminar Press.

Guyote, M.J. and Sternberg, R.J.(1978) A transitive-chain theory of syllogistic reasoning. Technical Report No.5, Department of psychology, Yale University. Revised version in: *Cognitive Psychology*, 13, p461-525.1981.

Inder, R. (1986) The computer simulation of syllogism solving using restricted mental models. PhD Thesis, Centre for Cognitive Science, Edinburgh University.

Jarvella (1971) Syntactic processing of connected speech. *Journal of Verbal Learning and Verbal Behaviour*, 10, p409-416.

Johnson-Laird, P.N. (1983) *Mental Models*. C.U.P.

Johnson-Laird, P.N. and Bara,B. (1982) The figural effect in syllogistic reasoning. Mimeo, Laboratory of Experimental Psychology,Sussex.

Johnson-Laird, P.N. and Steedman, M. (1978) The psychology of syllogisms. *Cognitive Psychology*, 10, p64-99.

Kieras, D.E. (1980) Initial mention as a signal to thematic content in technical passages. *Memory and Cognition*, 8, p345-353.

Kieras, D.E. (1981) The role of major referents and sentence topics in the construction of passage macrostructure, *Discourse Processes*, 4, p1-15.

Kintsch, W. et al (1990) Sentence memory: a theoretical analysis. *Memory and Language*, 29 no. 2, p135-151.

Kintsch, W. (1984) *The representation of meaning in memory*. Hillsdale NJ.: Erlbaum.

Kintsch, W. and van Dijk, T.A. (1983) Strategies of discourse comprehension. New York Academic Press.

Kintsch, W. and van Dijk, T.A. (1978) Towards a model of discourse comprehension and production. *Psychological Review*, 85, p363-394.

Lea, R.B. et al (1990) Predicting propositional logic inferences in text comprehension. *Memory and Language*, 29 no. 3, p361-375.

Miller, G.A. (1956) The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychology Review*, 63, p81-96.

Newell, A.(1981) Reasoning, problem solving and decision processes: the problem space as a fundamental category. In R. Nickerson (ed.) *Attention and Performance, Vol.8.*. Hillsdale, NJ: Erlbaum.

Newell, A. and Simon, H. A. (1956) The logic theory machine. A complex information processing system. *Transactions on Information Theory*, 3, p61-79.

Newstead, S.E. and Griggs, R.A.(1983) Drawing inferences from quantified statements; a study of the square of opposition. *Journal of Verbal Learning and Verbal Behaviour*, 22, p535-546.

Newstead, S.E. (1989) Interpretational errors in syllogistic reasoning. *Memory and Language*, 28, p78-92.

Oakhill, J. et al (1989) Believability and Syllogistic Reasoning. *Cognition*, 31, p117-140.

Oaksford, M. (1988) Cognition and Inquiry: Conditionals and Pragmatics. PhD Thesis, Centre for Cognitive Science, Edinburgh University.

Pezzoli, J.A. and Frase, L.T. (1968) Mediated facilitation of of syllogistic reasoning. *Journal of Experimental Psychology*, 78, p228-232.

Politzer, G. (1990) Immediate deduction between quantified sentences in Gilhooley, K.J., Keane, M.T.G., Logie, R. H., and Erdos, G. (eds.) *Lines of Thinking*, vol. 1. John Wiley and Sons Ltd.

Politzer, G. (1989) A neo-atmospheric approach to syllogistic reasoning. Paper presented at British Psychological Society Cognitive Section Meeting, Cambridge.

Ratcliffe, R. and McKoon, G. (1989) Similarity information versus relational information: differences in the time course of retrieval. *Cognitive Psychology*, 21, p135-155.

Revlin, R. and Leirer, V.O. (1980) The belief bias effect in formal reasoning. The influence of knowledge on topic. *Memory and Cognition*, 8, p447-458.

Revlis, R. (1975) Two models of syllogistic reasoning: feature selection and conversion. *Journal of verbal learning and verbal behavior*, 14, p180-95.

Sachs, J. S. (1967) Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics* 2 p437-442.

Stenning, K. (1986) On making models: a study of constructive memory. Chapter 7. in Myers, T., Brown, K. and McGonigle, B. (eds.) *Reasoning and Discourse Processes*, p165-185. London: Academic Press.

Stenning, K. (1989) Modelling memory for models. In Ezquerro, J and Larrazabal, J.M. (eds.) *First International Colloquium on Cognitive Science*, Universidad del Pais Vasco, San Sebastian, 1989: Kluwer

Stenning, K.(1990) Distinguishing Conceptual and Empirical Issues about Mental Models

Stenning, K. and Oberlander, J.(1991) Reasoning with words, pictures and calculi: computation versus justification. In Barwise, J. et. al., (eds.) *Second International Conference on Situation Theory and its Applications*, Kinloch Rannoch, Scotland, September 1990, CSLI: Stanford.

Stenning, K., Shepherd, M. and Levy, J.(1988) On the construction of representations for individuals from descriptions in text. *Language and Cognitive Processes*, 2, p129-164.

Tyler, S.W. et al (1979) Cognitive effort and memory. *Journal of Experimental Psychology: Human Learning and Memory*, 5, p607-617.

Wason, P.C. (1959) The processing of positive and negative information. *Quarterly Journal of Experimental Psychology*, 11, p92-107.

Wason, P.C. (1961) Responses to Affirmative and negative binary statements. *British Journal of Psychology*, 52, p133-142.

Wason, P.C. (1963) Negatives: denotation and connotation

Wason, P.C. (1983) Rationality and the selection task. In J. St. B. T. Evans (ed.) *Thinking and Reasoning*. London: Routledge and Kegan Paul.

Wason, P.C. (1965) Contexts of plausible denial. *Journal of Verbal Learning and Verbal Behaviour*, 4, p7-11.

Wetherick, N.E. and Gilhooley, K.J. (1990) Syllogistic reasoning: effects of premise order. in Gilhooley, K.J., Keane, M.T.G., Logie, R. H., and

Erdos, G. (eds.) *Lines of Thinking*, vol. 1. John Wiley and Sons Ltd.

Wilkins, M.C.(1928) The effect of changed material on the ability to formal syllogistic reasoning. *Archives of Psychology*, 16 No.102.

Woodworth, R.S.and Sells, S.B. (1935) An atmosphere effect in formal syllogistic reasoning. *Journal of Experimental Psychology*, 18, p451-60.

Appendix A

Experiment One

Correct responses are given in bold type, incorrect in light type. The number of subjects giving each response is noted in brackets.

	All	Some	None	Some..not
All	**All A are C(18)** NVC(3) Some A are C(4)	**Some A are C(21)** **Some C are A(3)** All A are C(1)	**Some C not A(4)** NVC(8) No A are C(7) No C are A(3) Some C are (1) Some A are C(2)	**NVC(9)** Some A not C(7) Some A are C(5) Some C not A(3) Some C are A(1)
Some	**NVC(14)** Some A are C(5) Some C are A(4) Some C not A(2)	**NVC(18)** Some A are C(4) Some C are A(1) Some C not A (1) Some A not C(1)	**Some C not A(6)** NVC(11) Some A not C(6) No A are C(2)	**NVC(18)** Some A not C(4) Some A are C(2) Some C not A(1)
None	**No A are C(19)** **No C are A(3)** NVC(1) All A are C(1) All C are A(1)	**Some A not C(17)** NVC(7) Some C not A(1)	**NVC(23)** No A are C(2)	**NVC(13)** Some A not C(11) Some C not A(1)
Some not	**NVC(15)** Some A not C(8) Some C not A(1) Some A are C(1)	**NVC(16)** Some A not C(7) Some C are A(1) All C are A(1)	NVC(19) Some C not A(4) Some A not C(1) Some A are C(1)	**NVC(21)** Some A not C(3) Some A are C(1)

Figure A.1: Responses for the figure AB–BC: Experiment One

	All	Some	None	Some..not
All	**All C are A(17)** All A are C(6) NVC(1) Some A are C(1)	**NVC(15)** Some C are A(5) Some A are C(3) All C are A(1) Some C not A(1)	**No C are A(18)** **No A are C(6)**	**NVC(9)** Some C not A(12) Some C are A(2) Some A not C(1) Some A are C(1)
Some	**Some C are A(21)** **Some A are C(1)** Some C not A(2) NVC(1)	**NVC(22)** Some C are A(2) Some C not A(1)	**Some C not A(15)** NVC(6) Some C are A(3) No A are C(1)	**NVC(18)** Some C not A(4) Some A not C(2) Some C are A(1)
None	**Some A not C(3)** NVC(15) No A are C(4) No C are A(3)	**Some A not C(5)** NVC(12) Some C not A(5) Some A are C(2) No C are A(1)	**NVC(23)** Some A are C(1) No C are A(1)	**NVC(21)** Some C not A(2) Some A not C(2)
Some not	**NVC(4)** Some C not A(10) Some C are A(8) Some A not C(2) Some A are C(1)	**NVC(20)** Some C are A(3) Some A not C(2)	**NVC(13)** Some C not A(6) Some C are A(3) Some A not C(2) Some A are C(1)	NVC(21) Some C not A(3) Some A not C(1)

Figure A.2: Responses for the figure BA–CB: Experiment One

	All	Some	None	Some..not
All	**NVC(16)** All A are C(5) All C are A(3) Some C are A(1)	**NVC(13)** Some A are C(6) Some C are A(4) Some C not A(1) Some A not C(1)	**No C are A(12)** **No A are C(9)** NVC(4)	**Some A not C(11)** NVC(10) Some A are C(3) Some C not A(1)
Some	**NVC(15)** Some C are A(5) Some A not C(2) Some A are C(2) SOme C not A(1)	**NVC(19)** Some A are C(3) Some C are A(1) Some A not C(1) Some C not A(1)	**Some C not A(10)** NVC(9) Some A not C(3) No A are C(3)	**NVC(19)** Some C not A(5) Some A not C(1)
None	**No A are C(16)** **No C are A(7)** NVC(2)	**Some A not C(8)** NVC(15) No C are A(1) Some A are C(1)	**NVC(20)** No A are C(2) No C are A(2) Some C are A(1)	**NVC(20)** Some C not A(2) Some A not C(2) No C are A(1)
Some not	**Some C not A(7)** NVC(11) Some C are A(3) SOme A not C(3)	**NVC(20)** Some C not A(4) Some C are A(1)	**NVC(13)** Some C not A(9) Some A not C(1) Some C are A(1) Some A are C(1)	NVC(23) Some C are A(1) Some A not C(1)

Figure A.3: Responses for the figure AB–CB: Experiment One

	All	Some	None	Some..not
All	**Some A are C(8)** **Some C are A(5)** All A are C(6) All C are A(4) NVC(2)	**Some C are A(12)** **Some A are C(12)** Some A not C(1)	**Some C not A(8)** No A are C(7) No C are A(6) NVC(4)	**Some C not A(9)** Some A are C(6) NVC(4) Some C are A(4) Some A not C(2)
Some	**Some A are C(16)** **Some C are A(8)** Some C not A(1)	**NVC(20)** Some A not C(2) Some A are C(2) Some C are A(1)	**Some C not A(14)** NVC(8) Some C not A(1) Some A not C(1) No C are A(1)	**NVC(21)** Some C not A(3) Some A not C(1)
None	**Some A not C(9)** No A are C(8) NVC(6) No C are A(2)	**Some A not C(12)** NVC(12) No A are C(1)	**NVC(22)** No A are C(1) No C are A(1) Some A not C(1)	**NVC(10)** Some A not C(12) Some C not A(3)
Some not	**Some A not C(15)** NVC(4) Some A are C(2) Some C are A(2) Some C not A(2)	**NVC(21)** Some C not A(2) Some A not C(2)	**NVC(14)** Some C not A(7) Some A not C(4)	NVC(24) Some A not C(1)

Figure A.4: Responses for the figure BA–BC: Experiment One

Appendix B

Experiment Two

Correct responses are given in bold type, incorrect in light type. The number of subjects giving each response is noted in brackets. Not all responses are given as for many cases these are widely dispersed, but in cases where the error rate was greater than 4 out of 22 responses, the next one or two most frequent answers are given.

	All	Some	None	Some..not
All	**+A+B+C(22)**	**+A+B+C(21)**	**-A+B+C(16)** NVC(3)	**NVC(12)** -A+B+C(3)
Some	**NVC(9)** +A+B+C(13)	**NVC(18)**	**-A+B+C(17)** NVC(3)	**NVC(18)**
None	**+A+B-C(21)**	**+A+B-C(17)** NVC(3)	**-A+B-C(11)** NVC(6)	**NVC(16)** +A-B+C(2) +A+B-C(2)
Some not	**NVC(11)** +A+B-C(10)	**NVC(16)** +A+B-C(4)	**-A+B-C(13)** NVC(7)	**NVC(20)**

Figure B.1: Responses for the figure AB–BC: Experiment Two

	All	Some	None	Some..not
All	**+A+B+C(20)**	**NVC(11)** +A+B+C(11)	**-A+B+C(18)**	**NVC(9)** -A+B+C(11)
Some	**+A+B+C(20)**	**NVC(18)**	**-A+B+C(15)** NVC(2)	**NVC(16)** -A+B+C(5)
None	**+A+B-C(14)** NVC(5) -A-B+C(3)	**+A+B-C(13)** NVC(8)	**-A+B-C(8)** NVC(10)	**-A+B-C(13)** NVC(8)
Some not	**NVC(15)** +A+B-C(5)	**NVC(19)**	**NVC(17)** -A+B-C(3)	**NVC(19)**

Figure B.2: Responses for the figure BA–CB: Experiment Two

	All	Some	None	Some..not
All	**NVC(11)** +A+B+C(9)	**NVC(15)** +A+B+C(6)	**-A+B+C(16)** +A-B-C(3)	**+A-B-C(7)** NVC(7) -A+B+C(4)
Some	**NVC(16)** +A+B+C(6)	**NVC(18)**	**-A+B+C(15)** NVC(3)	**NVC(19)**
None	**+A+B-C(13)** -A-B+C(6)	**+A+B-C(7)** not counted(11)	**-A+B-C(10)** NVC(9)	**NVC(16)** +A-B-C(3)
Some not	**-A-B+C(11)** NVC(8)	**NVC(19)**	**NVC(19)**	**NVC(19)**

Figure B.3: Responses for the figure AB–CB: Experiment Two

	All	Some	None	Some..not
All	**+A+B+C(21)**	**+A+B+C(21)**	**-A+B+C(20)**	**-A+B+C(19)**
Some	**+A+B+C(20)**	**NVC(16)** **+A+B+C(6)**	**-A+B+C(15)** NVC(2) +A-B-C(2)	**NVC(18)**
None	**+A+B-C(20)**	**+A+B-C(19)**	**-A+B-C(14)** NVC(4)	**-A+B-C(16)** NVC(3)
Some not	**+A+B-C(18)**	**NVC(18)**	**-A+B-C(14)** NVC(4)	**NVC(20)**

Figure B.4: Responses for the figure BA–BC: Experiment Two

Appendix C

Conclusions, Cancellation and Double-Hatching

For each syllogism an indication is given as to whether the syllogism possesses a valid conclusion (VC or NVC), cancels (C or NC) or has double-hatching (DH or NDH).

Figure AB–BC	All	Some	None	Some..not
All	VC	VC	VC	NVC
	DH	DH	DH	NDH
	C	C	C	NC
Some	NVC	NVC	VC	NVC
	NDH	NDH	DH	NDH
	C	C	C	NC
None	VC	VC	NVC	NVC
	DH	DH	DH	NDH
	C	C	C	NC
Some	NVC	NVC	NVC	NVC
..not	DH	DH	DH	NDH
	C	C	C	NC

Figure C.1: Features of syllogisms with figure AB–BC

Figure BA–CB	All	Some	None	Some..not
All	VC	NVC	VC	NVC
	DH	NDH	DH	DH
	C	C	C	C
Some	VC	NVC	VC	NVC
	DH	NDH	DH	DH
	C	C	C	C
None	VC	VC	NVC	NVC
	DH	DH	DH	DH
	C	C	C	C
Some	NVC	NVC	NVC	NVC
..not	NDH	NDH	NDH	NDH
	NC	NC	NC	NC

Figure C.2: Features of syllogisms with figure BA–CB

Figure AB–CB	All	Some	None	Some..not
All	NVC	NVC	VC	VC
	NDH	NDH	DH	NDH
	C	C	C	NC
Some	NVC	NVC	VC	NVC
	NDH	NDH	DH	NDH
	C	C	C	NC
None	VC	VC	NVC	NVC
	DH	DH	DH	NDH
	C	C	C	NC
Some	VC	NVC	NVC	NVC
..not	NDH	NDH	NDH	NDH
	NC	NC	NC	C

Figure C.3: Features of syllogisms with figure AB–CB

Figure BA–BC	All	Some	None	Some..not
All	VC	VC	VC	VC
	DH	DH	DH	DH
	C	C	C	C
Some	VC	NVC	VC	NVC
	DH	NDH	DH	DH
	C	C	C	C
None	VC	VC	NVC	NVC
	DH	DH	DH	DH
	C	C	C	C
Some	VC	NVC	NVC	NVC
..not	DH	DH	DH	NDH
	C	C	C	C

Figure C.4: Features of syllogisms with figure BA–BC

Appendix D

Instructions to Subjects - Experiment One

The following instructions were given in written form to subjects undertaking Experiment One.

INSTRUCTIONS FOR THE SYLLOGISMS EXPERIMENT

To begin you must log-on to the practice session. To do this type *I AM CATH.SY99. The computer will chain the program, tell you to wait while it reads in the material and then will instruct you to press the [SPACE] bar when you are ready. This is the large bar at the very bottom of the keyboard. Do this and you will be presented with the first sentence of the syllogism - "All the Welsh are busdrivers". Once you have read this, press the [SPACE] bar once more for the second sentence - "All the busdrivers are moslems". Once you have finished reading this, press the [SPACE] bar again.

Now you will be asked if you can draw a conclusion from the syllogism you have just read. Think about the syllogism that you have encountered, can you find a relationship between the Welsh and the moslems that must always be true according to what you have read?

In this case you can; if all the Welsh are busdrivers and all th busdrivers are moslems, then it must be the case tht all the Welsh are moslems. But notice that it is not necessarily the case that all the moslems are Welsh - there may be moslems that are not busdrivers and are therefore not Welsh without violating the terms of the syllogism.

Having decided that there is a conclusion, press the [Y] key for Yes. You will then be presented with a menu and asked to draw that conclusion. The top item in the menu - "all" - is highlighted, if you press the [SPACE] bar you will see it is written up at the top of the screen. Press the [DELETE] key (at the bottom on the right hand side) and the "all" will disappear. Choose "all" as this is the beginning of the conclusion. Now press the [X] key and the highlighted item will move down one place. Press it again, until you reach the item "Welsh" and select this item. If the item "moslems" is above "Welsh" then use the [Z] key to move the highlighted item up, if below, use the [X] key again. Choose the item "moslems". You now see your conclusion

all
Welsh
moslems

written up at the top of the screen. This stands for the sentence "all the Welsh are moslems". To tell the computer that you have finished press the [RETURN] key (large button on the right). It will ask younif you are sure of your response - if you answer no (by pressing [N]) it will give you the chance to alter your response by using the [DELETE] key to remove your responses and choosing new ones as before. If you answer [Y] it will write your data and present you with the second menu.

This second menu is for you to reproduce as exactly as possible the syllogism that you have just read. The menu works exactly as before, but this time you will have to wirte up six items. The answer you produce should look like this:-

all
Welsh
busdrivers
all
busdrivers

moslems

This stands for "All the Welsh are busdrivers","All the busdrivers are moslems". It is important that you get the words in each sentence in the same order that they were in the original test, as has been shown above, so try and remember this too, but you can recall the two sentences of the syllogism in any order you like. If there are things you can't remember, then make a guess. The computer will make you choose six items and of course two of them will be from the first four items in the list, all; some; none; some..not, in first and fourth position in your answer and the other four will be from the last three items in the list, making up the rest of your answer.

Again, once you have finished, press [RETURN]. As before you will be given a chance to correct your answer, if you think you have got as close as you can, press [Y].

Now you are ready for the next syllogism. The procedure is as before, but let's look at the conclusions to the next three syllogisms in detail.

Read the next syllogism and stop when you are asked if you can draw a conclusion or not. In the situation described it might equally be the case that none of the vegetarians are Algerians, but equally they might all be, without any of them being teachers. The same goes for all or no Algerians being vegetarians and also for the cases where some are or where some are not. In other words the syllogism places no real restriction on the relationship between the vegetarians and the Algerians, so there is no conclusion that can be drawn as there is no relationship that must always be the case. Therefore you should press the [N] key. The computer will write No Valid Conclusion and then will give you the menu for recalling the syllogism.

WARNING - ONCE YOU HAVE PRESSED THE [Y] OR [N] KEY FOR THE QUESTION "CAN YOU DRAW A VALID CONCLUSION" THERE IS NO GOING BACK, SO THINK CAREFULLY BEFORE YOU ANSWER.

Complete the menu as before and then you are ready for the next text.

Read syllogism number three. Does this one have a valid conclusion? If none of the rugbyplayers are Ethiopians and some of the farmers are rugbyplayers then the farmers that are rugby players cannot be Ethiopians. You can therefore draw the conclusion that some of the farmers are not Ethiopians. This is a valid conclusion because it is always true - there is no situation in which all of the farmers could be Ethiopians and the relationships described by the syllogism also be true.

Carry on until you have finished the third syllogism and are ready to answer the fourth and final one. In this situation we are told that some of the French are teetotallers and that all of the butchers are French. Can there be a valid conclusion for this syllogism. One possible conclusion might be that some of the butchers are teetotallers. However, it could be the case that the butchers happen to be those French people that are not teetotallers, as "some" implies that there may be French people who not teetotallers as well as some that are. In that case could we say that some of the butchers are not teetotallers? Again it might be the case that all of the butchers happen to be those French people that are teetotallers. Similar arguments apply for conclusions using "all" or "none" for this syllogism. So there is no conclusion that must always be true, as there is always another possible situation that means that the conclusion does not hold, although the situation described by the syllogism is still captured. Therefore we must say that there is no valid conclusion for this syllogism.

Continue as before until you have finished recalling the syllogism you have just read. There are no more syllogisms to read but there is one more thing you have to do and that is to try to remember all the syllogisms you have read in the order you read them. The computer will ask you to remember syllogism number one and will give you a menu to choose from, as before. Use the menu just as you did before and then you will be given the menu for the second syllogism and so on until you have recalled all four. Then the computer will say that it is writing your data and after a short pause will come back with the prompt.

Now it is time for you to start the experiment proper. You log-on as before wirh *I AM CATH.SY? where ? is your subject number, which will have been given to you. There are two sessions each has 32 syllogisms. Once you have finished the first session you can get the second one by logging-on as before. Don't worry if you find it hard to remember the syllogisms, especially at the final recall stage - you can remember more than you think you can.

GOOD LUCK AND THANK YOU!

Appendix E

Instructions to Subjects - Experiment Two

The following instructions were given orally to subjects undertaking Experiment Two.

"This is an experiment about solving problems. I am going to ask you to try to solve some problems called syllogisms. Written here on this piece of paper is an example syllogism. You could imagine the situation described by these two sentences as being about people that have or don't have certain characteristics. To solve the syllogism you have to decide if there must exist a particular kind of person that either has or doesn't have certain characteristics. For example you might find that there is a person that must be German, a butcher and a football player. You might find for another syllogism that there is a person who is not German, but is still a football player and a butcher. For a lot of syllogisms it isn't possible to find a person that must always exist. This is because when you think about the kinds of person that could exist according to the syllogism, you can always imagine another situation where that person doesn't exist, but where the situation descrobed by the syllogism is still true.

Try the first syllogism and write down your answer below it."

(The subject tries the first syllogism and writes an answer).

If the subject is right: "Good, well done. You can now do the rest of the syllogisms. I'm afraid I can't help you, so just carry on and do your best until you've finished."

If the subject is wrong: "Can you explain why you think that is the answer? The right answer is actually X and that is because...(an explanation is given, only orally, couched in the same terms as the explanation above). Do you understand now? I can let you try one more syllogism if you like." (If desired the procedure can be repeated exactly as before, one more time only).

"Now carry on and try the rest of the syllogisms. I'm afraid I can't help you, so just carry on and do your best until you've finished."

"Have you finished? Thank you very much, this has been a great help."