

**WORKING MEMORY AND HUMAN REASONING:
AN INDIVIDUAL DIFFERENCES APPROACH**

by

ALISON JAYNE CAPON

A thesis submitted to the University of Plymouth

in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY

Department of Psychology

Faculty of Human Science

September 2000

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without the author's prior consent.

**WORKING MEMORY AND HUMAN REASONING:
AN INDIVIDUAL DIFFERENCES APPROACH**

ABSTRACT

Experiments 1-3 investigated the relationship between working memory and syllogistic and five-term series spatial inference. A secondary aim was to replicate the findings of Shah and Miyake (1996) who suggested the use of separate central resources of working memory for spatial and verbal ability. The correlational analysis showed that the complex verbal and spatial working memory span tasks were associated together and consistently predicted reasoning performance in both verbal and visual modalities. The confirmatory factor analysis showed that three factors best accounted for the data - a verbal, a spatial, and a general resource. All the span tasks and most of the reasoning tasks significantly and consistently loaded the general factor. Experiments 4-6 investigated the relationship between working memory and a range of reasoning tasks - identified as either propositional, spatial, or quantifiable tasks. These experiments were based on the work of Stanovich and West (1998) who found that a range of reasoning tasks were predicted by cognitive ability and a reasoner's thinking style. The correlational analysis showed that the complex verbal and spatial working memory span tasks were associated together and consistently predicted reasoning performance. Two clusters of reasoning task emerged from the correlational analysis - one cluster related to the propositional and simple spatial reasoning tasks, whilst the other related to the quantifiable and complex spatial reasoning tasks. The confirmatory factor analysis showed that four factors best accounted for the data - a verbal, a spatial, a general, and a thinking style resource. All the span tasks and the reasoning tasks loaded the general factor, and most of the reasoning tasks further loaded the thinking disposition factor. These results are discussed in light of models of working memory, theories of reasoning, and how to best characterise factor 3 (executive function) and factor 4 (thinking style) from the factor analysis.

LIST OF CONTENTS

CHAPTER 1 - A REVIEW OF WORKING MEMORY

1 . 1 INTRODUCTION TO THE REVIEW	1
1 . 1 . 1 What is Working Memory?	2
1 . 1 . 2 From Short-Term Memory to Working Memory	5
1 . 2 THE BADDELEY WORKING MEMORY MODEL	11
1 . 2 . 1 Introduction	11
1 . 2 . 2 The Phonological Loop	13
1 . 2 . 3 The Visuo-Spatial Sketch Pad	16
1 . 2 . 4 The Central Executive	19
1 . 2 . 5 Summary	24
1 . 3 INDIVIDUAL DIFFERENCES AND WORKING MEMORY	26
1 . 3 . 1 Introduction	26
1 . 3 . 2 Working Memory Capacity and Reading Comprehension	30
1 . 3 . 3 Working Memory Capacity and Cognitive Ability	39
1 . 3 . 4 Is Working Memory Capacity Domain-General or Domain-Specific?	43
1 . 4 CONCLUSION FOR THE WORKING MEMORY REVIEW	48

CHAPTER 2 - A REVIEW OF DEDUCTIVE REASONING

2 . 1 INTRODUCTION TO THE REVIEW	52
2 . 1 . 1 What is Deduction?	53

2 . 2 SYLLOGISTIC REASONING	54
2 . 2 . 1 Syllogisms	54
2 . 2 . 2 Syllogistic Reasoning Phenomenon	58
2 . 2 . 3 Theories of Syllogistic Reasoning	61
2 . 3 SPATIAL REASONING	81
2 . 3 . 1 Spatial Inference	81
2 . 3 . 2 Spatial Reasoning Phenomenon	84
2 . 3 . 3 Theories of Spatial Reasoning	93
2 . 4 REASONING AND WORKING MEMORY	98
2 . 4 . 1 Dual-Task Work	99
2 . 4 . 2 Difficulties of Interpreting Dual-Task Work	105
2 . 4 . 3 Individual Differences in Reasoning and Working Memory Capacity	108
2 . 4 . 4 Strategy Effects	110
2 . 5 RATIONALE FOR THESIS - REASONING WITH A LIMITED WORKING MEMORY CAPACITY – IDs WORK	114
CHAPTER 3 - WORKING MEMORY WITH SYLLOGISTIC AND SPATIAL INFERENCE	
3 . 1 GENERAL INTRODUCTION FOR EXPERIMENTS 1-3	119
3 . 2 EXPERIMENT 1	126
3 . 2 . 1 Aim	126
3 . 2 . 2 Method	127
3 . 2 . 3 Results	146

3 . 2 . 3 . 1	Descriptive Statistics	147
3 . 2 . 3 . 2	The Span Measures	150
3 . 2 . 3 . 3	The Span and Reasoning Measures	152
3 . 2 . 3 . 4	Further Analysis	155
3 . 2 . 4	Discussion	156
3 . 3	EXPERIMENT 2	160
3 . 3 . 1	Aim	160
3 . 3 . 2	Method	162
3 . 3 . 3	Results	166
3 . 3 . 3 . 1	Descriptive Statistics	167
3 . 3 . 3 . 2	The Span Measures	169
3 . 3 . 3 . 3	The Span and Reasoning Measures	170
3 . 3 . 3 . 4	Further Analysis	172
3 . 3 . 4	Discussion	174
3 . 4	EXPERIMENT 3	177
3 . 4 . 1	Aim	177
3 . 4 . 2	Method	178
3 . 4 . 3	Results	181
3 . 4 . 3 . 1	Descriptive Statistics	181
3 . 4 . 3 . 2	The Span Measures	183
3 . 4 . 3 . 3	The Span and Reasoning Measures	186
3 . 4 . 3 . 4	Further Analysis	188
3 . 4 . 4	Discussion	189

3 . 5 CONFIRMATORY FACTOR ANALYSIS - EXPERIMENTS 1, 2 AND 3 ...	192
3 . 5 . 1 Rationale for the Factor Analysis	192
3 . 5 . 2 Method and Results	194
3 . 5 . 3 Factor Analysis Discussion	203
3 . 6 GENERAL DISCUSSION FOR EXPERIMENTS 1-3	213
CHAPTER 4 - WORKING MEMORY AND DEDUCTIVE REASONING: COGNITIVE CAPACITY AND THINKING STYLE	
4 . 1 GENERAL INTRODUCTION FOR EXPERIMENTS 4 -6	220
4 . 2 EXPERIMENT 4	230
4 . 2 . 1 Aim	230
4 . 2 . 2 Method	231
4 . 2 . 3 Results	235
4 . 2 . 3 . 1 Descriptive Statistics	236
4 . 2 . 3 . 2 The Span Measures	238
4 . 2 . 3 . 3 The Span and Reasoning Measures	239
4 . 2 . 3 . 4 Further Analysis	240
4 . 2 . 4 Discussion	242
4 . 3 EXPERIMENT 5	245
4 . 3 . 1 Aim	245
4 . 3 . 2 Method	249
4 . 3 . 3 Results	258
4 . 3 . 3 . 1 Descriptive Statistics	259
4 . 3 . 3 . 2 The Span Measures	260

4 . 3 . 3 . 3	The Span and Reasoning Measures	262
4 . 3 . 3 . 4	The Reasoning Measures	264
4 . 3 . 3 . 5	Further Analysis	266
4 . 3 . 4	Discussion	268
4 . 4	EXPERIMENT 6	271
4 . 4 . 1	Aim	271
4 . 4 . 2	Method	273
4 . 4 . 3	Results	275
4 . 4 . 3 . 1	Descriptive Statistics	276
4 . 4 . 3 . 2	The Span Measures	278
4 . 4 . 3 . 3	The Span and Reasoning Measures	279
4 . 4 . 3 . 4	The Reasoning Measures	282
4 . 4 . 3 . 5	Further Analysis	284
4 . 4 . 4	Discussion	285
4 . 5	CONFIRMATORY FACTOR ANALYSIS - EXPERIMENTS 4, 5 AND 6 ...	290
4 . 5 . 1	Rationale for the Factor Analysis	290
4 . 5 . 2	Method and Results	292
4 . 5 . 3	Factor Analysis Discussion	307
4 . 6	GENERAL DISCUSSION FOR EXPERIMENTS 4-6	312
 CHAPTER 5 - GENERAL DISCUSSION FOR EXPERIMENTS 1-6		
5 . 1	INTRODUCTION	322

5 . 2 SUMMARY OF EXPERIMENTAL FINDINGS	322
5 . 3 MODELS OF WORKING MEMORY	327
5 . 4 INTELLIGENCE	332
5 . 5 REASONING THEORY	335
5 . 5 . 1 A Fourth Factor?	341
5 . 6 METHODOLOGICAL IMPLICATIONS	342
5 . 7 FUTURE DIRECTIONS	345
5 . 8 CONCLUSION	349
REFERENCES	352
APPENDICES A-B	381

LIST OF TABLES

CHAPTER 2

Table 2 . 1

The four moods of the syllogism 55

Table 2 . 2

The four figures of the syllogism 56

Table 2 . 3

The four orientations of two-dimensional spatial deductions 83

CHAPTER 3

Table 3 . 1

Descriptive statistics for Experiment 1 tasks (n=45) 148

Table 3 . 2

Experiment 1 correlations between the five span tasks 151

Table 3 . 3

Experiment 1 correlations between the span, reasoning and ability tasks ... 153

Table 3 . 4

Experiment 1 % of correct responses to one model and
multi model problems 156

Table 3 . 5

Descriptive statistics for Experiment 2 tasks (n=47) 167

Table 3 . 6

Experiment 2 correlations between the four span tasks 169

Table 3 . 7

Experiment 2 correlations between the span and reasoning tasks 171

Table 3 . 8	Experiment 2 % of correct responses to one model and multi model problems	173
Table 3 . 9	Descriptive statistics for Experiment 3 tasks (n=46)	182
Table 3 . 10	Experiment 3 correlations between the five span tasks	184
Table 3 . 11	Experiment 3 correlations between the span and reasoning tasks	186
Table 3 . 12	Experiment 3 % of correct responses to one model and multi model problems	188
Table 3 . 13	Set up of two-factor model for Experiments 1-3	195
Table 3 . 14	Goodness of fit summary (two-factor model, uncorrelated) - multiple group analysis of Experiments 1-3	196
Table 3 . 15	Goodness of fit summary (two-factor model, correlated) - multiple group analysis of Experiments 1-3	197
Table 3 . 16	Set up of orthogonal three-factor model for Experiments 1-3	198
Table 3 . 17	Goodness of fit summary (three-factor model, uncorrelated) - multiple group analysis of Experiments 1-3	199
Table 3 . 17A	Experiment 1 factor loadings for the three-factor uncorrelated model	199
Table 3 . 17B	Experiment 2 factor loadings for the three-factor uncorrelated model	200

Table 3 . 17C

Experiment 3 factor loadings for the three-factor uncorrelated model200

Table 3 . 18A

Goodness of fit summary (two-factor model, uncorrelated) -
analysis of Shah and Miyake's (1996) data 204

Table 3 . 18B

Shah and Miyake's (1996) data factor loadings,
a two-factor orthogonal model205

Table 3 . 19A

Goodness of fit summary (three-factor model, uncorrelated) -
re-analysis of Shah and Miyake's (1996) data206

Table 3 . 19B

Shah and Miyake's (1996) data factor loadings,
a three-factor orthogonal model 207

CHAPTER 4

Table 4 . 1

Descriptive statistics for Experiment 4 tasks (n=47) 236

Table 4 . 2

Experiment 4 correlations between the four span tasks 238

Table 4 . 3

Experiment 4 correlations between the span and reasoning tasks 239

Table 4 . 4

Experiment 4 % of correct responses to one model,
multi model and indeterminate problems 241

Table 4 . 5

The logical structure of conditional inferences 252

Table 4 . 6

Descriptive statistics for Experiment 5 tasks (n=51) 259

Table 4 . 7	
	Experiment 5 correlations between the two span tasks 261
Table 4 . 8	
	Experiment 5 correlations between the span and reasoning tasks 262
Table 4 . 9	
	Experiment 5 correlations between the reasoning tasks 264
Table 4 . 10	
	Experiment 5 % of correct responses to one model, multi model and invalid problems 266
Table 4 . 11	
	Descriptive statistics for Experiment 6 tasks (n=48) 276
Table 4 . 12	
	Experiment 6 correlations between the four span tasks 278
Table 4 . 13	
	Experiment 6 correlations between the span and reasoning tasks 280
Table 4 . 14	
	Experiment 6 correlations between the reasoning tasks 283
Table 4 . 15	
	Experiment 6 % of correct responses to one model, multi model and invalid problems 284
Table 4 . 16	
	Set up of the orthogonal three-factor model for Experiments 4-6 294
Table 4 . 17	
	Goodness of fit summary (three-factor model, uncorrelated) - multiple group analysis of Experiments 4-6 295
Table 4 . 17A	
	Experiment 4 factor loadings for the three-factor uncorrelated model 295

Table 4 . 17B

Experiment 5 factor loadings for the three-factor uncorrelated model 296

Table 4 . 17C

Experiment 6 factor loadings for the three-factor uncorrelated model 296

Table 4 . 18

Set up of orthogonal four-factor model for Experiments 4-6 300

Table 4 . 19

Goodness of fit summary (four-factor model, uncorrelated) -
multiple group analysis of Experiments 4-6 301

Table 4 . 19A

Experiment 4 factor loadings for the four-factor uncorrelated model 302

Table 4 . 19B

Experiment 5 factor loadings for the four-factor uncorrelated model 302

Table 4 . 19C

Experiment 6 factor loadings for the four-factor uncorrelated model 303

LIST OF FIGURES

CHAPTER 1

Figure 1 . 1

A simplified representation of the working memory model
(as Baddeley, 1997) 10

CHAPTER 2

Figure 2 . 1

A set of inference rules for a syllogistic problem (Example 2.6) 64

Figure 2 . 2

Mental model representation for a single model syllogistic problem
(Example 2 . 8) 70

Figure 2 . 3

Mental model representation for a multi model syllogistic problem
(Example 2 . 9) 71

Figure 2 . 4

Mental model representation for a single model spatial problem 96

Figure 2 . 5

Mental model representation for a multi model spatial problem 96

CHAPTER 3

Figure 3 . 1

Arrow span, representation of arrow 1 130

Figure 3 . 2

Arrow span, representation of arrow 2 131

Figure 3 . 3

Arrow span, recall grid for arrows 131

Figure 3 . 4	
	Letter span, representation of letter 1 135
Figure 3 . 5	
	Letter span, representation of letter 2 136
Figure 3 . 6	
	Letter span, recall grid for letters 136
Figure 3 . 7	
	TTT span, representation of winning line 1139
Figure 3 . 8	
	TTT span, representation of winning line 2139
Figure 3 . 9	
	TTT span, recall grid for winning lines 140
Figure 3 . 10	
	Four possible orientations for five-term series spatial problems 140
Figure 3 . 11	
	A one model spatial problem 141
Figure 3 . 12	
	A multi model spatial problem 141
Figure 3 . 13	
	A one model syllogistic problem 143
Figure 3 . 14	
	A multi model syllogistic problem 143

CHAPTER 4

Figure 4 . 1	
	A one model temporal problem 233

Figure 4 . 2

A multiple model temporal problem 234

Figure 4 . 3

An abstract problem251

Figure 4 . 4

A conditional problem 253

Figure 4 . 5

The six possible orders for three-term series problems 254

Figure 4 . 6

A valid three-term series problem 255

Figure 4 . 7

A multiple quantified problem 257

ACKNOWLEDGEMENTS

I would firstly like to acknowledge the contribution that my supervisory team made to this thesis. I wish to thank Doctor Simon Handley for his patience, understanding, time and effort. This thesis would not have been possible without his enthusiasm and support. I would also like to thank Doctor Ian Dennis who was invaluable to this project. Without his expert knowledge of statistics and, specifically, factor analytic methodology, the interpretations made here would have lacked clarity.

Secondly, I would like to acknowledge the Economic and Social Research Council for their award of a studentship. This research project would not have been possible without this grant.

Lastly, many thanks go out to all my friends and family. They have been a long-suffering and tolerant crew that stuck by me when times were hard. In particular, I would like to give a special thank you to my mum, Ann. Without her, there is no way I would have had the courage and determination (or financial means!) to see this project through.

AUTHOR'S DECLARATION

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award except the permitted Postgraduate Diploma in Psychological Research Methods and Statistics.

This study was financed with the aid of a studentship from the Economic and Social Research Council (Reference:) from October 1st 1996 to October 1st 1999.

A programme of advanced study was undertaken, which included readings in working memory and reasoning, guided by my supervisory team. I also completed the Postgraduate Diploma in Psychological Research Methods and Statistics. Professional conferences and seminars, in relation to my own area of work, were also attended and presented at.

Presentations and Conferences Attended

Capon, A. J., Handley, S. J. & Dennis, I. (1999). Working memory and individual differences in reasoning. BPS Annual (Cognitive) Conference, University of York.

Capon, A. J., Handley, S. J. & Dennis, I. (1999). Working memory and individual differences in reasoning. ESCOP, University of Ghent.

Capon, A. J., Handley, S. J. & Dennis, I. (1999). Working memory and individual differences in reasoning. Centre for Thinking and Language, University of Plymouth.

Capon, A. J. & Handley, S. J. (1998). Working memory and individual differences in syllogistic and spatial inference. BPS Annual (Cognitive) Conference, University of Bristol.

Capon, A. J. & Handley, S. J. (1997). Working memory and individual differences in syllogistic and spatial inference. BPS Annual (Cognitive) Conference, University of Bristol.

Signed.....

Date.....18-01-2001.....

CHAPTER 1 - A REVIEW OF WORKING MEMORY

1 . 1 INTRODUCTION TO THE REVIEW

This thesis addresses the question of whether working memory plays a role in the variation seen in performance on deductive reasoning tasks. Example 1.1 is an illustration of a complex deductive argument of the type to be used in the following chapters. This problem requires the ability to temporarily store and integrate information in order to correctly respond.

The square is to the right of the triangle

The circle is to the left of the triangle

The rectangle is in front of the circle

The crescent is in front of the triangle

What is the relation between the rectangle and the crescent?

Example 1 . 1

Example 1.1 involves the use of a temporary mental workspace in order to store intermediate steps of the task. For instance, the information in the second premise must be integrated with the information in the first premise. This information needs to be kept in mind whilst, additionally, integrating the information in the third and fourth premises. Only after total integration can a logical conclusion be made. Most accounts of the reasoning process rest on the assumption of limiting factors in performance. Consequently, the more capacity a reasoner has in order to store and process this type of problem, the better their performance will be. Thus the aim of the next five chapters is to investigate a range of reasoning tasks and measures of variations in peoples working memory capacity to test the relationship between the two.

The first chapter reviews the working memory literature. The development of a model of working memory will be summarised. Chapter 1 investigates the development of specific components of the working memory model and the methodology that has been used to draw out these claims. Chapter 1 then goes on to give an alternative approach to studying the relationship between working memory and reasoning, an individual differences methodology. The review will summarise what this approach has been able to uncover about the processing and storage components of working memory and their relationship to cognitive tasks. Chapter 2 introduces the reasoning literature. The chapter focuses on two types of reasoning task that are consistently investigated throughout the thesis - syllogistic and spatial inferences. These tasks will be discussed in light of reasoning theory and the link with working memory (through a dual-task methodology). Chapter 3 presents three experiments that investigate the relationship between a range of working memory capacity measures and syllogistic and spatial reasoning tasks. The second experimental chapter (Chapter 4) extends this to a wider range of reasoning tasks. Chapter 5 contains a general discussion in the context of the current literature and recommendations for future study.

1 . 1 . 1 What is Working Memory?

Gilhooly (1998) points out that information processing cannot occur without the reliance on both a long-term memory base and temporary mental workspace. We need to use knowledge of the world stored in our permanent memories (long-term), together with our understanding of the situation at hand (short-term), in order to perform even the simplest of tasks. Important information connected to the situation at hand must be successfully passed on to short-term memory. This allows us to store information long enough to be able to use it, and, for this reason, it is often referred to as 'working memory'.

There are many examples from everyday life in which we employ memory as a tool. A simple example would be an arithmetic problem. For instance, Example 1.2 shows a simple sum.

$$(3 * 6) + 2 = ?$$

Example 1 . 2

In order to correctly solve this problem, an individual must use a short-term memory buffer to store intermediate steps of the task. The first and second numbers must be multiplied together prior to the addition of the third number. Therefore, the sum of the first two numbers must be stored somewhere before being recalled to add to the third. Example 1.2 illustrates the necessity of a storage resource in order to perform the task and reach the correct conclusion.

Other instances of memory use from everyday life are more complex in nature. An illustration of one would be cooking a meal - creating the desired look, taste, and timing of a dish. An individual must follow steps in a recipe in the right order and at the right times. This requires them to process in parallel. They must remember many parts of the recipe at the same time in order to create the desired effect. Even buying the ingredients for the meal would involve the individual recalling what they must buy, what they already have at home, and when they must have completed the shop run by. Despite the time scale involved with cooking a meal (elongated in time), it involves formulating and retaining a plan and keeping track of its execution.

These examples all require the ability to temporarily store in mind information related to the specific task. The examples illustrate that, as complexity increases, then the need for memory space increases. Therefore some tasks are more difficult than others. It might be

the case that limitations in the ability to hold information in short-term memory constrain individual performance of the task. So as the complexity of an individual's plan increases, and the activities related to that plan increase (processing), then they are more likely to lose track of the goal. For instance, in the example of cooking a meal, one might go to the shops with a mental list of ingredients but get way laid by a trip to the bank which in turn causes you to forget something needed for the recipe. Information needed to execute a plan can be lost due to interference from other sources. Thus working memory holds the plan during execution and all the intervening steps. Working memory is a temporary memory work space with limited capacity.

Working memory has been implicated as playing a major role in cognitive processing (e.g., Newell & Simon, 1972; Sanford, 1985). Theorists suggest that the limitation in working memory capacity is a factor contributing to difficulty in problem solving (e.g., Johnson-Laird, 1983). Daneman and Carpenter (1980) point out that information processing activities compete for this limited capacity and consequently, a task that has heavy processing requirements should decrease the amount of additional information that can be maintained. If working memory is a limiting factor on information processing activities, it may be implicated in patterns of individual differences across cognitive tasks. This thesis addresses these questions. Individual differences in a range of cognitive tasks that require storage and maintenance will be investigated in terms of their reliance on a short term resource.

There have been two major approaches to the study of working memory and cognition. The first approach is the traditional one, set in the experimental vein (e.g., Baddeley & Hitch, 1974). The strategy underlying these experiments rests on the assumption that working

memory has limited capacity. If a substantial amount of that capacity is taken up by a supplementary task, then performance will deteriorate substantially. An alternative to this approach is one based upon individual differences (e.g., Shah & Miyake, 1996). Individuals can be given a range of tasks (e.g., memory and reasoning measures) and a correlational approach can be used to identify which aspects of cognitive processing appear to be most closely associated with the measures of memory capacity. The objective of this chapter is to provide an integrated account of both these approaches in order to illuminate the role of working memory in cognitive tasks. The review begins with a summary of the precursors of the present working memory model.

1 . 1 . 2 From Short-Term Memory to Working Memory

The study of short-term memory has been popular since the 1950s. In the 1950s many studies seemed to point to the need to separate long and short-term memory (e.g., Brown, 1958; Peterson & Peterson, 1959). Following this were years of experimentation aimed at developing a model of short-term memory that accounted for all the experimental findings. Section 1.1.2 summarises the main findings of the research that led to the present model of working memory.

There have been many diverse models of short-term memory (Hunter, 1957; Newell & Simon, 1972). Most of the earlier conceptions of short-term memory had a good deal in common, and approximated more or less to the one proposed by Atkinson and Shiffrin (1968, 1971). This model was referred to as the multi-store or modal model and assumed three major components to memory. Firstly, there was a bank of relatively peripheral sensory stores, capable of parallel processing. These sensory buffers fed information into a short-term store which was responsible for encoding the incoming material to stop it being

forgotten. This was a necessary intermediate stage in the process of transferring information to the third component, the long-term memory store. The probability of learning anything permanently was assumed to be a direct function of the amount of time an item resided in the short-term store or rehearsal buffer.

There were numerous research studies carried out to evaluate and strengthen support for the modal model (e.g., Glanzer & Cunitz, 1966; Peterson & Peterson, 1959; Conrad, 1964; Milner, 1971). The evidence that follows relates to arguments that provide support for a separation of short-term and long-term memory. The modal model could explain many of the effects found empirically, from neuropsychological evidence to serial position curves.

One line of evidence in support of the modal model came from the study of brain-damaged patients. If a long-term and short-term store are distinct systems, then evidence that storage in one system is intact, whilst storage in the other is damaged would provide support to the modal model. Milner (1966) worked with a patient named H.M., who was incapable of forming new memories, but could learn new skills. H.M. performed very poorly on standard tests of verbal learning (long-term memory), but his memory span (short-term memory) was normal. Baddeley and Warrington (1970) studied a range of amnesic patients and found that their performance was intact on numerous tasks assumed to rely on a short-term store. These tasks included memory span and the recency effect in free recall. These patients also showed grossly defective long-term memory performance. This evidence suggested that patients could have normal short-term store coupled with a defective long-term store. The patterns of results seemed to suggest that those patients were incapable of transferring information from the short-term to the long-term store. Thus, these studies

appeared to support the distinction between short-term and long-term memory, assuming the need for a separate short-term store.

At about the same time, Shallice and Warrington (1970) reported the case of a patient K.F. with exactly the opposite pattern of defects, namely grossly defective short-term memory coupled with unimpaired long-term learning. This patient was mildly dysphasic and marginally dyslexic, but overall had no general intellectual impairment. His digit span, however, was grossly impaired, being limited to about two items with auditory presentation, but was reliably better when the presentation was visual. His long-term learning ability, as measured by performance on a paired-associate task, was quite normal, suggesting a very specific defect to an auditory-verbal short-term storage system. All these neurological studies pointed to the need to separate short-term and long-term stores.

In studies on 'normal' participants, when they were presented with a free recall test of a list of words, performance on the task produces a characteristic serial position curve (e.g., Glanzer & Cunitz, 1966). Immediate recall of initial items are good (primacy effect), middle items are recalled less well, whilst the last few items are recalled very well (recency effect). After a brief filled delay (rehearsal-preventing task), however, the recency effect disappears, while performance on earlier items in the curve is relatively unaffected by the delay. It should be noted that a variety of variables have different effects on the early and middle parts of the list as opposed to the effects on the end of the list. Long-term learning was assumed to depend on holding information in a temporary short-term store until it was transferred to a long-term store. The probability of learning was assumed to be a direct function of the amount of time an item resided in the short-term store. The interpretation suggested that the recency items are held in a temporary short-term store, while earlier items

are recalled from long-term memory. This is one of the strongest arguments for the separation of long-term and short-term stores due to the fact that free recall tasks appear to have two separable and quite different components.

Another argument for the separation of short-term and long-term stores comes from evidence suggesting that the short-term store has limited capacity, but relatively rapid input and retrieval. The long-term store, on the other hand, has an enormous capacity, but tends to be slower to register information and retrieve it. Arguments for the limited capacity of the short-term store come principally from tasks such as the digit span, in which the participant appears to be able to hold about seven chunks of information (Miller, 1956).

More evidence for the modal model came from examining the type of coding each memory store relied upon. Sachs (1967) investigated this by studying the retention of prose passages. Sachs (1967) found that provided the sentence was tested immediately, participants were relatively good at detecting all changes, whether in meaning or syntax. After one or more intervening sentences, however, the participant's capacity to remember the syntactic and surface features of the prose dropped dramatically, while retention of the meaning remained good. The generalisation here was that short-term storage relies upon phonological coding while long-term memory is more influenced by meaning.

However, there were problems with the modal model, and it suffered as an account of the memory system in the following ways. The model indicated that patients with short-term memory deficits should also have problems in long-term learning; such deficits were not apparent (Shallice & Warrington, 1970). The assumption that maintaining an item in a short-term store would ensure its transfer to a long-term store proved to be poorly supported

(Morton, 1967). The existence of long-term recency effects, and the absence of a disruption of recency in free recall by a concurrent memory span task were also both inconsistent with the modal model's interpretation of recency (Baddeley & Hitch, 1977). Finally, the assumption that the short-term store relies on acoustic coding, and the long-term store on semantic, was clearly over-simplified (e.g., Baddeley & Levy, 1971).

Baddeley's working memory approach, beginning with Baddeley and Hitch (1974), sought to address some of the problems of the modal model. They were interested in the role of short-term memory in general cognition. They did not reject the modal model's view of short-term memory, as rehearsing incoming information for transfer to long-term memory, but claimed that it had more important roles beyond this. The original working memory model has been itself modified and elaborated (Baddeley, 1981, 1986) and in its 1986 form consists of a central executive, at the top of the hierarchy, controlling or directing the activities of two other components, the phonological loop and the visuo-spatial sketch pad.

Baddeley and Hitch (1974) identified active (processing) and passive (storage) components of a limited working memory capacity important in performing higher level cognitive tasks. They proposed that part of the limited capacity in the system (i.e., the central executive CE) was used for processing incoming information and internally generated information. The remaining resources (i.e., the phonological loop PL and visuo-spatial sketch pad VSSP) were proposed to be used for storage of the products of that processing. Figure 1.1 shows a simple representation of this working memory model.

The PL is assumed to comprise two components, a phonological store that is capable of holding speech-based information, and an articulatory control process based on inner

speech. Memory traces within the phonological store are assumed to fade after about one-and-a-half seconds. The memory trace can however be refreshed by a process of reading off the trace into the articulatory control process, which then feeds back into the store. This is the process underlying subvocal rehearsal. The articulatory control process is also capable of taking written material, converting it into a phonological code, and registering it in the phonological store. The VSSP has been conceived as a spatial analogue of the PL. Like the loop, it can be fed either directly through perception, in this case visual perception, or indirectly, in this case through the generation of a visual image. While processing information, the CE is considered a controller, selecting the most advantageous strategies for integrating information from several different sources. Thus, the CE is a central processor, requiring two peripheral or slave sub-systems (PL and VSSP) to store the information being processed.

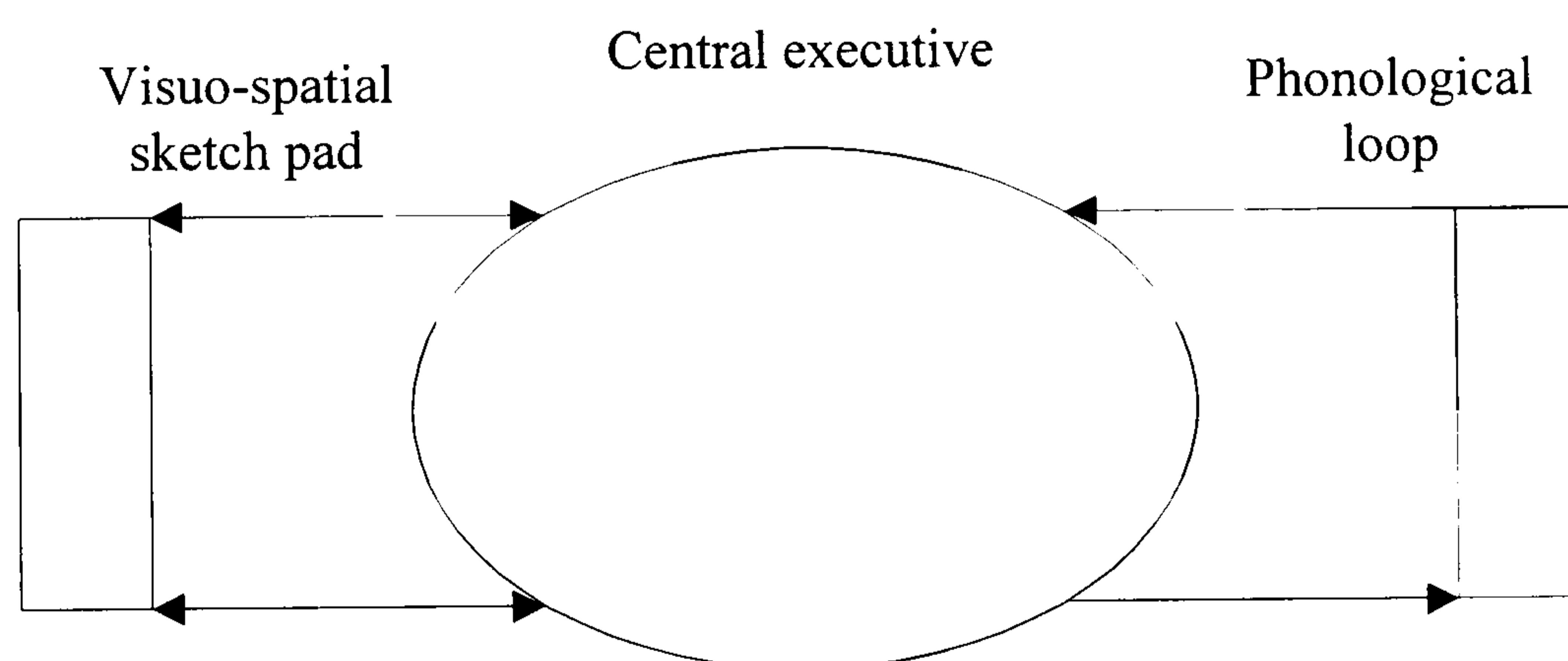


Figure 1 . 1 - A simplified representation of the working memory model (as Baddeley, 1997)

The working memory model refers to a system for the temporary holding and manipulation of information during the performance of a range of cognitive tasks such as comprehension, learning and reasoning. The initial research indicated that working memory was involved in temporary storage for a wide range of information processing skills (Baddeley, 1986, 1997).

The working memory model was testable and productive. Section 1.2 systematically explores this model by examining the experimental evidence for each of the three components.

1 . 2 THE BADDELEY WORKING MEMORY MODEL

1 . 2 . 1 Introduction

This section summarises the many research studies that have been carried out to systematically test the characteristics of each component of the working memory model.

Firstly the main choice of experimental methodology will be explained - the secondary task or dual-task methodology. This will be followed by the work, using this approach, that has tested the attributes of each working memory component - the PL, the VSSP and the CE.

The strategy underlying research in this field rests on the assumption that working memory has limited capacity. If a substantial amount of that capacity is taken up by a supplementary task, then performance on a primary task (such as reading comprehension) should deteriorate substantially. Such deterioration should be shown across a range of cognitive tasks even though such tasks do not have an obvious short-term memory component. An obvious short-term memory component can be seen easily in recall tasks, or memory span tasks. This component is less obvious in tasks such as comprehension or problem-solving.

In the case of processing verbal materials, there is abundant evidence that effects of working memory load does occur. As an example of this, Baddeley, Eldridge, Lewis and Thomson (1984) showed that a concurrent task involving retaining sequences of digits consistently slowed down the verification of simple sentences. Retaining a sequence of digits must rely

on a short-term store in order to be able to recall them at a later date. If performance on verifying sentences is degraded, when carried out concurrently with this, the standard assumption has been that the two tasks are competing for a shared resource.

Working memory has been studied mainly by the use of such concurrent or interference (dual-) task methodology. For example, Baddeley and Hitch (1974) presented participants with visual sequences of three to six digits whilst they listened to a prose passage. Statistical analysis indicated that comprehension was not significantly impaired by a concurrent load of three digits, but was reliably impaired by a six digit load. Earlier experiments (Hitch & Baddeley, 1976) demonstrated that participants are capable of holding a substantial number of digits in short-term store at the same time as performing the complex cognitive operations involved in learning, comprehension, and reasoning. In terms of latency however, a clear effect of concurrent load on a range of cognitive tasks was found. That is, the speed of performing a primary task involving learning, comprehension or reasoning, was found to be poorer when carried out with a digit load.

A special case of dual-tasking involves the use of articulatory suppression (repeating out loud the word 'the'). Concurrent articulation has been utilised to block the use of the PL. Thus, if concurrent articulation produces poorer performance on another, concurrent, task, then it might be inferred that this second task also uses the PL (Eysenck, 1986). What follows is a summary of the experimental evidence relating to the three components of the working memory model.

1 . 2 . 2 The Phonological Loop

Baddeley and Hitch (1974) suggest that the PL system comprises a phonological store and an articulatory control process. In its initial formulation, the PL was assumed to function like a tape loop of limited duration. The loop was assumed to hold about 1.5 sec of speech-based material in temporary storage and to be capable of maintaining this by means of articulatory rehearsal. This simple notion of the PL was able to account for a wide range of results. These phenomena are often thought of as the rudiments of verbal short-term memory, and collectively they provide converging evidence for the characteristics of the PL.

The phonological similarity effect indicated a poor immediate memory for phonologically similar items (Conrad & Hull, 1964; Baddeley, 1966), and it is the best known phenomena associated with verbal short-term memory. Specifically, this effect refers to the fact that recall of a series of words or letters is more difficult when the words or letters for recall sound alike. This could be explained by assuming that the PL is based on phonological coding. Similar items have easily confusable codes, leading to impaired performance. Thus, the phonological similarity effect arises because items contained within the phonological store will become confused when they are phonologically similar to one another. As verbal sequences are held in this store whether they are heard or read, the phonological similarity effect appears for both visually presented and auditorily presented material.

Verbal serial recall is also disrupted by the concurrent presentation of irrelevant speech (Colle & Welsh, 1976). This disruptive effect is even greater when the irrelevant speech comprises words that are phonologically similar to the words for recall. Salame and

Baddeley (1982) gave participants a sequence of visually presented digits accompanied by either a spoken word, nonsense syllable, or with no distraction. Participants showed an increase in errors when attempting to ignore either words or nonsense syllables. An explanation of this finding was that participants were subvocally articulating the visual digits, and were therefore using their phonological store. The effect of irrelevant speech arises from the fact that heard speech directly accesses the phonological store, thereby disrupting its current contents.

A further phenomenon in verbal serial recall is that sequences of long words can be recalled rather less well than sequences of short words. Baddeley, Thomson and Buchanan (1975) investigated this word length effect and found that memory span for short words is greater than for long. This could be explained by assuming that short words are better remembered simply because they can be articulated more rapidly. Spoken duration was seen as critical in explaining the articulatory control process. The word length effect was interpreted as reflecting the operation of subvocal rehearsal. Words that take a longer time to say are more difficult to rehearse, and therefore will be less well retained.

Finally, retaining a verbal sequence is dramatically impaired when participants are simultaneously required to repeat aloud an irrelevant speech sound such as 'the, the, the' (Levy, 1971, 1975). Murray (1968) found that this suppression impaired performance on a memory task. Articulatory suppression also removed the word length effect for visual and auditory presentation, and removed the phonological similarity effect but only when the list for recall was presented visually (Baddeley, Lewis & Vallar, 1984). The model accounted for the word length effect by suggesting that articulatory suppression blocked the operation of subvocal rehearsal, thereby removing the potential for rehearsing words of different

length. The model accounted for the phonological similarity effect by suggesting that with articulatory suppression, visually presented material could not be transferred into the phonological store and the phonological similarity effect disappeared. As auditory presentation results in direct input into the phonological store, the potential for phonological confusion remained even with articulatory suppression.

As stated in Section 1.1.2, the modal model was unable to fully explain why patients could have a defective short-term memory but normal long term learning. The modal model theorists suggested that long term learning depended crucially upon a short-term store. Thus, if this short-term store were damaged, then it would follow that long term learning would also suffer. Atkinson and Shiffrin (1971) suggested that those patients suffering from a defective short-term memory should also suffer a failure of long-term memory. This was found to be unsupported in many cases (Shallice & Warrington, 1970). This type of neuropsychological evidence has been applied to a working memory model, and is briefly described below.

The working memory hypothesis would argue that patients in this position had suffered damage to one of the subsidiary slave systems of working memory rather than to the CE. A wide variety of neuropsychological literature supports the notion that there are distinct working memory systems for spatial and verbal information (Milner, 1971; Paivio, 1971). For instance, if one assumes a deficit in the phonological store, then this is able to explain both their impaired memory span and their comparatively normal cognitive performance on other tasks such as long-term verbal learning, where one might expect semantic coding to be more important than phonological. Vallar and Baddeley (1984) studied a patient named P.V., who had a very pure and specific deficit in auditory short-term memory performance.

With the exception of this specific deficit, P.V. appeared to be intellectually entirely normal, with a high level of verbal and performance I.Q., excellent long-term memory, and no apparent problems of speech or language (Basso, Spinnler, Vallar & Zanolio, 1982). It was suggested that P.V.'s deficit was an impairment, though not complete disruption, of the phonological storage component of the PL.

The findings summarised above describe the PL as comprising a phonological store and an articulatory control process. Due to the effects found, research supports the notion of information being stored in a phonological, or as a verbally based code. The findings informed researchers of the structure of a working memory model and PL component. The PL model appears to offer an adequate and useful account of the available evidence.

1 . 2 . 3 The Visuo-Spatial Sketch Pad

Normal memory and neuropsychological research suggests the need to separate visuo-spatial and verbal processes of memory (De Renzi, 1982; Milner, 1971; Paivio, 1971). Thus Baddeley's (1983) concept of a VSSP is in line with other fields of research. He conceived the VSSP as analogous to the PL, providing temporary storage and maintenance of visuo-spatial rather than verbal material. There appears to be good evidence for the occurrence of a temporary visuo-spatial sketch pad that is capable of retaining and manipulating images, and is susceptible to disruption by concurrent spatial processing (Baddeley, Grant, Wight & Thomson, 1975). This simple notion of the VSSP was able to account for a range of dual-task results.

Baddeley, Grant, Wight and Thomson (1975) used a tracking task that required participants to keep a stylus on a spot of light, the performance being measured by the total amount of

time on the target. Tracking was combined with a paired-associate learning task. Participants were required to associate adjectives with nouns for highly imageable pairs (strawberry-ripe) and for abstract pairs (gratitude-infinite) of words. Participants were required to learn and recall these lists either concurrent with pursuit tracking or in a control condition. If the VSSP utilises images, then using concurrent tracking might reduce the advantage seen on an imageable list learning task. Baddeley et al. (1975) showed that when either an imagery or an abstract learning task was performed concurrently with visual pursuit tracking, the former had a much greater disruptive influence on tracking performance. This could be explained by a visuo-spatial subsystem of working memory responsible for imagery, which is implicated in both the use of imagery in recall tasks and in spatial tracking.

Baddeley and Lieberman (1980) investigated whether the subsystem responsible for imagery was visual or spatial in nature. They found that a spatial tracking task with no visual component had an effect similar to that of visual pursuit tracking on performance of visual and verbal memory tests. However, there was no effect on imagery if the visual task had no spatial component. This added strength to the notion of a subsystem based upon spatial coding.

Phillips and Christie (1977) postulated that if a separate visuo-spatial subsystem exists, then it should be possible to devise a task that places heavy demands upon this subsystem but minimal demands upon the CE. Farmer, Berman and Fletcher (1986) took this proposal and used spatial suppression (sequential tapping of four targets) as the concurrent task, which should place little demand on central processes. Results indicated that sequential tapping did interfere with performance on a spatial, but not a verbal, reasoning task.

Neuropsychological evidence has also been interpreted as suggesting the need for a storage system that relies on visuo-spatial information. De Renzi and Spinnler (1967) studied a group of patients all of whom suffered from colour-blindness as a result of cortical damage. Such patients report the inability to answer such questions as, 'What colour is the banana?' and are not able to select the appropriate colour from a set of crayons. The converse pattern of visual processing deficit also occurs, with the patient able to localise an object accurately, but not able to recognise it, a deficit known as agnosia (Humphreys & Riddoch, 1987). It has been argued that two separate locations exist for visual imagery. One, which is primarily concerned with visual aspects of imagery is dependent on the occipital lobes. The second system is one that depends principally on spatial coding, and is dependent on the functioning of the parietal lobes. Thus imagery has related but separable visual and spatial components.

The visual-spatial distinction has been fleshed out more recently. Logie (1995) suggested that a dissociation could be made between a capacity for retaining visual patterns (the visual cache) and that for retaining sequences of movements (the inner scribe). Through dual-task studies, it has been demonstrated that the retention of spatial patterns, but not visual information, was disrupted by arm movements during a retention interval. On the other hand, retention of visual information, but not spatial patterns, was disrupted by a visual interference task inserted between presentation and retrieval (e.g., Logie & Marchetti, 1991). Thus the VSSP has been fractionated, as was the PL, into a passive visual cache and an active spatially based rehearsal system called the inner scribe.

Baddeley (1986) points out the analogy between the PL and the VSSP. Both systems appear to take advantage of an essentially passive perceptual input store. In both cases, the

problem of coping with rapid decay from the store appears to have been solved by an active control process based on a response system, articulation in the case of the PL and eye movement in the case of the VSSP. These allow the transformation of a passive perceptual store into an active memory system that enables the organism to take information out of the relevant input store and to feed it back, thereby continuously refreshing the trace and minimising forgetting.

So far the findings from research are in relation to the hypothesised PL and the VSSP.

These two components of the working memory system have been conceived of as relatively passive in nature. Remembering a list of words would be considered a passive act - reliant on a system that could store them in order to recall them at a later date (PL). Remembering a matrix of randomly filled cells would be considered a passive act - reliant on a system that could store them in order to recall them at a later date (VSSP). Evidence has strengthened the view that there need to be passive storage systems for phonological and visuo-spatial coding in memory. These two systems could deal with simple processing, such as the passive storage of words, or the use of simple imagery in recall. However, alone, they could not deal with the problems illustrated in Examples 1.1 and 1.2. A more active system, capable of processing, storing and co-ordinating activities, would be required in these instances, and Baddeley identified it as the CE.

1 . 2 . 4 The Central Executive

The CE is the most crucial component of working memory. This is because it is postulated to be responsible for the selection, initiation, and termination of processing routines (e.g., encoding, storing, retrieving). The term CE was also used to imply some type of supervisor, capable of selecting strategies and integrating information from several different sources.

The evidence put forward to substantiate the claim for a CE in the former years was less well formulated than the evidence to support both slave systems of working memory.

Baddeley (1986, 1990) equates the CE with the supervisory attentional system (SAS) described by Norman and Shallice (1980) and by Shallice (1982). According to Shallice (1982), the SAS is a limited capacity system and is used for a variety of purposes. It is involved in tasks that need planning or decision making, as well as for trouble shooting in situations in which the automatic processes appear to be running into difficulty. The SAS would be used in novel situations, and in situations that are dangerous or technically difficult. Norman and Shallice (1980) apply their model to a number of phenomena, including those situations in which participants are able to perform two apparently demanding tasks simultaneously. The model assumed the need to resolve conflicts so as not to overload the system, and that overall supervisory control was exercised by the SAS system.

The function of the CE/SAS can perhaps be most clearly illustrated in studies of patients suffering from frontal lobe syndrome. Rylander (1939) characterised the deficit as,

'disturbed attention, increased distractability, a difficulty in grasping the whole of a complicated state of affairs ... well able to work along old routine lines ... (but) ... cannot learn to master new types of task, in new situations ... (the patient is) ... at a loss'.

One of the tasks claimed to be dependent on the frontal lobes is the Wisconsin Card Sorting Test (WCST). This involves presenting the participant with a pack of cards, on each of which is a pattern made up from various numbers of shapes which vary in colour, size and surround. The patient is instructed to sort the cards into piles on the basis of some rule.

Once a rule has been acquired it is changed, until all the cards have been sorted on the basis of all the rules. Patients with frontal lobe damage tend to learn the first rule but appear to be unable to escape that rule, with a very high proportion of their errors being based on the old rule. Shallice (1982) explains that since the SAS system is not functioning properly, the patient is at the mercy of the currently active schemata. If a situation exists in which one schema is clearly dominant, then it will continue to dominate, leading to perseveration. Although the SAS research is not directly related to the distinct research area investigating CE function, it is of some use in summarising some common characteristics between the two. Understanding the function of the frontal lobes seems likely to offer fruitful challenges of the potential applicability to the concept of working memory.

Baddeley (1986) suggests a crucial role of the CE in understanding the processes of ageing, dementia, the role of the frontal lobes, and its role in intelligence. If the CE is a system capable of attentional control, selecting and operating control processes, then damage to this system would have far reaching effects on any information processing activity.

As with the PL and VSSP components of working memory, secondary task methodology has been used to investigate the characteristics of the CE. If learning and/or retrieval were limited by the amount of available attentional capacity, then requiring a participant to perform a second attention-demanding task during learning or retrieval should cause performance impairment. Hitch and Baddeley (1976) used a digit load as a secondary task and found that a two-digit load did not affect performance on a verification task. This was presumably because the load placed only minor demands on the PL component. A longer series of digits would occupy more of the available capacity up to a point at which capacity is exceeded and errors creep in. Indeed, additional memory load substantially interfered

with the verification task (a six-digit load). The concurrent load forced the participant to divide his/her attention between the separate processes of verification and short-term storage, hence the interference in this situation. The evidence suggested that a limited capacity executive processing system must be brought into play when dealing with larger memory loads.

Many experiments have found that a decrement occurs in learning when carried out concurrently with a six-digit load. These decrements in learning can be seen across a range of tasks, including paired-associate learning (Baddeley, Eldridge, Lewis & Thomson, 1984) and the retention of prose (Baddeley & Hitch, 1974). A secondary task that diverts attention from learning will tend to impair performance, with the degree of impairment increasing as the degree of diversion increases (Murdock, 1965).

An alternative secondary task assumed to tax CE resources of working memory is random number or letter generation. Random generation methods require participants to generate sequences of numbers or letters, making the order as random as possible (Baddeley, 1966). Random number generation is a more common method than random letter generation, although the same points apply to both. A person's success in carrying out this task can be measured in three ways. The first is to count the frequency with which each number is emitted - the more biased the distribution, the more redundant the output. The second measure involves scoring the number of different number pairs generated in a hundred responses. The third measure uses the number of stereotyped number pairs such as 1 and 2, or 7 and 8 that follow the chronological sequence.

Baddeley (1966) investigated random generation as a secondary task methodology. He combined choice reaction time and random letter generation. The participant's primary task was to sort playing cards into 1, 2, 4, or 8 categories. The participant was required to make one response every two seconds, a rate that allowed perfect performance even in the most demanding 8 choice condition. It was expected that the more difficult the sorting response, the more of their available supervisory capacity should be absorbed, given that the card sorting task was not highly over-learned, and the less random their output of letters should become. This indeed was found and the results were highly consistent and broadly fitted the conceptualisation that the process of random generation depends on a system of limited informational capacity - hence the more rapid the rate, the less random the output, and the larger the set of selection alternatives, the slower the maximum generation rate.

Baddeley (1966) suggested that random generation requires a mechanism with a clear selection process that behaves lawfully and has limited capacity. Random generation disrupts the operation of the CE by its demands for the constant switching of retrieval plans, involving the generation of a set of digits which correspond to randomness. Baddeley (1996) suggested that an adequate model of random generation is still needed and concluded that it is unclear to what extent the load imposed by generation stems from 1) the need to switch strategies, 2) the problem of accessing new strategies, or 3) the monitoring of the response output. Thus random generation appears to involve many processes assumed to rely on executive function. These operations involve retrieving and control processes which map onto the function of the CE as proposed by Baddeley and Hitch (1974).

Shiffrin and Schneider (1977) looked at the concept of automaticity, which adds to the justification to use random generation as a research tool. This concept shows that the repeated pairing of a specific stimulus with the same response will gradually reduce the

attentional demand of responding to the stimulus. Random generation could be seen as the opposite end of this continuum, where the aim is to generate a response that is minimally associated with what went before. The patterns of results produced by Baddeley (1966) are consistent with the idea that, even with practice, random generation continues to place heavy demands on the underlying mechanism of control processes.

The CE offers the mechanism for control processes in working memory, including the coordination of the subsidiary memory systems, the control of encoding and retrieval strategies, the switching of attention, and the mental manipulation of material held in the slave systems (Baddeley, 1996). The CE itself was originally conceptualised as being equipped with a supplementary storage capacity. The detailed nature of the processes attributed to the CE were derived empirically and have been briefly explained above. The organisation of these processes remains an open question and is the subject of ongoing empirical exploration.

1 . 2 . 5 Summary

The working memory model outlined above contains three components - the PL, the VSSP, and the CE. Dual tasks have been used to load different components of the memory system in order to see the effects of these when carried out concurrently with primary cognitive tasks. This methodology allowed the investigation of the importance of the working memory system for a variety of information processing activities. Articulatory suppression has been used to disrupt the PL component of working memory. This was shown to rely on the encoding of speech based material maintained by rehearsal. Tracking tasks were used to disrupt the VSSP component of working memory. These have been shown to rely on spatial processes that would disrupt any competing spatial material. Random generation tasks have

been used to disrupt the CE component of working memory. These tasks are assumed to rely on selection and retrieval processes. If performance on any of these secondary tasks disrupted performance on a primary cognitive task, they were said to share a common resource.

Neuropsychological evidence has also been forwarded to support the idea of the working memory model. Selective impairment on some tasks, and not others, suggested that part of the short-term system could be affected by damage, whilst leaving other areas unimpaired. One could assume a deficit in a component of the short-term memory system which was able to explain why patients were impaired on particular activities, but not others. The neuropsychological evidence could be interpreted as supporting the existence of the three subsystems of working memory (PL, VSSP and CE).

The model, and associated assumptions about concurrent tasks and neuropsychological work, has been used to investigate cognitive tasks that require both the processing and storage of information. What role do the working memory sub-components play in particular cognitive tasks? One approach to studying the role of working memory in complex cognition is to use dual task methodology, and this will be summarised in detail in Chapter 2. These studies have revealed knowledge about the role of different components of working memory in a variety of cognitive information processing activities, following the Baddeley model perspective. An alternative method of the study of working memory and complex cognition is an individual differences approach. This approach uses various working memory span tasks as research tools. These span tasks are designed to represent the working memory demands during the performance of complex cognitive tasks by placing simultaneous demands on both processing and storage. This individual differences

approach specifies the role of working memory in complex cognition by correlating participants' performance on these span measures with that on other target tasks.

1 . 3 INDIVIDUAL DIFFERENCES AND WORKING MEMORY

1 . 3 . 1 Introduction

Individual differences and working memory research addresses the same issues as the dual-task work, but in a contrasting style. An individual differences approach is an alternative way of thinking about working memory. This tradition, in contrast to the dual-task work, has not adopted the Baddeley framework. The dual-task work indicated working memory involvement in performing higher level cognitive tasks - a decrement was seen in performance on a primary task when performed concurrently with a secondary task, assumed to load one of the working memory systems. The individual differences approach tries to identify components which we can measure independently and in which there is independent individual variation. This approach is an alternative way of seeing which components are involved in cognitive processing, for instance reasoning.

The initial research concentrated on fluent reading ability, where individual differences are known to exist between people. Early studies attempted to test the hypothesis that fluent reading ability depends upon working memory. For instance, Perfetti and Lesgold (1977) used a standard digit span measure, which required participants to remember a set of digits and then to recall them in the order in which they had heard them. The more digits a person could remember, the higher was their short-term memory capacity score. They found only a weak relationship between this and reading skill. Daneman and Carpenter (1980) suggested that working memory contains procedures for both processing as well as storage functions.

They suggested that the digit span used by Perfetti and Lesgold (1977) may have been a reliable measure of a certain type of storage, but required little in the way of simultaneous storage and processing. It might be that the standard digit span was a good predictor of an individual's ability to store numbers in the PL, but was not a good predictor of executive or active processes related to the role of the CE. Thus, tasks developed since these early studies recognise the fact that a span task, which taxes both the storage and processing of information simultaneously, is measuring something different from passive span tasks, such as the digit span.

Baddeley and Hitch (1974) argued for the importance of storage and processing of activities within working memory. The limitation in capacity is believed to reside in simultaneously satisfying both the processing and storage demands that a given task imposes. Baddeley (1986) conceptualised working memory as an active system for temporarily storing and manipulating information. As we have seen, the two storage systems (the PL and VSSP) are assumed to be relatively passive slave systems primarily responsible for the temporary storage of verbal and spatial information. A memory task that taxes the participants capability to store verbal or spatial information may lead to results telling us the capacity that person has in respect to these systems. The CE is conceptualised as very active and responsible for the selection, initiation, and termination of processing routines (e.g., encoding, storing, and retrieving). A memory task that taxes the participants capability to store and process information concurrently may tell us about the capacity that person has in respect to this system.

There are two types of memory span of interest here - simple and complex measures. The simple measures refer to those with no obvious processing requirement. They only require

the storage of information for later recall. The complex measures refer to those with both a storage and a processing requirement. They require both the storage and processing of information. The simple or passive measures can be mapped onto Baddeley's two slave systems, where information is stored for future recall. The complex or active measures can be mapped onto Baddeley's CE system, where information is actively processed.

Engle, Kane and Tuholski (1999) state that when they refer to 'working memory capacity' they mean the capabilities of the limited capacity attention mechanism which Baddeley and Hitch (1974) called the CE. Kane and Engle (1998) made the argument that the construct of working memory capacity is isomorphic with the capacity for controlled processing, which has a strong relationship to general fluid intelligence. Kane and Engle (1998) tested participants on a variety of complex and simple working memory span tasks, together with two tests of fluid intelligence. Confirmatory factor analysis confirmed the need to separate the simple and complex span tasks even though the two constructs were strongly related. The complex span tasks were shown to connect strongly with the measures of general intelligence, however the simple spans were not connected in this way. This finding lends support to the idea that the component of the working memory tasks that is important to higher-order functioning is controlled attention. As for Baddeley and Hitch (1974), the results here suggest that what is important in the complex span measures is processing capacity, not storage.

Some of the passive measures of working memory capacity can be related to the PL component. As previously mentioned, the standard digit span was assumed to investigate whether individual differences in verbal ability reflected differences in passive verbal working memory capacity. The word span is another task that has been suggested to

measure this relationship (Shah & Miyake, 1996). This task requires participants to remember a set of words in a set. At the end of a set, their task is to recall each word in the order in which they have seen them. This task can be mapped onto the PL component due to the fact that it only requires the storage of verbal information, and there is no processing involved.

Other passive measures, for instance the simple arrow span, were developed to investigate whether individual differences in spatial ability may reflect differences in passive spatial working memory capacity (Shah & Miyake, 1996). This task requires participants to remember the orientation of a set of arrows in a set. At the end of a set, their task is to recall the orientations on a recall grid, in the order in which they have seen them. This task can be mapped onto the VSSP component due to the fact that it only requires the storage of spatial information, and there is no processing involved.

Another span measure is the complex reading span which was developed to investigate whether individual differences in reading comprehension may reflect differences in complex working memory capacity. This complex span task contained both processing and storage components. The reading span task was designed as a span measure of functional working memory capacity for language. The task requires the simultaneous maintenance and processing of verbal information. A single trial consists of the individual presentation of a small set of sentences. The participant is asked to read each sentence within a set (processing component), whilst remembering the last word of each sentence (storage component). At the end of a set of sentences, participants are asked to recall the last word of each sentence in the order in which they read them. The number of sentences per set is incremented from trial to trial and the participant's reading span is defined as the maximum

number he/she could read while maintaining perfect recall of the final words. In the context of the Baddeley model, this task is assumed to measure the capacity of the CE component of working memory, due to the additional processing function (Daneman & Carpenter, 1980).

A major stream of research studies in working memory has been concerned with the role of working memory in aspects of reading. Advantage can be taken of the knowledge that individual differences between people occur in reading ability. Participants can be given a range of tasks (memory and comprehension tasks), and a correlational approach can be used to identify which components of language appear to be associated most closely with different kinds of memory capacity. Although language comprehension is not being studied in this thesis, these studies are important, as many working memory measures were devised in light of their findings.

1 . 3 . 2 Working Memory Capacity and Reading Comprehension

Using span measures as a technique is an artificial laboratory-based process. Thus, it is fortunate that in educational and clinical settings, it is of ecological validity (Baddeley, 1979). The most obvious relevance is in relation to learning the complex but important skill of reading. This section begins with some evidence of the applicability of using memory span as a research tool in the area of reading development. This will be followed by a range of studies of adult reading performance. Reading comprehension has been included because the main tradition of individual differences research stems from here.

The evidence seems to suggest that the PL, or some similar system, plays an important role in learning to read (Jorm, 1983). Case, Kurland and Goldberg (1982) researched the development of memory span in children by studying the relationship between vocabulary,

digit span (remembering a sequence of six digits) and the speed with which a child could process digits. Their results showed that a child's capacity on all these measures increased systematically with age. They argue that the primary cause for increase in memory span stems not from a change in strategy adopted, but rather from an increased efficiency in carrying out the relevant control processes. The findings were suggested to show that as a child develops, their language skills become more efficient. This allows them to perform operations such as rehearsal more easily making fewer demands on the limited attentional capacity, and hence allowing more items to be stored. The Baddeley framework can be applied here in two ways. The first explanation could be that the memory span is dependent on a limited capacity CE. The executive might use processes such as the PL system to store information, thereby freeing capacity for storing more items, either directly within the CE, or indirectly by the more efficient use of control processes. As a child develops, the PL system will become more and more efficient, and as such will require progressively less monitoring by the CE. The second explanation might be that the effects could be explained entirely in terms of the PL, without recourse to further assumptions about the CE. These findings for the development of reading skills in children should be reflected in the findings from adult studies, summarised next.

Using an individual differences approach, Daneman and Carpenter (1980) attempted to test the hypothesis that adult comprehension depends upon working memory capacity. Earlier studies (Guyer & Friedman, 1975) had attempted to explore this question using standard digit span measures, and found only a weak relationship between digit span and comprehension. Daneman and Carpenter (1980) argued that the digit span might be a reasonable measure of a certain type of storage, but requires little in the way of

simultaneous storage and processing. They argued that the essence of working memory is that it divides its capacity between storage and processing.

Daneman and Carpenter (1980) gave participants the reading span task (both visually and auditorily presented), together with a task involving the reading of passages of prose about which they had subsequently to answer questions. They also gave their participants the Scholastic Aptitude Test (SAT), a more general test of intelligence. It was found that people who had higher span scores performed better at comprehension and the capacity to draw inferences and to integrate information (a correlation of .72). There was also a healthy correlation (.59) between working memory span and performance on the SAT. These findings have subsequently been replicated (e.g., Daneman & Carpenter, 1983; Just & Carpenter, 1992).

Using the reading span as a working memory capacity task was clearly a very powerful predictor of reading comprehension. However, both its strength and its weakness stem from the fact that it is itself a relatively complex task, probably involving strategy selection, the PL, and knowledge of vocabulary, as well as the capacity to co-ordinate these various aspects of memory. A more varied approach to the study of comprehension came from Jane Oakhill and colleagues. Oakhill, Yuill and Parkin (1986) tested two groups of children on a test based on Daneman and Carpenter's (1980) reading span task. One group of children were normal on tests of vocabulary and the ability to read single words, but were poor performers on tests of comprehension. This group was tested against a group of children who performed normally overall. Oakhill et al. (1986) modified the reading span task, replacing the sentences with groups of three numbers (alleviates specific language difficulties). The children were instructed to read out the groups of three numbers, and then

recall the last number from each group. The reasoners were presented with two, three, or four groups of numbers. It was found that the children in the high comprehension group did better than children in the low comprehension group, and this difference increased as the number of digit groups became greater. Even without a language content to the task, the differences between groups still occurred. This finding was similar to that found by Daneman and Carpenter (1980) and the assumption was that the two groups differed in the attentional capacity of the CE.

Other aspects of reading comprehension have also been investigated, including syntactic ambiguity. Just and Carpenter (1992) uncovered a number of systematic individual differences in reading comprehension that were related to working memory capacity for language. They used the reading span task devised by Daneman and Carpenter (1980), together with a garden path task devised by Ferreira and Clifton (1986). The garden path task was constructed in such a way that the reader may initially assign incorrect structure to a sentence and then have to go back and re-analyse it again. The reader could avoid being led down a garden path only by making immediate use of information not based on grammar alone (semantics). Their results showed qualitative and quantitative differences between readers in their ability to accurately comprehend certain sentences as a function of working memory capacity. The qualitative differences among readers included the permeability of their syntactic processing to pragmatic information, and in their representing one versus two interpretations of a syntactic ambiguity. The quantitative differences among readers included the time course of comprehension and the accuracy of comprehension. The results showed that reading slows down at just that point in a sentence that introduces a computational demand, and slows down more for low span than high span participants. One explanation of this was because people with larger working memory capacities for language

were able to draw on a larger supply of resources. According to this view, working memory constraints exist in the maximum amount of activation that one has available for allocation to the processing and storage functions and manifest themselves mainly in the form of processing slowdown, the gradual loss of critical information, or both under capacity-demanding situations (Shah & Miyake, 1996). In conclusion, this study showed that performance differences in the comprehension of syntactically ambiguous sentences could also be explained in large part in terms of working memory.

A question that arises is whether these patterns of correlations between working memory and comprehension are due to storage or processing functions, or both. Daneman and Tardif (1987) argue that task-specific processing skill, not storage, is the real cause of individual differences in working memory capacity and the accompanying correlation with comprehension. They studied the relation between verbal abilities and three different span tasks: a verbal span, a math span, and a spatial span. All spans were similar at surface value, and all had a processing and storage component that could be assessed separately. The only factor that seemed to be important in determining whether a particular task would or would not predict reading skill was the domain of the processes that task tapped. For the task to predict reading skill, the processes had to be symbolic, that is, involve the manipulation of words (verbal processes) and, to a lesser extent, numbers (maths processes). The verbal span task again predicted a participant's ability to perform a reading comprehension task, as well as a general aptitude test (vocabulary test). None of the spatial measures correlated above zero with verbal ability. The most striking finding, according to Daneman and Tardif (1987), was that individual differences in processing accounted for all the interesting individual differences in comprehension and verbal ability. This study

showed that the processing component of a task is as, or more, important than its storage component and must be taken into account when looking at working memory capacity.

In a similar vein, Turner and Engle (1989) report two experiments which investigated whether correlations between the complex span and reading comprehension depend on the nature of the processing component and individual skill in that task. This and other studies (e.g., Engle, Cantor & Carullo, 1992), crossed the types of processing required in a span task (sentence verification or numerical equation verification) with types of information to be remembered (words or digits). The main findings can be summarised as follows. The span tasks correlated with verbal ability scores equally well when the processing component of the span task involved language processing and when it involved numerical processing. Again, it can be seen that the complex spans predicted performance of a participant's ability for reading comprehension. This study showed the importance of looking at both storage and processing components of tasks and is consistent with Daneman and Tardif's (1987) findings where comprehension correlated with both verbal and math spans. This study is also consistent with the notion that CE working memory resources are required when comprehending, due to the correlation with the complex span measures. The span measures required both storage and processing of information which might be afforded by a centrally controlled resource.

Although Daneman and Carpenter (1980) showed that simple span measures do not correlate with comprehension, there is contrasting evidence that they do. La Pointe and Engle (1990) investigated the relationship between a simple word span and the reading span task and reading comprehension. Their results showed evidence for a relationship between both span measures (simple and complex) and reading comprehension. An explanation for

this may have been that the PL was also involved in comprehension. La Pointe and Engle (1990) conclude that the reading span does not measure a working memory specific to reading. They found that reading comprehension was predicted as well by a complex span task involving arithmetic as by one involving reading. More importantly, the simple word span task also significantly predicted comprehension and, in some cases, did so as well as did the complex span task. This suggested that the complex and simple span tasks may not have been greatly different in what they measure.

La Pointe and Engle (1990) found a correlation between simple word span and comprehension whereas previous studies by Daneman and Carpenter (1980) and Turner and Engle (1989) did not find these correlations to be significant. The strategy used by La Pointe et al. (1990) was to see whether simple spans correlated with reading comprehension under conditions in which the presentation modality was more similar to that used with the reading span task (both visually presented). La Pointe et al. (1990) suggested that the reading span task does not have a special relationship with reading comprehension as Daneman and Carpenter (1980) proposed. However, this was only when similar conditions were upheld for both the simple and complex span measures. La Pointe et al. (1990) argued that one factor that both span tasks have in common is a reliance on verbal knowledge. They suggested that it might be some form of articulatory coding that is the critical factor.

In contrast, however, Waters and Caplan (1996) were interested in the measurement of verbal working memory capacity and its relation to reading comprehension. Participants carried out the following tasks - Daneman and Carpenter's (1980) reading span task, four versions of a related sentence span task in which reaction times and accuracy on sentence processing were measured along with sentence-final word recall, two number generation

tasks designed to test working memory, digit span, and two shape-generation tasks designed to measure visuo-spatial working memory, plus standard vocabulary and reading tests. Their results showed that it is the processing component of the sentence span tests that is the major determinant of the correlation between these tests and comprehension. The results suggested that sentence span tasks are unreliable unless measurements are made of both their sentence processing and recall components, and that the predictive value of these tasks for reading comprehension abilities lies in the overlap of operations rather than in limitations in verbal working memory that apply to both.

This section outlined the initial individual differences research in relation to comprehension and reading ability. The reading span task was devised with both a storage and processing component which might reflect the operations of a CE of working memory. These types of span task were favoured over and above the former span measures used, such as the digit span. The standard digit span type task shows little relationship with reading comprehension arguably because it relies only on a storage component. It can be seen that the only span measures that consistently showed positive and significant relationships with adult comprehension were those that relied on both storage and processing components. The evidence put forward might suggest that a working memory capacity reliant on both storing and processing of information is required for the process of reading comprehension. This conclusion could be fulfilled by the notion of a centrally controlled processor for language based ability. The research also shows that comprehension and reading ability have been linked with a working memory capacity for language due to the correlations with verbal abilities, but not with spatial ones. This section also presented evidence that span measures relating to letters or numbers showed positive relationships with reading ability. It

was only the spatial spans that were not predictive of comprehension (Daneman & Tardif, 1987).

The findings presented in this section, although not based upon Baddeley's model, can be mapped in this way. Research into the development of reading showed that it relied on a system for speech based coding - as articulation skills improve, so does reading performance. These studies were based on using a digit memory span as a measure of simple working memory capacity. Memory span was found to be related to the ability to read. These findings could be taken in conjunction with Baddeley's working memory model by recourse to the PL component. Research into the ability to read in adults also showed that it relied on a system for verbally based information due to the fact that comprehension correlated with complex spans based upon verbal material. These findings could be interpreted in line with Baddeley's idea of a CE component of working memory.

Thus, reading comprehension seems to be predicted consistently by complex span tasks assumed to measure the CE systems of working memory. The findings showed that individual differences in working memory capacity can account for variations in performance on comprehension and reading skill tasks. Looking at individual differences in working memory and reading has been a successful approach. It is encouraging that the application of this approach has been so fruitful, and lends more support to applying the same approach in relation to reasoning. If it is the case that these measures only relate to language abilities, they may not be as useful for examining the role of working memory in reasoning. However, it appears that this is not the case and Section 1.3.3 follows on in the individual differences vein covering cognitive ability measures and working memory from this perspective.

1 . 3 . 3 Working Memory Capacity and Cognitive Ability

Cognitive psychologists have examined the role that simple information-processing abilities play in moderating individual differences in complex cognition. Areas of study include comprehension (Daneman & Carpenter, 1980; Just & Carpenter, 1992), which has already been discussed, as well as reasoning (Kyllonen & Christal, 1990) and academic achievement (Engle, Cantor & Carullo, 1992). One common thread in this work is the importance placed on the role of working memory. These investigations link two central constructs - working memory capacity and cognitive ability - which arise from two distinct bodies of literature on individual differences in cognition - the information-processing and the psychometric traditions, respectively.

Perhaps one of the most comprehensive studies of working memory capacity and cognitive ability was carried out by Kyllonen and Christal (1990). They used a battery of 25 tests to 723 military recruits to investigate the relationship between reasoning ability and working memory capacity. They used six tests of working memory capacity, together with four general knowledge tests, two processing speed tests, and 15 reasoning tests. The working memory span tasks were all based upon either alphabetic or arithmetic problems where participants had to recall the inferred relationships between letters or numbers in series. The span tasks required concurrent processing and storage and could be mapped onto the CE system of working memory. The reasoning tests were selected from the paper-and-pencil Armed Services Vocational Aptitude Battery (ASVAB). The reasoning tasks were based upon three types of reasoning factor proposed by Carroll (1989) - deductive serial reasoning (nonsense syllogisms), inductive (verbal analogies) and quantitative (arithmetic) reasoning. The results demonstrated a consistent and remarkably high correlation between reasoning ability and working memory capacity. Confirmatory factor analysis yielded

consistently high estimates of the correlation between these two factors ($r = .80$ to $.90$).

Kyllonen and Christal's (1990) measures of reasoning were standard cognitive ability tests and these were found to have strong relationships with measures of working memory capacity. Kyllonen and Christal (1990) interpret their results as an indication that individual differences in reasoning ability reflect differences in working memory capacity (e.g., Kyllonen, 1996).

Similarly, Jurden (1995) was interested in individual differences in working memory and complex cognition. In two studies, participants completed two working memory span tasks designed to assess verbal and nonverbal working memory, as well as assessments of verbal intelligence, nonverbal intelligence, and academic achievement. Verbal working memory was assessed with an adaptation of the reading span, whilst nonverbal working memory was assessed using the computational span task (Salthouse & Babcock, 1990). The verbal span required participants to read sets of sentences and answer a multiple choice question about each, as well as remembering the last word of each sentence in a set. In contrast, the nonverbal span required participants to solve simple arithmetic problems with a recall component. Verbal working memory had no relationship with nonverbal intelligence but did relate to verbal intelligence, whereas nonverbal working memory had no relationship with verbal intelligence and academic achievement, but did relate to nonverbal intelligence. Kyllonen and Christal's (1990) data were also reanalysed. Autonomous verbal and nonverbal working memory factors were each identified with multiple indicators. Verbal working memory was more highly correlated with verbal assessments than with nonverbal assessments; the opposite was true for quantitative working memory. These results, together with Jurden's (1995) experiments, suggest that working memory capacity can

predict the performance of cognitive tasks, other than comprehension, and that a dissociation can be seen between verbal and nonverbal working memory.

Shah and Miyake (1996) have also identified dissociations between different types of working memory capacity. They investigated spatial thinking and language comprehension by comparing the magnitude of correlations between individuals performance on a spatial span task and a reading span task with other spatial and verbal ability measures. They also examined the relative contribution of the processing and storage components of spatial and verbal working memory for predicting performance on spatial and language tasks. The results from their first experiment showed a correlation between the complex spatial span task and the spatial ability measures, but not with the verbal ability measures. In contrast, the complex verbal span task correlated with the verbal ability measures, but not with the spatial ability measures. Again, the working memory tasks predicted specific performances on the ability tests.

In their second experiment in this series, Shah and Miyake (1996) developed four different working memory span measures by crossing the type of processing requirements (mental rotation or sentence verification) with the type of information to be maintained for later recall (spatial orientations indicated by arrows or two-syllable words). This interference paradigm was used to demonstrate that both the processing and storage requirements in a working memory span task contribute to the pattern of correlations with complex cognitive tasks. Participants were either in the arrow or word conditions, where the span tasks involved remembering spatial or verbal information, respectively. The same battery of cognitive ability tests were used for participants in both conditions, in conjunction with the span measures. They compared the pattern of correlations for the span tasks involving the

processing and storage of same-modality information, with the patterns obtained for the span tasks based on different-modality information, and with the simple spans that required no explicit processing. The results showed that only the span tasks involving the same storage requirement (i.e., spatial or verbal information) showed similar correlations with the same domain ability tests. The results replicated the dissociation between verbal and spatial working memory found in Experiment one, and further demonstrated that both the processing and storage demands of working memory tasks are important for predicting performance on spatial thinking and language processing tasks.

Shah and Miyake (1996) did not measure a participant's accuracy in performing mental rotation or sentence verification (processing component). They suggested that there was no indication of participants trading processing accuracy for better span scores. They investigated the role of processing in Experiment 2 by manipulating the processing requirement to same or different modalities in relation to the ability measures. These findings were based on the storage components of the span tasks measured, and it must be acknowledged that this study made no separate assessment of the processing component of the span tasks.

This section has summarised a number of studies that have investigated the links between working memory capacity and cognitive ability. Many different span tasks have been devised and can be interpreted as measuring individual differences in the components of Baddeley's (1986) working memory model. The span measures are assumed to reflect either the passive storage of the subsystems, the PL or VSSP, or to reflect active processing within working memory, the CE. The span tasks are based on the storing and/or processing of information and they measure verbal and nonverbal material. The findings from this

research showed a confirmation of the link between working memory capacity and cognitive ability. The studies illustrated links between verbal spans and verbal abilities, not nonverbal ones. It is worth pointing out that the nonverbal tasks were mostly based on numerals, not spatial information. The majority of the findings in this section lend support to the view that the processes involved in complex cognition tasks could reflect the resources afforded by a CE of working memory. Some of the research showed that the spans with different processing and storage components actually correlated with one another, especially when the information was based on verbal or mathematical abilities. This was assumed to reflect a unitary working memory system, where the spans measured this capacity. However, other studies have shown a dissociation between span tasks with different processing and storage components. These findings might suggest a separation of central resources of working memory.

Very little of the individual differences research has been conducted strictly within the Baddeley framework, although as discussed, most of the findings can be reconciled with his model. However, one aspect of the findings that may not be so easily reconciled is evidence for the separation of resources at the CE level - that there may be a dissociation between different types of working memory capacity. Therefore, Section 1.3.4 discusses what research does exist about the idea of a unitary executive and questions whether it could be a range of different systems for varying information.

1 . 3 . 4 Is Working Memory Capacity Domain-General or Domain-Specific?

Many of the studies described in Section 1.3.3 examined working memory capacity in relation to whether the CE's capacity could be separated for different types of information. If a task dependent on verbal or spatial ability could be predicted equally well from both

verbal and spatial spans then a general architecture explanation could be accepted. Put another way, this finding would support the view that a domain-independent explanation of working memory CE resources fits the data best. If a task dependent on verbal or spatial ability could be predicted differently from a verbal or spatial span task, then a specific architecture explanation could be supported. Put another way, this finding would support the view that a domain-specific explanation of working memory CE resources fits the data best. This section will outline the traditional unitary view of working memory capacity first. This notion supports the view that all higher level cognitive thinking is reliant on a single pool of general purpose resources. The section will go on to examine a more contemporary separability view of working memory capacity. This notion supports the view that there are separate pools of resources that different processes and representations rely upon.

Turner and Engle (1989), as explained previously, asked the question as to whether working memory is task dependent. The span tasks correlated with verbal ability scores equally well when the processing component of the span task involved language processing and when it involved numerical processing, suggesting some domain-generality of working memory. This work indicated that, when the processing component of the span task was arithmetic-related, it led to correlations with reading comprehension similar to those found when the processing component was reading. This added more evidence, claimed Turner and Engle (1989) and others (e.g., Engle, Cantor & Carullo, 1992), that working memory resources are domain general. However, it should be noted that Turner and Engle (1989) only looked at linguistic versus numerical processing. They did not look at specifically spatial tasks, only nonverbal tasks, in relation to verbal tasks.

Jurden (1995) also investigated the separation of working memory capacity for different types of information and his study was summarised in Section 1.3.3. Verbal working memory was shown to have no relationship with nonverbal intelligence, whereas nonverbal working memory was shown to have no relationship with verbal intelligence and academic achievement. In Kyllonen and Christal's (1990) study, confirmatory factor analysis suggested the existence of a single working memory factor - a verbal versus quantitative content factor. Jurden's (1995) results, together with the re-analysis of Kyllonen and Christal's (1990) study, suggest that rather than being a unitary system, the working memory system may be a parallel processing system. The subsystems may operate in tandem, possess some degree of autonomy, and have their own control processes. It is of note here that many of these studies have used a separation of information by recourse to verbal versus nonverbal assessments. However, as also seen for Turner and Engle (1989), nonverbal does not necessarily mean specifically spatial and this should be noted.

Daneman and Tardif (1987) argue that individual differences in working memory capacity for different types of information should be examined. As we have seen, they studied the relation between verbal abilities and three different span tasks: a verbal span, a math span, and a spatial span. The processing component of the spatial span required participants to view sets of cards. Each card contained three 3 x 3 tic-tac-toe grids, with red and blue tokens occupying certain cells. The participants were to imagine the three grids as representing the top, middle, and bottom layers of a three-dimensional tic-tac-toe game. They were to mentally combine the three layers and locate the winning line by touching the three tokens with their finger. After a set of two to four cards, the storage component of the spatial-span task required the participants to recall the winning lines for all the cards in the set. If the math span and spatial span measures predicted reading skill as well as the verbal

span measure then there would be support for a general working memory system. If only the verbal span measure predicted reading skill then there would be support for a language-specific processor. Daneman and Tardif (1987) showed that only the math and verbal span tasks, not the spatial span task, predicted verbal ability. They concluded, therefore, that there are separate working-memory resources for language- and nonlanguage-based information processing. They suggest that rather than the PL and the VSSP acting as temporary storage systems with processing functions limited primarily to rehearsal, they should be considered as the processors themselves.

As stated in Section 1.3.3, Shah and Miyake (1996) tested the separability of working memory resources for spatial thinking and language comprehension. In Experiment one, all participants completed a complex spatial and verbal working memory span, plus a simple spatial span measure. The complex and simple spatial spans were developed as spatial analogues to the Daneman and Carpenter (1980) reading span task and the digit span, respectively. In the complex spatial span measure, the task was to decide whether each one of a set of letters was normal or mirror imaged while simultaneously keeping track of the orientation of each of the letters in the set. In the simple spatial span measure, the task was to keep track of the orientations of a set of arrows. They also completed a battery of tests including tests of spatial visualisation, perceptual speed, and verbal and quantitative tasks.

Their results showed a correlation between the complex spatial span task and the spatial ability measures, and a correlation between the complex verbal span task and the verbal ability measures. The spatial spans were not correlated with verbal ability, the verbal span was not correlated with spatial ability. An exploratory factor analysis was also performed on this data, with a resulting two factor model (uncorrelated). The first factor was loaded

highly on the three spatial ability measures, the two spatial span tasks, but not the verbal measures, indicating that this was a spatial factor. The second factor was loaded highly on the two verbal measures, but not on the spatial ability measures, suggesting that this was a verbal factor. Shah and Miyake (1996) interpreted this result as evidence for domain specific components of active working memory, beyond the scope of the subsystems identified in Baddeley and Hitch's (1974) model.

Similar conclusions regarding the non-unitary view of CE resources have been drawn from work investigating working memory span and executive function tasks. Lehto (1996) was interested in the relationship between working memory capacity and three executive function tests - the Wisconsin Card Sorting Test (WCST) plus The Tower of Hanoi (TOH) and Goal Search tasks were studied. The digit span and word span tasks were used as measures of simple working memory capacity (PL). The sentence-word span, operation-word span and memory-updating and backward digit span were used as measures of complex working memory capacity (CE). The executive tasks used are commonly employed to measure executive function amongst patients with frontal lobe injury. The WCST correlated significantly with working memory tasks, the storage function of working memory probably being a limiting factor in card sorting. The two other executive tasks did not correlate with working memory tasks. In addition, none of the executive tasks exhibited any significant intercorrelations. The present results and evidence from earlier studies suggest that there does not exist a unitary, limited capacity central executive. The absence of interrelationships between tasks led to the conclusion that these tasks tap into separate functions that are independent of one another and unrelated to working memory capacity. Lehto (1996) only used span measures based upon verbal information, not spatial. It might

be that if Lehto (1996) used working memory spans based upon spatial information, links may have been seen between it and the spatial executive tasks, such as the TOH.

This final section has opened up the question of whether working memory central resources are domain general or domain specific. The field of research is split between traditional views that suggest a unitary executive and those that suggest a separability of executive processes for varying information. Many studies only concentrated on verbal span measures with verbal and nonverbal cognitive tasks. Only Daneman and Tardif (1987) and Shah and Miyake (1996) used spatial capacity measures based on both the storage and processing of spatial information. More research is needed to systematically examine the role of both complex verbal and spatial capacity measures and their relationships to individual differences in complex cognition, this being one of the primary issues with which this thesis is concerned.

1 . 4 CONCLUSION FOR THE WORKING MEMORY REVIEW

This review started with a look at the precursors to the contemporary model of working memory. The modal model (Atkinson & Shiffrin, 1971) was found to be a too simplistic account of the available evidence, thus the working memory model (Baddeley & Hitch, 1974) was developed. Researchers have tried to establish that firstly the capacity of this system is limited, and secondly that using a substantial amount of this available capacity should have broadly comparable effects across a range of different cognitive tasks.

Much of the evidence for a tripartite model of working memory came from dual-task work. These studies examined detriments in performance on a primary task when carried out concurrently with a secondary task assumed to load one of the working memory subsystems (e.g., Hitch & Baddeley, 1976; Vallar & Baddeley, 1982). Working memory has been

conceptualised as an active system for temporarily storing and manipulating information needed in the execution of complex cognitive tasks (e.g., learning, reasoning, and comprehension).

The PL has been conceptualised as responsible for storing speech based information (e.g., Baddeley, 1966; Baddeley, Thomson & Buchanan, 1975). The VSSP has been conceptualised as responsible for the manipulation and temporary storage of visual and spatial information (e.g., Baddeley & Lieberman, 1980). The CE is postulated to be responsible for the encoding, storing, and retrieving of information (e.g., Baddeley, 1966).

Many studies have been carried out to draw out the characteristics of the three systems that make up the working memory model. Research has provided evidence that the PL is a speech-based system and is important for learning to speak and to read and for comprehending spoken discourse. Work using the dual-task paradigm suggests the separability of a system for setting up and manipulating images, the VSSP. The pattern of results suggests separate visual and spatial components of imagery. Evidence put forward for the CE system has been less clear. However, it has been suggested that the Norman and Shallice model gives a good account of the functioning of the CE. Experimental evidence (random generation) supports the claim that the system involves processing and storage resources, and that scheduling and planning also play a role. The experimental evidence gives only a general picture of the characteristics of this system and it is the least defined component of the working memory model.

The review continued by describing an alternative way of looking at the role of working memory capacity in cognition - through an individual differences approach. Working

memory span tasks have been used in conjunction with a range of cognitive tasks to evaluate the relationship between them. A range of working memory span tasks have been shown to be predictive of performance on comprehension tasks (e.g., Daneman & Carpenter, 1980; Just & Carpenter, 1992) and cognitive ability tasks (e.g., Kyllonen & Christal, 1990; Jurden, 1995).

The review ended with a look at the contemporary view of the structure of working memory. Span tasks considered to represent either the storage of information in the PL and VSSP, or processing and storage in the CE show differing results. One perspective showed evidence for a unitary CE (e.g., Turner & Engle, 1989), whilst another showed evidence for separate pools of CE resources (e.g., Shah & Miyake, 1996).

The value of the individual differences approach must be emphasised due to the wide range of results that can be accounted for when predicting from capacity measures. This approach has informed us about the role of working memory in cognitive tasks. It can be seen from this review that working memory capacity was a significant predictor of an individual's performance on a range of cognitive tasks, including comprehension and cognitive ability. A range of measures have been employed in looking at individual differences in cognitive ability. However, the individual differences work, as yet, has not tackled the relationship between complex deductive tasks and working memory capacity tasks. The dual-task work has investigated a much wider range of ability and complex deductive reasoning tasks and these complex tasks need to be addressed from the complementary individual differences perspective. Deductive reasoning tasks, such as syllogistic and spatial reasoning, have not been investigated in relation to a participant's working memory capacity for different sorts of information.

The evidence concerning the unitary versus non-unitary view of CE resources also raises interesting research questions. If it is the case that verbal and spatial executive resources are distinct, this has obvious implications for the structure of the working memory model (Baddeley & Hitch, 1974). This thesis will further examine the evidence for this proposal. In addition, this thesis will examine the extent to which working memory capacity predicts performance on a range of reasoning tasks. It will also examine whether different reasoning tasks draw more on spatial or verbal capacity resources. To what extent are individual differences in reasoning explained by individual differences in working memory? Which reasoning problems seem to rely on a verbal working memory resource, and which on a spatial resource? There are a number of reasoning theories, that specify different representations and processes, and it may be possible using this methodology to differentiate between these theoretical accounts. In the chapter that follows, a summary is provided of these theoretical perspectives and the evidence for the role of working memory in human reasoning is examined.

CHAPTER 2 - A REVIEW OF DEDUCTIVE REASONING

2 . 1 INTRODUCTION TO THE REVIEW

As mentioned in Chapter 1, this thesis is interested in the question of whether working memory plays a role in performance on deductive reasoning problems. From Chapter 1 it could be seen that, through both a dual-task and individual differences methodology, working memory was implicated in performance seen on many cognitive tasks. This experimental series addresses the relationship between complex deductive reasoning tasks and working memory capacity tasks, taking an individual differences approach. The two tasks that have been consistently used throughout this thesis are syllogistic and five-term series spatial inference problems. Therefore this review of deductive reasoning will concentrate on these tasks and use them as examples for discussion.

Chapter 2 will introduce the literature on deductive reasoning. After a short introduction, the literature relating to syllogistic inference will be summarised. The nature of syllogistic problems, the phenomena associated with them, and the theories that have been forwarded to explain these phenomena will be outlined. This will be followed by an explanation of spatial inference. Chapter 2 then goes on to review the literature concerning the connection between reasoning and working memory. Research that has examined the role of working memory in reasoning using dual-task methods will be examined first. This will be followed by an alternative way of looking at working memory and reasoning from an individual differences approach. Chapter 2 will end with a more detailed rationale for the first set of experiments.

2 . 1 . 1 What is Deduction?

The word 'reason' can be defined as "the intellectual faculty by which conclusions can be drawn from premises" (The Oxford English Dictionary, 1995). Deduction is central to human intelligence and without it we would not be able to learn and apply general principles, or generalise our experience from one situation to another. To be able to reason deductively, the reasoner must attend to and process particular steps of the task in mind, and this is where working memory comes in. To be able to reason is a fundamental attribute and individuals often reason to draw novel inferences using background knowledge. Johnson-Laird and Byrne (1991) suggest that deduction is used in all sorts of situations including the formulation of plans and the evaluation of actions, as well as to determine the consequences of assumptions and hypotheses.

Evans, Newstead and Byrne (1993) give some examples of deductive arguments. Working out whether we qualify for a home improvement grant or a tax rebate, or whether a child can choose the combination of school subjects that they wish, are all examples of deductive reasoning. These examples involve understanding rules and regulations and applying them to personal circumstances. Thus deductive reasoning involves a process where the goal is to draw a valid consequence from the premise information given. This thesis indirectly examines the question of how deductive reasoning problems are represented, stored and processed; and directly examines whether they rely on a limited working memory capacity.

As mentioned in the brief introduction above, Chapter 2 will be concentrating on two types of deductive task - syllogistic and spatial inference. The past research concerning both these types of task have uncovered certain phenomena that needs to be accounted for by any reasoning theory. Theories of deductive reasoning should be able to explain why, in

general, people show evidence of logical competence on these tasks. They should also provide theoretical accounts of the errors that people make in reasoning and the way in which people's knowledge influences their patterns of response.

This review chapter will describe the phenomenon associated with deductive reasoning tasks and discuss claims concerning the underlying mental processes. The focus will be on syllogistic and five-term series spatial reasoning problems in terms of their nature, performance on them and theoretical explanations for them. In this way, reasoning theory will be introduced in light of the findings for these two tasks. The chapter will move on to a discussion of the research that has examined the role of working memory and reasoning. This will be done by firstly examining research that has employed a dual-task methodology. An alternative approach will then be proposed, using an individual differences methodology to examine the relationship between deductive reasoning and working memory capacity tasks. Finally, the rationale for the thesis in general will be introduced. Section 2.2 outlines the nature, patterns of performance and theoretical accounts of these performance patterns for the syllogistic task.

2 . 2 SYLLOGISTIC REASONING

2 . 2 . 1 Syllogisms

A definition of a syllogism is,

a form of reasoning in which a conclusion is drawn from two given or assumed propositions (premises): a common or middle term is present in the two premises but not in the conclusion, which may be invalid. (The Oxford English Dictionary, 1995).

A syllogism consists of two premises and a conclusion that link three classes. The following is an illustrative example of a syllogism (Example 2.1)

Some of the actors (*a*) are bakers (*b*)

All of the bakers (*b*) are carpenters (*c*)

Therefore,

Some of the actors (*a*) are carpenters (*c*).

Example 2 . 1

The premises contain three terms (*a*, *b*, and *c*) one of which, the middle term *b*, occurs in both premises. The conclusion connects the other two terms (*a* and *c*). Each premise can take one of four forms or contain one of four quantifiers - All, Some, None, and Some not. The form of the quantifiers determines the mood of the syllogism as shown in Table 2.1. Table 2.1 shows the moods of the premise information as described by Evans, Newstead and Byrne (1993).

Quantifier or Mood	Evans Classification
All <i>as are bs</i>	A universal
Some <i>as are bs</i>	I particular
No <i>as are bs</i>	E universal
Some <i>as are not bs</i>	O particular

Table 2 . 1 - The four moods of the syllogism

The mood of the whole syllogism is determined by the premises and conclusion, and an example of an IEO problem (shown in Table 2.1) would be Example 2.2.

Some bakers are athletes

No bakers are carpenters

Therefore,

Some carpenters are not athletes.

Example 2 . 2

Table 2.2 illustrates the figure of a syllogism as determined by the position of the three terms within the two premises. Figures 1 and 2 are asymmetrical because B is located in different places in the two premises. Figures 3 and 4 are symmetrical because B is located in the same place in the two premises.

	Figure 1	Figure 2	Figure 3	Figure 4
Premise 1	A B	B A	A B	B A
Premise 2	B C	C B	C B	B C

Table 2 . 2 - The four figures of the syllogism

Following Cohen and Nagel's (1934) scholastic logic, there are four figures, and in addition the two premises and conclusion can each contain one of four different quantifiers. This calculates to 256 different syllogisms (64 premises * 4 conclusions = 256 syllogisms).

Some of these syllogisms are valid arguments, they yield a conclusion that must be true given that its premises are true. Some of these syllogisms are invalid arguments, they do not yield a valid conclusion to the premise information given. Scholastic logicians accepted no logical effect of premise order and therefore ignored half of the possible syllogistic structures. Today all 512 structures are investigated in Psychology, because performance on these problems may be affected in ways other than logic.

One of the first studies investigating all 64 forms of premise information was Dickstein (1978). Participants were presented with a syllogism and were required to indicate which of five conclusions followed (i.e., all, some, none, some...not, or impossible to tell). Thus the data collected referred to a participant evaluating rather than producing a conclusion. Participants did make many errors on this syllogistic task, and the rank ordering of the difficulty in solving each syllogism was evaluated. The invalid problems were found to be harder than the valid ones. Some of the valid syllogisms were solved accurately nearly all the time (e.g., All of the *bs* are *as*, All of the *cs* are *bs* resulted in 95% accuracy), whilst others were very rarely solved correctly (e.g., None of the *bs* are *as*, All of the *bs* are *cs* resulted in 24% accuracy).

Experiments that investigate syllogistic reasoning tasks follow very different methodologies and therefore may reflect different aspects of the reasoning process. The premises can be given followed by all the possible conclusions as response options. Within this evaluation approach there are many more variations in methodology. Participants can be presented with complete syllogisms, two premises and a conclusion, and asked to judge their validity. In this method, participants must evaluate rather than produce a conclusion. Participants can also be presented with the two premises and asked to generate their own conclusions.

Studies investigating performance on this task indicate a vast difference in difficulty between the valid syllogisms. Some are so easy that a child could solve them, some so hard that an adult has problems responding correctly. There are a range of phenomenon associated with syllogistic reasoning which any theoretical account must explain. The following section aims to summarise the findings of research investigating the nature of different syllogisms.

2 . 2 . 2 Syllogistic Reasoning Phenomenon

Theoretical accounts of syllogistic reasoning performance must be able to account for performance patterns found in syllogistic experimental data. As already mentioned, some syllogistic problems are solved nearly all the time, whilst others are hardly ever solved correctly. An example (Example 2.3) of an easier valid syllogism can be seen below. The following examples are taken from Johnson-Laird & Byrne (1991), where participants were asked to generate their own conclusions to the syllogistic statements presented.

All actors are bakers
All bakers are carpenters
Therefore,
All actors are carpenters.

Example 2 . 3

The conclusion to this syllogism was produced by 89% of participants. The following syllogism is an example (Example 2.4) of a harder valid problem (taken from Johnson-Laird & Byrne, 1991).

No actors are bakers
Some bakers are carpenters
Therefore,
Some carpenters are not actors.

Example 2 . 4

The conclusion to this syllogism was produced by only 20% of participants. The preferred conclusion that was systematically observed to be used by the other participants was that there was no valid conclusion. The differences in accurately responding to these types of

valid syllogism must be accounted for by theory. The following syllogism is an example (Example 2.5) of an invalid problem (taken from Johnson-Laird & Byrne, 1991).

Some athletes are not bakers

All bakers are carpenters

Therefore,

No valid conclusion.

Example 2.5

There is no one conclusion that interrelates the end terms here, so a logical conclusion is not possible. The preferred conclusion, however, that was produced by the majority of participants was that some athletes are not carpenters. The pattern of systematic errors observed in Examples 2.4 and 2.5, when individuals produce syllogistic conclusions, must be explained by theoretical accounts.

Many different aspects of the syllogistic problem have been explored, and have been found to have an effect on accurate responding. The mood and figure of the premise information affect the conclusions drawn. Typically, it has been found that syllogisms in Figure 1 are the easiest, those in Figure 4 the hardest, and those in Figures 2 and 3 somewhere in between (e.g., Frase, 1968). This has become known as the figural difficulty effect.

On occasions, participants seem to prefer conclusions in one direction rather than another, the directional effect of the conclusion (e.g., Johnson-Laird & Steedman, 1978). Studies have related these directionality effects to the four figures of syllogistic reasoning (e.g., Frase, 1968; Dickstein, 1978). These authors suggest that the natural order of processing in figure 1 (AB-BC) is from A to C, however in figure 4 (BA-BC) is from C to A. Johnson-

Laird and Bara (1984) proposed that reasoners have to carry out mental operations on syllogistic premises in order to bring the two middle terms into contiguity. Premises in figure 1 already occur in an order in which the two occurrences of 'B' follow one another, and so there is no need for any operation to bring them into contiguity. Premises in figure 4, however, requires the reasoner either to re-new their interpretation of the first premise in order to switch it round or else to switch round their interpretation of the second premise and then re-new their interpretation of the first premise. These, and other studies (e.g., Johnson-Laird & Steedman, 1978) demonstrated that people prefer to process information in a forward rather than a backward direction.

Thus, individuals are influenced by the order of the terms in the premises, however, they are also influenced by the content assigned to the premises and conclusion. Belief bias effects can be seen as errors that occur when reasoners make inferences on the basis of the believability of the conclusion, rather than the logical structure of the syllogism. There have been many studies that have investigated this claim (e.g., Kaufmann & Goldstein, 1967; Evans, Barston & Pollard, 1983; Oakhill & Johnson-Laird, 1985). There are three basic findings from these studies, highlighted by Evans, Newstead and Byrne (1993). It is suggested that believable conclusions are more readily accepted than unbelievable ones. Secondly, it has been shown that logically valid conclusions are more readily accepted than invalid ones. Thirdly, evidence reveals an interaction between logical validity and believability such that the effects of believability are more marked on invalid conclusions. All these factors must be taken into account whilst determining how individuals attempt and solve these types of problem.

In summary, there are four findings from the syllogistic studies that theoretical accounts must explain. The first finding is that reasoners show competence on the syllogistic task - more valid inferences are drawn than invalid ones. A sub-set of syllogisms is usually solved quickly and correctly, whilst another sub-set is found to be very difficult to solve accurately. The second finding is that the ordering of the premises affects the difficulty of correctly responding - syllogisms in figure 1 are easier to solve than those in figure 4. The third finding is that the order of the terms in the premises influences the preference of the conclusion generated - the preferred conclusion for figure 1 is A-C, the preferred conclusion for figure 4 is C-A. The fourth finding is that the believability of the conclusion in evaluation tasks influences the extent to which reasoners endorse the given conclusion. There is an interaction between validity and believability - a greater effect of believability can be seen for the invalid problems. In the following section a range of theories of syllogistic reasoning will be examined. In addition, the extent to which each of these theories can explain the data will also be summarised.

2 . 2 . 3 Theories of Syllogistic Reasoning

There are many explanations proposed for the errors individuals produce on syllogistic tasks. The reasoner must interpret the premises, combine the premises, and produce a response. Error may occur at any of these stages. This section will discuss different theories that have tried to explain the syllogistic phenomena described above. The first set of theories loosely fit into approaches based upon verbal processes and include interpretational, rule and verbal reasoning accounts. The second set of theories loosely fit into approaches based upon processes that involve the manipulation of mental models or analogical representations. The final set of theories are based upon heuristics, where syllogistic reasoning performance and errors are explained through the application of simple

strategies. What follows is a summary of how each of the three classes of theory describe and account for the data.

- **Verbal processing accounts**

The first class of explanations suggest that people make errors in interpreting the premise information. For instance, given the premise 'All as are bs', participants may draw the incorrect inference that 'All bs are as'. According to standard logical criteria, this reversing of terms does not hold. Some authors suggest that people may illicitly convert the quantifiers in each premise (e.g., Chapman & Chapman, 1959; Revlis, 1975; Newstead, 1989). There is still an incomplete understanding of the role conversion plays in the interpretation of the premise information, although it may be part of the explanation of accurate performance. Other authors suggest that people may interpret the quantifiers according to their meaning in everyday language rather than in a logical way (e.g., Begg & Harris, 1982; Politzer, 1990). These studies produced conflicting evidence for the errors being based on Gricean maxims. Gricean maxims refer to rules governing the use of language in everyday conversation. For instance, it would be wrong to say 'some' when you know 'all' to be the case. In conclusion, it seems appropriate to propose that there is some evidence to suggest that Gricean maxims could be responsible for some of the errors which occur on syllogistic reasoning tasks (conversion). Although this theory does not fully explain the effects of figure and belief, considered earlier. The studies have demonstrated that indeed these types of error occur, but there is little evidence to suggest this as the major limitation in solving syllogistic problems (Newstead, 1995).

As in other areas of reasoning, there is a theoretical debate as to whether the fundamental processes involved in syllogistic reasoning are based upon propositions, mental logic, or

deduction rules (Rips, 1994), and those that are based upon arrays, spatial representations, or mental models (Johnson-Laird, 1983).

The mental logic approach to reasoning is a well established theory and attempts have been made to explain syllogistic reasoning performance (see Rips, 1986). It is suggested that people reason by constructing proofs using an internal set of abstract reasoning rules. The logic is abstract and general purpose and all real-world problems must be translated into some kind of abstract code in order for reasoning to occur.

The most complete formal rule theoretic explanation for syllogistic reasoning comes from Rips (1994) (although, see Braine & O'Brien, 1998, who have also developed an account of predicate logic). Rips (1994) suggested that people first convert the premises into an abstract propositional form. They then apply the appropriate inference rules to produce the conclusion generated. Errors are explained by the absence of an appropriate rule, or by the total length of the reasoning process, or by failure to encode the premises correctly.

Following is an example of a syllogistic argument (Example 2.6) given as an example by Rips (1994).

All square blocks are green blocks

Some big blocks are square blocks

Thus,

Some big blocks are green blocks.

Example 2.6

A set of inference rules can be applied to the quantified statements, showing the argument to be deducible. The quantifier rules (Elimination and Introduction) for this example are

shown in Figure 2.1. As the syllogistic argument becomes more difficult, then the set of rules associated with processing it will also increase.

<i>Deductive Processes</i>	<i>Rules</i>
a. + (FOR ALLx)(IF Square-block(x) THEN Green-block(x)).	Premise
b. + (FOR SOMEEx)(Big-block(x) AND Square-block(x)).	Premise
c. Big-block(b) AND Square-block(b).	FOR SOME Elim.
d. Big-block(b).	AND Elim.
e. Square-block(b).	AND Elim.
f. IF Square-block(b) THEN Green-block(b).	FOR ALL Elim.
g. Green-block(b).	IF Elim.
h. Big-block(b) AND Green-block(b).	AND Intro.
i. (FOR SOMEEx)(Big-block(x) AND Green-block(x)).	FOR SOME Intro.

Figure 2 . 1 - A set of inference rules for a syllogistic problem (Example 2.6)

Rip's (1994) work clearly captures competence in syllogistic reasoning by equipping individuals with deductive processes and inferential rules that allow them to accurately evaluate syllogistic conclusions. Individuals can solve valid problems because they possess and apply these rules. Invalid conclusions are correctly rejected because there are no rules to form a proof.

Rips proposed that systematic errors made on syllogistic problems are due to the inability to access certain rules. So the difficulty in solving any deductive argument is related to the number of rules required and the availability of those rules. The directional effect is explained by a guessing heuristic. Whilst this enables a description of a preference for certain conclusions, the heuristic is ad hoc and based on the data. The issue of figure is explained as a by-product of implicature and rule use rather than the order of the terms

themselves. When these are held constant Rips reports no figural differences in his data. Quite how figural difficulty is a product of rule use is not made explicit, and how this might explain the preference for conclusions in a certain direction is also left unexplained. In addition, Rips has been criticised due to his inability to explain ‘content effects’ in reasoning. The strong syntactic nature of his theory, which implements logical rules on logical forms, means that this account is not ideally suited to accommodate the influences of belief.

Some accounts propose formal inference rules prescribed by logic, whilst others propose verbal processes, similar to those used in verbal comprehension, as the major determinant in solving syllogistic problems. A theory that proposed these types of processes was put forward by Polk and Newell (1995) and is called the verbal reasoning (VR) theory. Polk and Newell (1995) suggest that representational processes alone can explain reasoning performance. VR is proposed to be accomplished exclusively by existing comprehension and production skills, in particular, by comprehending and recomprehending the premises, that is, by building an initial model and incrementally augmenting it. Polk and Newell (1995) proposed four processes of VR: initial encoding, conclusion generation, reencoding, and giving up. For example, consider the following syllogism (Example 2.7).

Some B are A

All B are C

Thus,

Some A are C

Example 2.7

The first stage, initial encoding, is when individuals encode direct knowledge from the premise information. The individual constructs a mental model of a situation in which the premises are true. These models identify properties corresponding to the topics of the propositions being encoded. In Example 2.7 this would involve encoding the first premise into an annotated model, followed by the second premise. For instance, 'some B are A' would be annotated as (B^*) , $(B^* A)$, where $*$ is an identifying property in relation to 'some B'. The second premise, 'all B are C' would be additionally annotated as $(B^* C)$, $(B^* A C)$, where $*$ is an identifying property in relation to 'all B'.

Once the premises have been encoded, an attempt is made at producing a conclusion based on the annotated model. This is done by using a simple generate-and-test procedure. First, simple propositions are proposed that are considered to be true based on the annotated model. The simple propositions that VR may propose include the legal syllogism conclusions but may also involve the middle term, other quantifiers, or both. If none of the proposed propositions are legal syllogism conclusions, generation fails and reencoding is evoked. For instance, with Example 2.7, the initial encoding might trigger the following conclusions - 'some B are A', and 'all B are C'. Neither of these conclusions is a legal one, and thus reencoding occurs.

If generation fails to produce a legal conclusion, an attempt is made to try to extract knowledge from the premises by reencoding. Unlike initial encoding, reencoding chooses a property based on the annotated model (the reference property) and then tries to extract additional knowledge about that property from any premise that mentions it. In Example 2.7 reasoners might reencode 'some B are A' using 'C' as the reference property ($(B^* C)$, $(B^* A C)$), or alternatively reencode 'some B are A' using 'A' as the reference property

((B* C), (A*), (B* A* C)). From the second reencoding, based on indirect knowledge about 'A', the extracted annotated model leads to the generation of a legal conclusion, 'some A are C'. If the generation of a conclusion does not relate the end terms (A and C), reencoding must occur until the reference property leads to a legal conclusion. It is assumed that most individuals give up only as a last resort, when they feel confident that repeated attempts to extract more knowledge will not lead anywhere.

The theory offered by Polk and Newell (1995) has a number of properties relevant to the question of functional constraints. The model performs the task of reading the two premises and attempting to produce a conclusion. The model is functional in that it draws on abilities, like language skills, that exist to provide the reasoner with certain functional capabilities (such as speaking, listening, reading) motivated independently of the particular task. Reasoner's difficulties with the task are explained as a consequence of this functional basis. Polk and Newell (1995) point out that although reasoners' everyday language skills can serve to perform the task, they are not perfectly adapted to that task, and so by themselves are not adequate for perfect performance. Thus, difficulties with the task are explained by recourse to functional, not resource, constraints.

Polk and Newell (1995) proposed that participants make errors in syllogistic tasks because they accept plausible conclusions. The acceptance of particular conclusions is not based upon logical validity due to the fact that only plausible conclusions are considered. An explanation of preferred conclusions from particular premise combinations are also covered. For instance, negative quantifiers in the premise information create negative associations between classes in the annotated model and tend to lead to negative conclusions. The theory also explains the directional effect to a point. It can explain why there is a preference

for conclusions in the A-C direction for figure 1, and C-A direction for figure 2. This is because the identifying end term for figure 1 is A and for figure 2 is C. Conclusions linked to these properties are likely to be tried first and consequently, are the preferred conclusion directions. Belief bias is also investigated. Verbal Reasoning attempts are often inconclusive, so other knowledge sources are used on which to base a decision. Also, because reasoning attempts can be faulty, results people would normally accept as valid are reconsidered if contradicted by beliefs.

Polk and Newell (1995) carried out extensive statistical modelling of existing data obtained from syllogistic reasoning tasks (as well as a discussion of five-term series spatial descriptions and other reasoning tasks). They suggest that accounts based on verbal reasoning could account for the reasoners' performance. They go on to say that assumptions about linguistic processes and their results provide the theoretical leverage in almost every explanation, including relational deductions. Therefore, Polk and Newell (1995) proposed that the central processes in deductive reasoning are linguistic or verbal comprehension processes. These included processes involved in encoding, re-encoding, and generation of information rather than reasoning-specific skills. This theory has been developed as a general theory of cognition rather than as a direct model of particular reasoning tasks. Thus, Polk and Newell (1995) proposed that reasoners use verbal comprehension processes that are common to reasoning and language tasks alike.

- **Analogical processing accounts**

An alternative view of syllogistic reasoning performance has been put forward by the mental model theorists. There have been many proponents of the models approach to reasoning (e.g., Erikson, 1978; Fisher, 1981; Sternberg & Turner, 1981). Johnson-Laird's

(1983) mental models theory is the best developed of this class of approaches and has been applied to many reasoning tasks subsequent to its initial strong association with explaining syllogistic reasoning. It is assumed that people form representations of the premises consisting of tokens of the terms linked together so as to exemplify the stated categorical relationships. Each token represents an individual element of the set. In the examples used previously, each token would represent a profession (e.g., actor, baker etc.). They then seek to combine the two premise representations into an integrated representation from which a possible conclusion can be read.

Participants must integrate the information contained in the premises, construct a model of that information, draw a putative conclusion, and then search for alternative models that might refute the initial conclusion chosen (falsification). In some cases there is claimed to be only one way in which the premise representations can be combined (single model), but in other cases there may be two or three possibilities (multiple models). If people do explore all possible premise combinations then they perform perfectly. Suboptimal performance is explained by failure to consider all ways of combining premise information due to working memory limitations. The theory predicts that syllogisms that can be represented in a single model will be easier than syllogisms that can be represented by multiple models, and this proves to be the case.

As an example, consider the following one-model problem (Example 2.8). In Johnson-Laird's (1983) mental models theory, the information in this syllogism would be represented in the following type of mental model, shown in Figure 2.2.

All of the actors are bakers
 All of the bakers are carpenters
 Therefore,
 All of the actors are carpenters.

Example 2 . 8

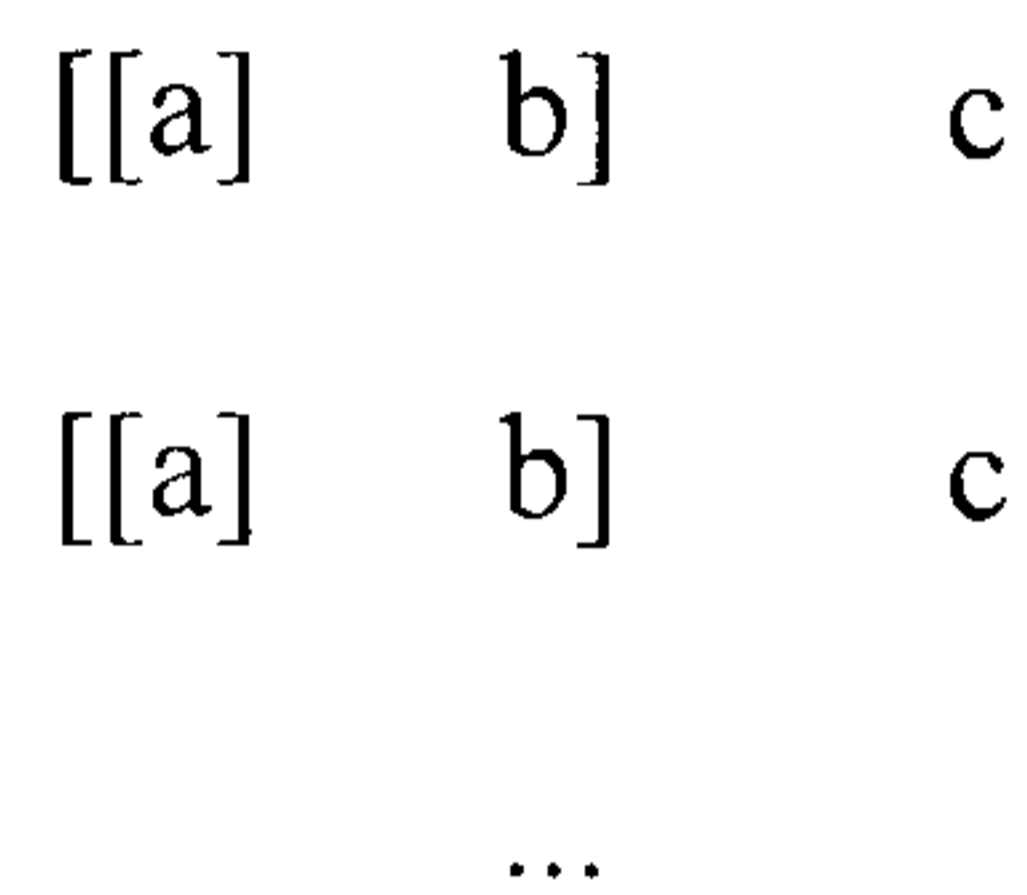


Figure 2 . 2 - Mental model representation for a single model syllogistic problem (Example 2 . 8)

The brackets shown in the annotated model in Figure 2.2 represent that *a* is exhausted in respect to *b*, and *b* is exhausted in respect to *c*. The three dots shown at the bottom of the annotated model indicate that there could be other individuals who are not yet represented in the model. With one-model syllogisms it is not necessary to flesh out the model in order to produce the conclusion that ‘all of the *a*s are *c*s’. This one-model syllogism is a valid argument and depends on the construction of this one model. There is no other way of constructing a model where the conclusion doesn’t hold. In contrast, the following problem is a multiple model syllogism (Example 2.9).

Some of the actors are bakers
 None of the bakers are carpenters
 Therefore,
 Some of the actors are not carpenters

Example 2 . 9

In Johnson-Laird's (1983) account, the information in this syllogism could be presented in any one of the following types of mental model (Figure 2.3).

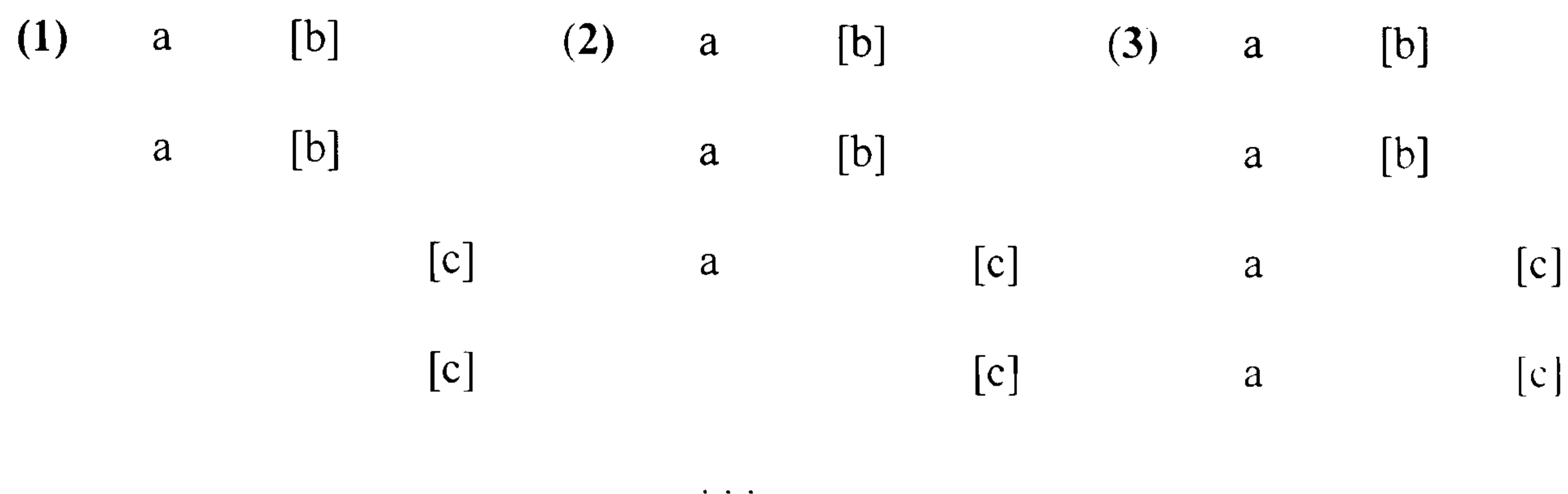


Figure 2 . 3 - Mental model representation for a multi model syllogistic problem (Example 2 . 9)

This multi-model syllogism is a valid argument that does depend on the construction of more than one model. In the context of the process of mental model theory, a participant would first integrate the premise information. They would then construct a model from this information, and the one constructed might well be Figure 2.3 (1). A putative conclusion would then be drawn from this initial model, for instance, that none of the actors are carpenters. Then they might search for alternative models, perhaps generating Figure 2.3 (2) and (3), which refute the initial conclusion. The only conclusion that holds across all three models is that some of the actors are not carpenters.

If the full set of models are not generated, erroneous conclusions may be produced that are consistent with a subset of the models generated. This means that once a possible conclusion has been established, the validity of this needs to be explored by constructing alternative models in which the premises are true but where the conclusion does not follow. If people fail in this exploration, then they adopt the initial conclusion as valid; but if they

succeed, they try to find another possible conclusion and, failing that, say no conclusion follows. Johnson-Laird and Bara (1984) found that the majority of their participants could correctly solve one-model syllogisms, whilst hardly any could correctly solve multi-model problems.

Many of the main syllogistic phenomenon described can be explained by recourse to the mental models theory. Easy syllogisms involve the construction of a single model.

Difficult syllogisms involve the construction of multiple models. If more than one model has to be constructed to reach the right response, then the problem will be more difficult than a one-model problem. This factor explains the difference in difficulty among the valid syllogisms, and it accounts for the characteristic errors made on the multi-model problems. Successful, or competent reasoners produce the correct conclusion due to the fact that they choose one supported by all possible models. Unsuccessful reasoners may only construct a sub-set of the models and produce erroneous conclusions based on this set.

Johnson-Laird and Byrne (1991) have provided an explanation of the figural effects on the conclusion production task in their re-analysis of Johnson-Laird and Bara (1984). They found fewer correct responses, and more erroneous conclusions that no valid conclusion follows, on figure 4 syllogisms (BA-BC). This was attributed to the directional preference in which reasoners construct mental models, since they tend to form a model of the first premise and then add the information in the second premise to this. If the conclusion is read off in a left-to-right direction, then the observed preferences would be expected. Because individuals prefer to process in a forward direction, A to C conclusions should be easiest to solve, and this has proved to be the case (Dickstein, 1978). Figural effects stem from the difficulty of rearranging the information so that the two middle terms occur together. In

figures 3 and 4 (see Table 2.2) the middle terms appear in the same positions in both premises, and so cause the most difficulty in the rearrangement of B. Conclusion preference is seen as a by-product of the way in which terms are integrated into working memory. Belief bias effects have also been explained. According to this account, unbelievable conclusions cue the search for counterexamples, so invalid unbelievable conclusions are more likely to be rejected. The search will also be cued by valid unbelievable conclusions, although there are no models in which the conclusion does not hold and therefore the conclusion will not be rejected. This proposal explains the interaction between belief and logic.

The mental models theory tries to encompass all stages of syllogistic reasoning, namely, interpretation, premise combination, and response production. The theory proposes that it explains both rational competence and the pattern of performance, including errors and difficulties. It should be noted that the mental model theory suggests that reasoning occurs primarily when there is a subsequent attempt to falsify conclusions generated from the models constructed. This is done, as stated previously, by attempting to construct alternative models in which the conclusion no longer holds true. This notion is at the heart of the mental models theory. However, it is worth pointing out that recent evidence suggests that falsification may not be a common strategy (Newstead, Handley & Buck, 1999; Evans, Handley, Johnson-Laird & Harper, 1999).

Another explanation, based on the idea that reasoners construct analogical models of the syllogistic premises, proposed that people use a mental analogue of Euler circles to represent set membership and the relations between sets (e.g., Erikson, 1978; Guyote & Sternberg, 1981). The Euler circles indicate the possible relationships between the three

classes *a*, *b* and *c*. Erikson (1974; 1978) assumed that individuals encode premises as Euler circles, but can only handle one diagram for each premise. Once the individual has formed these representations, they try to combine the two diagrams they have constructed. Erikson (1974; 1978) again assumed that people can handle no more than one composite representation. In combining premises in this way, clearly a number of representations are overlooked. This theory is difficult to evaluate due to the fact that testable predictions are hard to derive for new data sets. Thus Euler circles has not been a recent contender to explain syllogistic performance.

Another explanation sees performance as a search through the problem space, and that this latter is best characterised as involving representations analogous to Venn diagrams (e.g., Newell, 1981). Venn diagrams involve three overlapping circles, each of which represents one of the sets in the syllogism. Syllogisms can be solved by marking these circles in specified ways for different premises. While Venn diagrams are recommended by many as a good technique for solving syllogisms, it has not often been claimed that they provide a means by which people actually do solve syllogisms. Newell (1981) presented no empirical evidence to support his claim. However, Newell (1990) has more recently proposed that Venn diagrams are only used by reasoners who are expert in their use and that the problem space he introduced could be translated as a mental models explanation.

- **Heuristic processing accounts**

The last class of explanations to be considered are models that propose that error occurs at the response stage of reasoning and that performance is based upon heuristics or biases.

Woodworth and Sells (1935) were the pioneers of a response generation explanation of syllogistic reasoning performance called ‘atmosphere’. They suggested that people respond

on the basis of the “mood” of the quantifiers presented in the premises. not on the basis of logical analysis of the problem (Sells, 1936). Begg and Denny (1969) are also proponents of this view and put forward two rules by which people respond:

Quality Rule - if one or more premise is negative then the preferred conclusion will be negative.

Quantity Rule - if one or more premise is particular then the preferred conclusion will be particular.

For instance, an example of the use of the quality rule can be seen in Example 2.10. Since both the premises are positive and particular, there should be a strong preference for a positive, particular conclusion. The predicted conclusion ‘Some A are C’ is in fact given regularly by participants (90% of responses in Johnson-Laird and Bara, 1984).

Some A are B

Some B are C

Example 2 . 10

Overall, the atmosphere hypothesis has been found to account well for the response data in a number of studies (Begg & Denny, 1969; Revlis, 1975; Dickstein, 1978). Although atmosphere theory predicts the response data reasonably well, it does not explain why people respond according to the principles of quality and quantity.

The matching hypothesis (Wetherick, 1989) is another response generation explanation of syllogistic reasoning performance. It is simpler than the atmosphere notion and has been explored by Wetherick and Gilhooly (1990). It is suggested that reasoners match the quantifier in the conclusion to that used in the more conservative of the premises. Evidence

for the existence of matching as a strategy was reported by Wetherick and Gilhooly (1990).

Wetherick (1989) put forward the following rule -

Give as the conclusion a proposition of the same logical form as the more conservative of the premises, where the logical forms are ordered for conservatism, from most to least, "No", "Some Not", "Some" and "All".

When the two quantifiers are the same in the two premises, atmosphere and matching make exactly the same predictions. The predictions only differ with the premise pairs IE and OE, where atmosphere predicts an O conclusion, matching an E conclusion. Thus, matching adds little to atmosphere. Atmosphere and matching are examples of theories that only focus on predicting data and give no explanation of theoretical underpinning. Due to their inability to explain why reasoners adopt these heuristic strategies, they are not fully established models of syllogistic reasoning. Why is it that individuals should induce these heuristics? The phenomenon of syllogistic reasoning performance is not well explained, although an alternative does have a theoretical underpinning, and is called the Probability Heuristics Model (PHM, Chater & Oaksford, 1999).

Following in the same vein as the atmosphere and matching theories is the probability heuristics model (PHM) of Chater and Oaksford (1999). The PHM suggests that syllogistic reasoning performance may be determined by simple but rational informational strategies justified by probability theory rather than by logic. Oaksford and Chater (1992; 1993) argue that none of the currently proposed mechanisms for deductive competence (e.g. rules, models) are capable of accounting for real-world reasoning, because they place a high demand on the limited information processing capability of human beings.

Chater and Oaksford (1999) claim that the purpose of syllogistic argument is to link the “end terms” in the chain via their connection with the “middle term”. So, as in a chain, the strength of the conclusion that links the end terms depends on the strength of the weakest link. In other words, the heuristics, to be described below, rely on an ordering in the informativeness of quantified statements that serve as premises of syllogistic arguments. Chater and Oaksford (1999) specify the following order of informativeness, starting with the most highly informative: All > Most > Few > Some > None > Some...Are Not.

Following are the heuristics that implement the informativeness principle. At the algorithmic level, participants are suggested to rely on three generation heuristics - the *min*-heuristic, *p*-entailments, and *attachment*-heuristic. These will be explained briefly below.

The *min*-heuristic is explained as follows -

Choose the quantifier of the conclusion to be the same as the quantifier in the least informative premise (the ‘min-premise’). The max-heuristic refers to a participant’s confidence in the conclusion generated in proportion to the informativeness of the most informative premise (the ‘max-premise’).

The most informative conclusion that can validly follow from a pair of syllogistic premises always follows this rule. Furthermore, some conclusions probabilistically entail (*p*-entail’) other conclusions. For instance ‘Some’ entails ‘Some are not’. Thus, the second heuristic is, the *p*-entailments, explained as follows -

The next most preferred conclusion will be the p-entailment of the conclusion predicted by the min-heuristic (the ‘min-conclusion’).

The *p*-entailments include a family of heuristics corresponding to the probabilistic relationships between the various quantified statements that make up the premises of a

syllogism. These first two heuristics specify the subject of the conclusion. The third heuristic specifies the order of end terms in the conclusion, the *attachment*-heuristic, and is explained as follows -

If the min-heuristic has an end term as its subject, use this as the subject of the conclusion. Otherwise, use the end term of the max-premise as the subject of the conclusion.

The theory assumes that individuals first interpret quantified statements as probabilistic statements. They then use these probabilistic semantics in deriving an analysis of informativeness, which justifies the informativeness ordering over which the heuristics are defined. Thirdly, individuals do not explicitly consider validity, it being seen as a by-product of the process, defined over probabilistic statements.

All Y are X	(<i>max</i> -premise)
<u>Some Z are Y</u>	(<i>min</i> -heuristic)
I-type conclusion	(by <i>min</i>)
Some Z are X	(by <i>attachment</i>)

Example 2 . 11

The heuristics can be illustrated by Example 2.11. By the *min*-heuristic, the conclusion is ‘Some’. The *min*-premise has an end term (Z) as its subject. Therefore, by *attachment*, the conclusion will have Z as its subject term and the form ‘Some Z are X’.

Many of the phenomenon of syllogistic reasoning performance can be explained by recourse to the PHM. Logical competence is explained by individuals correctly applying the *min*-heuristic (and to some extent *p*-entailments when required) and the attachment-heuristic. These are performed without any knowledge of probability or validity. If there is a valid conclusion to be found, then individuals will tend to draw it. Errors are explained due to

individuals employing heuristics that produce a plausible, but inappropriate conclusion. The figural effects are explained by specific processes that determine conclusion direction (*attachment*-heuristic). Belief bias is explained as a response bias. False conclusions carry no informativeness, thus participants are biased towards accepting conclusions they believe because it is something that arises automatically in the context of everyday reasoning. However, this explanation of belief bias does not explain the belief-logic interaction, it only predicts a main effect of belief.

At the computational level, PHM follows recent work showing that simple heuristics can be highly adaptive insofar as they approximate optimal solutions (e.g., Gigerenzer & Goldstein, 1996). It is suggested that these heuristics are used in place of complex and sometimes computationally intractable optimal strategies, which are therefore denied any cognitive role. Rather than seeking a post hoc explanation of the success of processing principles observed in human performance, the PHM is suggested to describe the optimal theory and processing heuristics in tandem. PHM has been confirmed by two experiments with syllogisms involving ‘Most’ and ‘Few’ which are beyond the scope of other accounts. However, Chater and Oaksford (1999) make no attempt to consider the way people represent syllogistic premises which is where mental models (Johnson-Laird, 1983) fairs better.

- **Summary**

This section outlined many differing explanations for an individual’s performance on syllogistic reasoning tasks. The first set of approaches see verbal processes as most important. Performance on the syllogistic problems are suggested to rely upon propositional or verbal information. Some accounts suggested that the premises are

converted (e.g., Chapman & Chapman, 1959) and others suggested that the rules governing language, Gricean maxims, are involved when explaining syllogistic phenomenon (e.g., Politzer, 1990). Both of these theories can not fully explain all of the syllogistic phenomena and so are not complete accounts. Another set of theorists believe that syllogistic premise information is translated into an abstract propositional code and inference rules are applied in order to form a mental derivation of the conclusion from the premises (e.g., Rips, 1994). The last theory from this approach, verbal reasoning, suggested that there are four propositionally based processes that are used to process deductive arguments (e.g., Polk & Newell, 1995). Polk and Newell (1995) make no claim about special processes to support logical reasoning by constructing and manipulating multiple models.

The second set of approaches see the representation of syllogisms as spatial or analogue in nature. Performance on the problems are suggested to rely upon the construction of a mental analogue of the syllogistic premises. Some accounts suggest that the premises are represented by tokens in a mental model, which may be spatial or propositional in nature (e.g., Johnson-Laird, 1983). This theory can account for a variety of responses, including conversion and figural effects. Others suggest that individuals use models of the premises based upon either Euler circles (e.g., Erikson, 1978) or Venn diagrams (e.g., Newell, 1981), which require spatial representations. However, there is little empirical support for the claim that mental representations rely upon the construction of Euler circles or Venn diagrams (Evans, Newstead & Byrne, 1993).

The last set of approaches claim that syllogistic reasoning performance depends upon processes other than those proposed by formal logic. Performance is suggested to rely on strategies, heuristics or probability, rather than an underlying logical reasoning mechanism.

Some theories recognise that individuals may be led by the mood of the premise information, the atmosphere hypothesis (e.g., Begg & Denny, 1969). Others suggest that individuals match the quantifier in the conclusion to that used in the more conservative of the premises, the matching hypothesis (e.g., Wetherick, 1989). These response bias accounts lack theoretical underpinning. The last of the theories looked at in this section was called the PHM (e.g., Chater & Oaksford, 1999). They suggested that syllogistic reasoning performance may be determined by simple but rational informational strategies justified by probability theory rather than by logic.

The next section will introduce the task that is to be used as the spatial equivalent to the syllogistic reasoning problems. The task will be outlined and evidence will be put forward for particular phenomenon associated with it. These inferences depend on relational terms such as 'to the left of' and 'to the right of'. These inferences underlie many of our more complex, abstract skills, such as mathematical, scientific and probabilistic thinking (e.g., Kahneman, Slovic & Tversky, 1982).

2 . 3 SPATIAL REASONING

2 . 3 . 1 Spatial Inference

Spatial reasoning is widespread in our everyday activities and interaction with the world: it underlies our ability to plan a route and to imagine objects from descriptions of their arrangement. There are many types of reasoning problem based upon spatial information. In this thesis, five-term series spatial inference problems are investigated, although much research has investigated three-term series problems. The following is an example

(Example 2.12) of this where three terms can be arranged in a linear order, and support a determinate conclusion.

The duck (*a*) is to the right of the pig (*b*)

The sheep (*c*) is to the left of the pig (*b*)

Therefore

The duck (*a*) is to the right of the sheep (*c*)

Example 2 . 12

A more complex problem is the five-term series spatial inference which consists of four premises and a conclusion that link five classes. The following is an illustrative example of one where a determinate conclusion can be supported (Example 2.13).

The duck (*a*) is on the left of the pig (*b*)

The sheep (*c*) is on the right of the pig (*b*)

The cow (*d*) is in front of the duck (*a*)

The horse (*e*) is in front of the pig (*b*)

Therefore,

The cow (*d*) is on the left of the horse (*e*).

Example 2 . 13

The premises contain five terms (*a*, *b*, *c*, *d*, and *e*), two of which are not explicitly related in any of the premises given (always *d* and *e*). Byrne and Johnson-Laird (1989) based their two-dimensional spatial deductions on four different orientations, illustrated in Table 2.3.

Figure 1	Figure 2	Figure 3	Figure 4
A B C	E D	D C	A E
D E	A B C	B	B
		E A	C D

Table 2 . 3 - The four orientations of two-dimensional spatial deductions

It should be noted again that many differences exist between methodologies used to investigate spatial relations. The premises can be given followed by all the possible conclusions. Participants can be presented with a complete spatial inference and conclusion then asked to judge their validity. These are evaluation tasks. Participants can be presented with the premises and asked to generate their own conclusions. This is a production task.

The relational reasoning problems studied as part of my thesis concentrate on five-term series spatial descriptions (Byrne and Johnson-Laird, 1989). As with syllogistic reasoning exercises, similar results have been found in respect to spatial problems. Some individuals find certain relational inferences more difficult than others, they take longer to make some inferences, and they make more errors when they must make the inference in a short time. There has been a large body of research investigating how individuals solve this type of inference. The following section aims to summarise the findings of research, in relation to both three-term and five-term series problems, investigating the relative difficulty of spatial descriptions.

2 . 3 . 2 Spatial Reasoning Phenomenon

As previously mentioned, some spatial problems are solved nearly all the time, whilst others are hardly ever solved correctly. An illustration (Example 2.14) of an easier three-term series valid spatial inference can be seen below.

The antelope is to the right of the baboon

The camel is to the left of the baboon

Therefore,

The antelope is to the right of the camel.

Example 2 . 14

The next illustration (Example 2.15) shows a similar three-term series problem, but this time there is no determinate conclusion that can be drawn between the end terms.

The baboon is to the right of the antelope

The camel is to the left of the baboon

Therefore,

No valid answer.

Example 2 . 15

Individuals make more correct inferences from the determinate problems with a valid answer (69% correct) than from the indeterminate problems with no valid answer (19% correct; taken from Byrne & Johnson-Laird, 1989). The five-term series problems can be illustrated in the same way. Example 2.16 shows a five-term series description where the conclusion is determinate.

The baboon is on the right of the antelope

The cat is on the left of the antelope

The dog is in front of the cat

The emu is in front of the baboon

Hence, the dog is on the left of the emu.

Example 2 . 16

The following spatial inference is an example (Example 2.17) of a valid problem that is valid but indeterminate, supporting more than one distinct layout.

The baboon is on the right of the antelope

The cat is on the left of the baboon

The dog is in front of the cat

The emu is in front of the baboon

Hence, the dog is on the left of the emu.

Example 2 . 17

The last type of spatial inference (Example 2.18) to be investigated are the indeterminate problems, where no valid conclusion follows.

The baboon is on the right of the antelope

The cat is on the left of the baboon

The dog is in front of the cat

The emu is in front of the antelope

Hence, the dog is on the left of the emu.

Example 2 . 18

Participants make more correct inferences from the problems that were determinate and valid (70% correct), such as Example 2.16, than from those that were indeterminate and invalid, such as Example 2.18 (15% correct). The problems that were indeterminate but valid, such as Example 2.17, were intermediate in difficulty (46% correct; taken from Byrne & Johnson-Laird, 1989). The level of difficulty amongst spatial inference problems must be accounted for by any theory. Theory must also explain other factors of representing and processing these spatial problems, to be discussed next.

Many different aspects of the spatial inference have been explored, and have been found to have an effect on accurate responding. The phenomena of spatial reasoning has been reviewed by Evans, Newstead and Byrne (1993). The format of the following summary will follow the same vein. Three-term series problems were originally studied to investigate the representations individuals adopted in order to correctly respond. The initial research explored whether spatial reasoning was reliant on spatial arrays or linguistic representations, exploring the phenomena of markedness. Four phenomena will then be summarised that help further clarify the representation of spatial inferences as either spatial or verbal in nature. Lastly, two contemporary accounts of the mechanisms underlying more complex spatial reasoning will be explored.

The initial debate concerning different representations of three-term series problems centred around whether they were based on directional spatial arrays (e.g., De Soto, London & Handel, 1965) or based on linguistic compression (e.g., Clark, 1969a). The spatial array view suggests that individuals go beyond the propositional form of the premises and construct some representation of the described situation which is independent of the

describing language. The linguistic view suggests that individuals preserve the propositional form of the language used in the description.

Markedness refers to the adjectives used to describe the premise information in a spatial description. This phenomenon has been studied to try and distinguish between the spatial array and linguistic accounts. For instance, 'Sally is *better* than Tom' is an unmarked adjective because it only describes relative goodness, it does not provide absolute information about whether they are both good or bad. However, 'Sally is *worse* than Tom' is a marked adjective because it not only describes relative goodness, but also provides absolute information about the fact that they are both performing badly. The spatial array view suggests that marked and unmarked adjectives are represented in similar ways. In contrast, the linguistic view suggests that they are represented in different formats because marked adjectives require the encoding of additional information.

Potts and Scholz (1975) suggested that their participants discard the linguistic information once they represent the adjectives in a uniform format. According to this spatial array view, individuals prefer to construct arrays from the top-down, rather than the bottom-up. For instance, 'A is better than B' requires construction of the array,

A

B,

where as 'A is worse than B' requires construction of the array,

B

A.

De Soto, London & Handel (1965) found that judgements based on the first unmarked array ('better') were correctly answered 61% of the time, whilst the second marked array

(‘worse’) was only judged correctly 43% of the time. This study found that reasoners performed at a higher rate on the unmarked adjectives as opposed to the marked adjectives.

According to the linguistic view, individuals construct an economical linguistic representation of the premises. ‘Worse’ describes the relative badness of the two entities, and it provides the absolute information that they are both bad, and so requires more information to be stored than the unmarked adjective (Clark, 1969a). In support of this claim, Clark (1969) found that ‘A is better than B, B is better than C’ which is based upon unmarked adjectives was easier than problems such as ‘C is worse than B, B is worse than A’ which is based upon marked adjectives.

The spatial array view proposes that individuals represent the unmarked and marked adjectives in a uniform spatial array. The differences in difficulty arise because reasoners have a preference for constructing arrays in a top-down rather than a bottom-up direction.

The linguistic view proposes that individuals represent the unmarked and marked adjectives in two different linguistic formats. The differences in difficulty arise because they compress the information in their linguistic representation, sometimes eliminating information that must be recovered subsequently to make an inference. Similar predictions about the difficulty of problems are made by the spatial array theory’s top-down preference hypothesis and by the linguistic theory’s compression hypothesis.

More recent research is most consistent with a spatial array view, and this evidence will be considered next, relating to how an integrated representation of the premise information is constructed. Descriptions of a single array can differ in the difficulty with which people make the inferences they support. Each description may describe the same layout, but the

order of the terms in the description, and the relations used to express the arrangement, lead to differences in the ease with which people make inferences. Evans, Newstead and Byrne (1993) consider four lines of evidence to help clarify the nature of an individual's constructed representation of five-term series problems. This evidence relates to the integration of the premises, the continuity of information, the determinacy of the description, and the distance effect.

The first strand of research looks at how people might integrate premise information (e.g., Hunter, 1957; De Soto, London & Handel, 1965). Hunter (1957) proposed that the problems that were hardest to solve required the use of transformations in order to combine the premise information. It has been shown that indeed the order of terms affects the conclusions produced. For instance, the ordering AB-BC tends to show more conclusions produced in the A-C direction, whereas the ordering BA-CB tends to show more conclusions produced in the C-A direction (Johnson-Laird & Bara, 1984). One possibility is that the integration of the premises in working memory may be easier for certain orders (e.g., Trabasso, Riley & Wilson, 1975). These results, and others (e.g., Mayberry, Bain & Halford, 1986), suggest that people must attempt the integration of the premise information and deal with it as a whole not parts. This has implications for working memory due to the fact that information in the premises must be integrated into a unified seriation.

Baguley and Payne (1998) used the more complex five-term series spatial reasoning problems to clarify predictions about increased processing cost during premise integration. The following predictions were all upheld. When mental model construction entailed the conversion of a sentence in order to enter a new token in an existing model, reading times increased. When a description was rendered indeterminate (i.e. when a clash occurred)

reading times increased. Reading times for the initial sentence were greater than for subsequent sentences (because two tokens have to be added to the model). This study provided more evidence for the construction of a mental model by looking at premise integration specifically.

The second source of evidence regarding the processing and representation of relational tasks is the continuity of information given in the premise information (e.g., Potts, 1972; Smith & Foos, 1975). Ehrlich and Johnson-Laird (1982) found that a referentially continuous description, in which each adjacent pair of sentences had a referent in common, required less listening time and elicited more correct diagrams than a referentially discontinuous description, in which the first and second sentences had no referent in common. Ehrlich and Johnson-Laird (1982) suggested that participants try to integrate each incoming sentence into a single coherent mental model, and that those sentences which cannot be immediately integrated are represented in a propositional form. This evidence lends support to the idea that it is easier to understand descriptions that present the information in a continuous order, which facilitates the combination of the information in a series into an integrated representation.

The third source of research findings in respect to these relational problems is the determinacy of the description (e.g., Byrne & Johnson-Laird, 1989). A determinate description is one where valid relations can be made between objects in a set. An indeterminate description is one where no valid relation can be made. Mani and Johnson-Laird (1982) carried out two experiments that investigated how participants recalled their representation of spatial descriptions. They used determinate and indeterminate problems to study whether they were processed differently. In Experiment one participants remembered

the meanings of determinate descriptions very much better than those of grossly indeterminate descriptions. In the second experiment they found that although the semantic implications of a determinate description are better remembered than those of an indeterminate description, the verbatim details of an indeterminate description are easier to recall than are those of a determinate one.

Mani and Johnson-Laird (1982) suggest that these results indicate the existence of two types of encoding -- propositional representations and mental models. The propositional representations are harder to remember but correspond to sentences in the description. These representations were seen more for participants processing indeterminate, rather than determinate, problems. The mental models are easier to remember but are analogous to spatial arrays and accordingly poor in linguistic detail. These representations were seen more for participants processing determinate, rather than indeterminate, problems.

The fourth and final class of phenomenon regarding the nature of spatial relations suggests that people are faster and more accurate in their judgements of the truth of inferred information than of presented information (e.g., Potts, 1972). This has also been named the distance effect. For instance, given A-B, B-C, C-D, A-B is truly presented, and A-C is truly inferred. Studies have shown that the farther apart the two terms to be judged are, the easier it is to compute their distance (e.g., Moeser & Tarrant, 1977). Thus, judgements about inferable relations are faster than judgements about presented relations. This research lends more support to the notion that reasoners must combine premise information in order to produce accurate conclusions. These results make sense if individuals are using a spatial array representation. The array must be inspected whether the judgement to be made is inferred or presented. The farther apart the two terms are in the array, the easier it is to

compute their distance. This is because they construct an integrated representation that makes explicit the relations between the terms.

Other research trends have concentrated on strategy explanations of performance, which mask the underlying processes and representations. Strategists have worked on the possibility of people using both spatial and/or linguistic processing techniques (e.g., Johnson-Laird, 1972; Sternberg, 1981). Many factors may influence the strategy an individual may use, for instance the instructions given (e.g., Barclay, 1973), or with training (e.g., Wason & Johnson-Laird, 1972; Mynatt & Smith, 1977). An interest in strategies as used in performing deductive tasks has revived in past years (e.g., Byrne & Handley, 1997; Johnson-Laird & Byrne, 1990). However, they remain an unsolved issue in the present research field (e.g., Roberts, Gilmore & Wood, 1997).

This section has highlighted the findings from the original work on three-term series problems, and more recent work on five-term series. The two early theories of the mental representations and cognitive processes required to make relational inferences are that they depend on procedures that construct and manipulate spatial arrays or models, and that they depend on procedures that construct and manipulate linguistic representations. The spatial array account proposed that individuals construct a uniform array with a preference for top-down models. The linguistic view proposed that individuals represent marked and unmarked adjectives in different formats where difficulty arises due to information compression.

Four strands of research further clarified the nature of the representation of spatial inferences. The evidence suggests that individuals construct representations that integrate

information from the premises, where continuous descriptions are easier to understand. Determinate descriptions are easier than indeterminate ones, and inferred distances are easier than presented ones. The evidence presented suggests that individuals construct models based on the structure of the situation the premises describe, rather than on linguistic representations based on the language used to describe the premise information.

The earlier research on spatial reasoning studied three-term series problems and was problematic because no clear evidence was presented for either the spatial array or linguistic account. More promising results were found from research investigating five-term series problems, outlined above. The next section summarises theoretical accounts of three-, four- and five-term series spatial problems to see how individuals infer from spatial descriptions. The theories that follow have been developed to provide an explanation for performance on spatial inference tasks using the types of problem outlined above.

2 . 3 . 3 Theories of Spatial Reasoning

Two major theories have evolved in this area - one based on inference rules, and the other based on models. The inference rule theorists suggest that the process of deduction depends on the particular rules and postulates that are used in deriving the conclusion (e.g., Hagert, 1983, 1984). The models approach theorists suggest that an individual imagines the state of affairs described in the premises, draws a conclusion from such a mental model, and searches for alternative models that might refute that conclusion (e.g., Byrne & Johnson-Laird, 1989).

- **Inference rule accounts**

The theories in this section link with the rule theories put forward to explain syllogistic inference in Section 2.2.3 (verbal processing accounts). The inference-rule theorists postulate a mental logic consisting of formal rules of inference that are used to derive logical conclusions (e.g., see Osherson, 1975; Braine, 1978; Rips, 1983). It is suggested that people reason by constructing proofs using an internal set of abstract, general purpose reasoning rules. The logic is abstract and general purpose and all real-world problems must be translated into some kind of abstract code in order for reasoning to occur. There have been inference-rule theories put forward for this type of two-dimensional spatial reasoning problem (e.g., see Hagert, 1983; Ohlsson, 1981, 1984, 1988). Hagert (1984) proposed an inference-rule system for two-dimensional reasoning that uses the following sorts of rules. Consider the following premises in Example 2.19.

B is on the left of A
C is in front of B
Therefore,
A is behind and right of C.

Example 2 . 19

The first and second premise would be represented linguistically as:-

Left (B, A)
Front (C, B)

The information in these two premises would then be integrated according to the following rule:-

Left (x, y) and Front (z, x) implies Left (Front (z, x) y)

The conclusion could than be implied by the premises (referred to on the right-hand side of the rule):-

Left (Front (C, B) A).

This abstract rule must be translated back to natural language to be able to produce a conclusion. If more rules are required in order to produce a conclusion, the more errors are likely to affect the reasoning process.

- **Model accounts**

Byrne and Johnson-Laird (1989) proposed that human mental logic does not consist of formal rules of inference. Instead the mind contains procedures that manipulate mental models (Johnson-Laird, 1983). Byrne and Johnson-Laird (1989) distinguish between one-model problems and multi-model problems, as they did for syllogistic reasoning (Johnson-Laird & Bara, 1984).

The following is a one-model spatial two-dimensional inference (Figure 2.4). The next problem is a two-model spatial two-dimensional inference (Figure 2.5).

<i>Premises</i>	<i>Model</i>
A is on the right of B	
C is on the left of B	C B A
D is in front of C	D E
E is in front of B	
Therefore,	
D is on the left of E.	

Figure 2 . 4 - Mental model representation for a one model spatial problem

<i>Premises</i>	<i>Models</i>
B is on the right of A	
C is on the left of B	C A B or A C B
D is in front of C	D E D E
E is in front of B	
Therefore,	
D is on the left of E.	

Figure 2 . 5 - Mental model representation for a multi model spatial problem

Mental models theory explains that people naturally reason by imagining the state of affairs described in the premises, drawing a conclusion from such a mental model, and searching for alternative models that might refute that conclusion. Therefore, the first spatial problem should be easier due to the fact that the description is consistent with only a single model.

Byrne and Johnson-Laird (1989) proposed that reasoners can construct a model of the situation from their knowledge of the meaning of the relational terms. In their 1989 paper, Byrne and Johnson-Laird describe two experiments to investigate how people reason about the spatial relations among objects. Their experiment used two-dimensional layouts of spatial descriptions, a deductive reasoning task based on five-term series problems, previously unstudied. The experiments tested the predictions of the inference-rule theory against those of the mental models theory.

The mental models theory predicts that problems requiring only one model of the spatial layout to be constructed should be easier than those requiring more than one model to be constructed, even when the multiple model problems have valid conclusions. Contrasted with this is the theory based on rules of inference that predicts that problems based on fewer inferential steps should be easier than problems based on more steps. The first experiment held constant the number of inferential steps specified by the inference-rule theory, but varied the number of models required to make a valid response: the one-model problems were reliably easier than the problems requiring more than one model. The second experiment contrasted opposing predictions from the two theories, and once again the results supported the model based account.

Byrne and Johnson-Laird (1989) concluded that it is easier to draw a valid spatial inference when a description corresponds to just a single layout as opposed to two or more distinct layouts. This, they say, can readily be explained if people are naturally reasoning by imagining the state of affairs described in the premises, drawing a conclusion from such a mental model, and searching for alternative models that might refute that conclusion.

- **Summary**

This section has highlighted two major theories in explaining performance on two dimensional spatial inferences. Hagert (1984) proposed that premises are represented linguistically and integrated by inference rules. This is a well specified account where the representations are verbal in nature. Byrne and Johnson-Laird (1989) proposed that premises are integrated into representations based on procedures that construct alternative models. This account makes the claim that the representations are spatial in nature.

2 . 4 REASONING AND WORKING MEMORY

Three classes of theory emerged to explain a reasoner's performance and difficulties encountered on syllogistic and spatial inference tasks. The first class of theory was based upon verbal accounts of the processes involved. These theories suggested that deductive reasoning performance can be explained by recourse to linguistic, rule-based, or comprehension processes alone (e.g., Rips, 1994; Polk & Newell, 1995; Hagert, 1984). The second class of theory was based upon the notion that reasoners construct spatial or analogical representations of deductive arguments (e.g., Johnson-Laird, 1983; Byrne & Johnson-Laird, 1989). The third class of theory suggested that deductive reasoning performance can be explained by simple heuristics, although this was only seen for syllogistic problems (e.g., Wetherick & Gilhooly, 1990; Chater & Oaksford, 1999).

The question with which this thesis is concerned is to what extent do these deductive tasks draw on the notion of a limited working memory capacity as a source of difficulty in performing them? Without recourse to some sort of temporary mental workspace, the ability to reason deductively would be very difficult (see Example 1.1, Chapter 1). What most theories do have in common is their reliance on the notion that a major factor in

problem difficulty is a limited working memory capacity. These theories would make different predictions about the role of working memory in performance of the tasks. The inference rule theorists might expect stronger links between the syllogistic and spatial deduction tasks and verbal working memory capacity tasks, rather than with spatial capacity tasks. The model theorists might expect stronger links between the syllogistic and spatial deduction tasks and spatial working memory capacity tasks, rather than with verbal capacity tasks. The heuristic accounts might expect small relationships between deductive reasoning and working memory due to the reliance on simple strategies. There is limited research investigating these types of deductive reasoning task and working memory, although a dual-task methodology has been employed to some effect.

2 . 4 . 1 Dual-Task Work

From the outline of theory above it is clear that there is no generally agreed account of how people process deductive arguments, although many of these accounts do agree on the involvement of working memory capacity as a limiting factor (e.g., Fisher, 1981; Johnson-Laird, 1983; Sternberg & Turner, 1981). Gilhooly and Logie (1998) observed that the two areas, working memory and reasoning, have been studied in relative isolation. Therefore, the question of how working memory is implicated in the performance of complex cognitive tasks needs to be addressed. This section aims to explore the role working memory plays in accounting for the performance limitations associated with these deductive tasks.

Baddeley and Hitch's (1974) early work considered the role of working memory in simple deductive tasks, such as the AB task (A does precede B, given BA - True or False?).

Articulatory suppression was used as a secondary task to load the PL of working memory.

A detriment was seen in performance on the primary AB task when the two were carried out

concurrently. Farmer, Berman and Fletcher (1986) replicated these results - articulatory suppression, when carried out concurrently with the AB task, disrupted performance on the primary task. In contrast, spatial suppression (continuous sequential tapping) produced no reliable interference when carried out concurrently with the AB task. The Farmer et al. (1986) study adds to the earlier work by comparison between secondary tasks, strengthening the view that the AB task requires the PL of working memory for accurate performance.

Dual-task work has also investigated the role of working memory in performance on conditional reasoning problems (If p then q, p, therefore what follows?). Evans and Brooks (1981) used conditional reasoning problems which were performed concurrently with articulation of a set of digits, with or without memory load. Logical performance was not impaired by the competing tasks and the latency of responding was actually faster under concurrent articulation, without memory load, than in a control group. Halford, Bain and Mayberry (1984) questioned these findings, suggesting that the load imposed by the secondary tasks used by Evans et al. (1981) wasn't great enough. Halford et al. (1984) reported studies using algebraic problems (e.g., $(7 \div 3) \div 4 = 1$) performed concurrently with a short-term memory load. An interaction was found between difficulty of the algebra problem and concurrent memory load, but the point at which interference occurred was at or above span. These results support the contentions of Baddeley and Hitch (1974) and refute those of Evans and Brooks (1981).

Toms, Morris and Ward (1993) investigated the use of all three of the working memory components in conditional reasoning. Only the tasks assumed to load the CE (verbal memory load was given, not random number generation) disrupted performance. This indicated that the CE of working memory was involved in the performance on a conditional

reasoning task. This study further strengthens the links between the use of working memory in performing conditional reasoning tasks.

Klauer, Stegmaier and Meiser (1997) also used propositional reasoning problems as the primary task, together with the following secondary tasks: articulatory suppression (PL), verbal random number generation (CE), and spatial random number generation (CE). Each participant worked through the primary and secondary tasks both alone and together.

Articulatory suppression and random number generation degraded solution accuracy on the primary tasks. This study used a baseline comparison for both primary and secondary tasks, whereas Toms et al. (1993) did not. Thus seen together, these studies further implicate the role of the CE and PL components of working memory in propositional reasoning tasks.

The major impetus of dual-task work has been in investigating the role of working memory in syllogistic tasks. Gilhooly, Logie and Wynn (1999) reviewed a programme of research on syllogistic reasoning performance and working memory components. Categorical syllogisms were used as the primary task, together with a variety of secondary tasks: random number generation (verbal and spatial) (CE), articulatory suppression (PL), and tapping (VSSP). In one experimental series, all participants carried out 20 syllogisms in control and dual-task conditions (Gilhooly, Logie, & Wynn, 1999). Response accuracy on the primary task was looked at. Latency measures were also used - the time to read and formulate a conclusion (premise processing time), and the time to indicate the chosen conclusion from a menu (conclusion reporting time). They collated evidence in relation to the role the PL, VSSP and CE play in syllogistic reasoning in participants of varying skill levels and in participants after training in syllogistic reasoning.

Higher skill participants followed a high load strategy which loaded the CE, PL, and imagery subsystems. Lower skill participants followed a less demanding strategy (matching) which did not load working memory components as heavily. This finding is consistent with an earlier study by Gilhooly, Logie, Wetherick and Wynn (1993) who found that, in untrained or unskilled participants, working memory played only a minor role in performance of syllogistic reasoning tasks. This finding was interpreted as supporting the fact that most untrained participants did not reason logically and therefore did not place heavy demands on working memory. For the trained, or high skilled participants, in the later study, training seemed to result in better performance at the expense of an additional load on the executive resources of working memory. Overall, these studies were suggested to support the view that high accuracy strategies, such as those followed by the high skill and trained participants, load the CE to a marked degree and also load the PL subsystem to a significant but lesser degree. These results indicated that varying secondary task loads can bring about changes in strategy and complement the results of Klauer, Stegmaier and Meiser (1997), who found that changing strategy (increasing strategy load) affected the impact of a fixed secondary load (tapping). Many other studies have shown similar results (e.g., Sternberg & Turner, 1981, Johnson-Laird & Byrne, 1991; Gilhooly, Logie, Wetherick & Wynn, 1993) and conclude that the role of working memory components in reasoning performance need to be recognised.

Other dual-task research has focused on spatially based primary tasks, such as the manikin task (e.g., Farmer, Berman and Fletcher, 1986), and the space fortress task (e.g., Logie, Baddeley, Mane, Donchin, & Sheptak, 1989). Oakhill and Johnson-Laird (1984) investigated four-term series spatial problems by asking participant's to draw their representations of different problems whilst performing concurrent tasks suggested to load

working memory components. Referentially continuous and discontinuous descriptions were used. The first experiment investigated the affect of carrying out visuo-spatial tracking as a secondary task - a detriment in performance on the primary task was seen. The second experiment investigated the effect of remembering six random digits as a secondary task - a detriment again was seen in primary performance, but it affected discontinuous more than continuous descriptions. The results suggest that both continuous and discontinuous descriptions rely on the VSSP, but discontinuous descriptions tax The PL to a greater extent than do continuous descriptions. These results were taken as consistent with Baddeley's notion of two slave systems.

Klauer, Stegmaier and Meiser (1997) used three-term and five-term series spatial reasoning problems as the primary task, together with the following secondary tasks: articulatory suppression (PL), tapping (VSSP) and random number generation (CE). Each participant worked through the primary and secondary tasks both alone and carried out together. Articulatory suppression and random number generation degraded solution accuracy on the primary task.

Vandierendonck and De Vooght (1997) also looked at the performance on spatial descriptions and working memory load. However, unlike Klauer, Stegmaier and Meiser (1997), a distinction was made between processing the premise information and reporting the conclusion. They used 4-term series spatial reasoning problems together with the following secondary tasks: articulatory suppression (PL), tapping (VSSP), and RIR (CE). Random Interval Repetition (RIR) looked at the average time elapsed between the presentation of a beep and the occurrence of the response of the beep. Articulatory suppression, tapping, and RIR degraded primary task performance (solution accuracy).

Only tapping and RIR affected premise processing time. The results from this research, and the other non-verbal deductive task research, suggest a clear role of working memory components, especially the CE system, in the performance of reasoning problems.

In summary, there are some consistencies between conditional, syllogistic and spatial tasks in relation to the dual-task work. They all seem to involve the CE component of working memory. The conditionals and syllogistic tasks additionally seem to involve the PL component, whereas, the spatial task additionally seems to involve the VSSP component. Additionally, a distinction was found between simple verbal and simple spatial tasks and the detriments seen in performing them concurrently with PL and VSSP secondary tasks, respectively. Taken all together, the dual-task work summarised here supports the view that the components of working memory are involved in performance on a wide range of deductive reasoning tasks. A major role can be seen for some sort of executive resource in nearly all cases of deduction reviewed here, and a lesser role for both a verbally and spatially based resource for storage of the products of that processing. The Baddeley and Hitch (1974) working memory model seems supported in nearly all the research, although each study only answered certain questions, with differing tasks, differing designs and interpretations of data.

The dual-task findings can be explained in the context of the three classes of reasoning theory summarised in Sections 2.2.3 and 2.3.3. Most of the verbally based and analogical based processing accounts of deductive reasoning would expect working memory load to disrupt performance due to explicit processing involved in the tasks. However, the verbal reasoning theory of Polk and Newell (1995) is a special processing case. They suggest that verbal reasoning is accomplished exclusively by existing comprehension and production

skills. Polk and Newell (1995) suggested that there are no special reasoning processes - not only are there no logical proof procedures, but there are also no special processes to support logical reasoning by constructing and manipulating multiple models. Therefore, it might be expected that reasoning performance would not show significant relationships with measures of an explicit system, such as working memory. However, from the summary of research investigating verbal comprehension and working memory, it can be seen that these two constructs do correlate. Thus, if this account is to be supported, this correlation must be explained. Lastly, the heuristic accounts of deductive reasoning might expect less detriments in performance on a primary task when carried out with a working memory dual-task because the reasoner is assumed to use a less demanding heuristic based strategy. There may be some support for this claim from the training/skill evidence of Gilhooly, Logie and Wynn (1999). However, there are a number of difficulties in interpreting the dual-task work and the following section highlights some of these.

2 . 4 . 2 Difficulties of Interpreting Dual-Task Work

Navon (1984) showed that none of the methods assumed to measure demand on a common resource can be known to do so. Even if there is no obvious component common to two tasks that interfere, interference may still occur because doing one task changes the conditions under which the second must be performed. If researchers are using dual-task methodology, care should be taken to look at performance on both the primary and secondary tasks. This generates a question of how to measure performance on tasks such as concurrent articulation or random generation (Wagenaar, 1970; Evans, 1978).

Baddeley (1986) points out concerns about whether the two levels of difficulty in the concurrent tasks are sufficient to produce differential effects on performance. Baddeley

(1986) also points out a second problem. That is, that participants do not invest the same amount of attention in the secondary task across all conditions. Attempts have been made to counter these problems (e.g., Evans, 1978; Bourke, 1997), although completely alleviating them can never be achieved due to the nature of secondary task methodology.

Pashler (1994) also highlighted the problems associated with dual-task methodology. He asked the question why people have trouble performing two tasks at the same time. Pashler gives three classes of explanations. The first is capacity sharing - people may share processing capacity (or mental resources) among tasks. The second explanation is bottleneck models - parallel processing may be impossible for certain mental operations. The third explanation is cross-talk models - where two alternative explanations make differing predictions. It may be easier to perform two tasks concurrently when they involve similar inputs if this meant that the same set of processing machinery could be 'turned on' and used for both. Alternatively, it may be more difficult to perform two tasks when they involve similar information. Pashler (1994) identified the need to learn about how the mechanisms apparent on these simple tasks manifest themselves in more complex activities of comprehension, reasoning and thought. Do mental processes produce a bottleneck because they require the activity of a single brain structure or ensemble of structures or because of mutual inhibition, or do they do so for some as-yet unsuspected reason? Clearly these questions need to be answered before the dual-task data can be interpreted unequivocally.

Another difficulty with the dual-task method is that the working memory model was partly built on this tradition. Therefore, any problems associated with this method are also associated with the assumptions of a tripartite system of working memory. Starting with a

specific model could presuppose the type of interpretation made of the data. Using a specific model might also lead researchers to use particular dual-tasks, at the exclusion of others, and in this way may limit the results found.

Thus, problems have been indicated in the use of the dual-task methodology - both in the carrying out of these tasks and their measurement. Navon (1984) suggested that interference found between performance on primary and secondary tasks may not be due to a common resource pool. Other factors as simple as same modality tasks may also have had an effect. In other words, interference seen on a primary task might be associated with similar materials used in the secondary task, and not due to the use of the same system of working memory. Others have criticised existing dual-task work in the way they measure the interference shown (e.g., Wagenaar, 1970; Evans, 1978).

Are there any alternatives to the dual-task approach when investigating the role of working memory in reasoning? Another approach does alleviate the problem of being encumbered by a supplementary task and involves measuring the capacity of an individual's working memory system and relating capacity to individual performance on cognitive tasks. This approach takes advantage of the individual differences that occur in cognitive abilities. Participants can be given a range of tasks, and a correlational approach can be used to identify which components of cognitive processing appear to be associated most closely with capacity to carry out that cognitive task. The dual-task and individual differences work may be complementary in advancing knowledge about the role of working memory in reasoning. In the following section reasoning and working memory will be examined from the individual differences perspective.

2 . 4 . 3 Individual Differences in Reasoning and Working Memory Capacity

A question that remains from the research carried out on syllogistic and spatial inferences is how individual differences in performance on these tasks is explained. Stanovich and West (e.g., 1998) have studied differences that can be seen between the normative and descriptive accounts of human reasoning. These accounts refer to what would be expected due to traditional logic, as opposed to what is actually seen in practice. What Stanovich and West (1998) suggested is that these individual differences and their patterns of covariance might have implications for explanations of why human behaviour often departs from normative models. Stanovich and colleagues refute the claim that the gap between normative and descriptive explanations represent instances of human irrationality. Stanovich and West (1998) examined the performance errors hypothesis across a wide range of tasks from the reasoning and decision-making literature. For certain classes of task they found significant cross-task correlations. This suggested that participants giving the normative response on one task were usually significantly more likely to give it on another.

Stanovich and West (1998) investigated the extent to which measures of cognitive ability and thinking dispositions can predict discrepancies from normative responding on a variety of deductive tasks. These reasoning tasks included the selection task, belief bias in syllogistic reasoning, argument evaluation, and statistical reasoning. These tasks were given together with cognitive ability tasks. Verbal and mathematical SAT scores were reported, together with the Raven Matrices, and reading comprehension scores which were all combined to give a cognitive ability composite score. A thinking dispositions questionnaire was also completed. The questionnaire was made up of various items across a number of scales including actively open-minded thinking, counterfactual thinking, absolutism, dogmatism, paranormal beliefs and social desirability response bias. The

authors used the cognitive ability composite score as a measure of cognitive capacity. The authors used the thinking dispositions composite score as a measure of thinking style (a high score indicating that the individual was open-minded, cognitively flexible and had a sceptical attitude).

If the reasoning tasks correlated positively with these measures of cognitive ability then it might justify the use of the normative model being employed to evaluate performance.

Their results showed that indeed, cognitive ability measures did correlate with performance on a wide variety of deductive and problem solving tasks. Thus, the authors concluded that capacity limitations influenced the actual performance on the deductive tasks, as judged against normative criteria. It was also found that the thinking dispositions composite score was significantly correlated with each of the four reasoning tasks, although the correlations were smaller in magnitude than for those seen with cognitive ability. It was also found that there were significant correlations between the reasoning tasks, apart from one (between argument evaluation and statistical reasoning). Although the highest correlation was between the syllogistic and selection task performance (.36), correlations almost as strong were obtained between tasks deriving from the deductive literature (syllogistic reasoning and selection task) and the inductive reasoning literature (statistical reasoning). Jointly, cognitive ability and thinking dispositions accounted for a moderate amount of variance in overall rational thinking performance, although cognitive ability was a stronger unique predictor. Stanovich and West (1998) concluded that cognitive ability measures capacity, and that thinking style is a measure of what participants do with that capacity.

This notion has implications for patterns of individual differences across reasoning tasks (Stanovich, 1999). Stanovich (1999) suggests that if it were only performance errors which

accounted for the differences between the normative and descriptive accounts, there should be no significant correlations between disparate reasoning tasks. The positive relationships obtained between tasks indicated that departures from normative responding on each of them were due to systematic limitations in processing and not to non-systematic performance errors.

Two alternative ways of studying the relationship between working memory and reasoning have now been outlined. The first relied on the notion that if detriment is seen on a primary task when it is carried out concurrently with another, presumed to load a working memory subsystem, then they are sharing this common resource. The second relied on the notion that variation in performance on deductive reasoning tasks can be predicted by tasks suggested to measure the capacity of working memory subsystems. Both of these accounts are consistent with most of the processing theories of reasoning reviewed earlier in this chapter. These accounts can not be reconciled with those theories suggesting a heuristic explanation of the data. They suggest that implicit processing alone accounts for the reasoning performance. However, it is important to consider that other individual differences may exist, other than capacity limitations, that explain variations on reasoning tasks. Indeed, there is some evidence in the literature that the strategies people adopt may explain differences in performance.

2 . 4 . 4 Strategy Effects

Theorists have used different explanations as to why some people deviate from a normative logical response. Rule theorists propose that easy valid inferences are made by accessing a corresponding elementary inference rule. Difficult valid inferences have no corresponding inference rule, thus reasoners must construct a proof of several steps to derive the

conclusion. Mental models theory proposes that easy valid problems rely only on an initial set of models to be combined. More difficult problems require the initial set of models to be combined and fleshed out to search for alternative models. Roberts (1993) states that

If a theory of reasoning is being proposed that is intended to describe the processes used by all people for all reasoning tasks, then what is the status of this theory if it is subsequently found that not all people are using the same processes?

The existence of individual differences in strategies used to solve deductive reasoning problems casts doubt on the status of theories which propose general underlying cognitive processing. It may be that theories must account for the times at which people draw on short-cuts or different strategies in order to solve deductive reasoning tasks, and not just concentrate on the process the majority are shown to follow. Individuals may adopt different strategies that are verbal or spatial in nature and they may draw on working memory resources in different ways. There are many studies that have investigated what these different strategies are and how they may be accounted for by theory. What follows is an examination of the literature suggesting that individuals may rely on linguistic and/or spatial processes or strategies.

For instance studies have investigated strategy use for performing the picture verification task. Carpenter and Just (1975) produced data to suggest that participants solved these problems using a deduction rule strategy. Although, MacLeod, Hunt and Mathews (1978) found that a section of their participants appeared to be using a strategy based upon mental models. This finding was also replicated by Marquer and Pereira (1990). This research suggests that a combination of processes may be called upon when solving these problems.

Another example of this type of research investigated syllogistic deductions. Sternberg and Weil (1980) identified people that spontaneously used mental models strategy, deduction rule strategy, and a mixture of the two (the majority of participants). Galotti, Baron and Sabini (1986) claimed to have found evidence that formal rules could explain certain aspects of the performance of their participants in a syllogistic reasoning task. They found that some participants -- generally the ones classified as "good" reasoners -- started formulating their own rules as the task progressed. This is perhaps best classified as "strategies" that participants picked up and not formal rules present prior to the task. This, then, is not convincing evidence to support the formal rules approach. Ford (1994) also looked at strategy use when solving syllogistic problems. One group represented the relationship between classes in a spatial manner that was supplemented by a verbal representation. The other group used a primarily verbal representation. This research again lends support to the view that both linguistic and spatial procedures may be used in the pursuit of accurate performance.

There have been studies within the area of spatial reasoning tasks showing that people solve these problems by using mental models or deduction rules (e.g., Lohman & Kyllonen, 1983). Many studies have shown that individuals may progress from one strategy to another as they gain expertise (e.g., Johnson-Laird, 1972; Trabasso, Riley & Wilson, 1975). Wood, Shotton and Godden (1974) suggest that participants may abandon flexible strategies early on in experiments in favour of a quicker, short-cut strategy suited to only certain problem types. This, in effect, shutting out other possible solutions due to strategy limitations. In addition, Gilhooly, Logie and Wynn (1999) suggested that low skill participants used a simple strategy, whilst high skill participants used a more demanding one.

Other studies have shown that different individuals may use different strategies in relational reasoning (e.g., Ohlsson, 1984). Egan and Grimes-Farrow (1982) interviewed participants post three-term series spatial deductions, and revealed striking individual differences in the way problem representations were described. Two contrasting strategies were identified. The “concrete properties” thinkers concentrated on the physical properties of the objects in the description. The “abstract directional” thinkers concentrated on directional representations for every relation. Egan et al. (1982) suggest that different strategies were used, so a theory of reasoning must account for why individuals adopt one strategy over another.

Still other studies have suggested that people may use a combination of spatial and linguistic processes when solving spatial descriptions. For instance, Mani & Johnson-Laird (1982) suggest that their results indicated the existence of two types of encoding -- propositional representations and mental models - propositional representations that are relatively hard to remember but correspond closely to the sentences in the description, and mental models that are relatively easy to remember but are analogous to spatial arrays and accordingly poor in linguistic detail.

This section has highlighted the problem that individuals might use different strategies when performing deductive reasoning problems. It is important to note that these strategy effects exist and may occur with all the problems to be used in this thesis. It is important to bear these effects in mind when looking at the patterns of data produced on all the reasoning and working memory span tasks used in this series of experiments.

The focus of this thesis looks at the role of working memory from an individual differences perspective. This is because, although major links have been drawn between working memory and reasoning via dual-task methodology, no work looks at working memory capacity tasks and reasoning/deduction via correlational means. The next section aims to provide a rationale behind an individual differences approach, based on a general look at work in this area.

2 . 5 RATIONALE FOR THESIS

REASONING WITH A LIMITED WORKING MEMORY CAPACITY -

INDIVIDUAL DIFFERENCES WORK

Chapter 2 started with a summary of the phenomenon associated with two types of deductive reasoning task - syllogistic and spatial inference. It was found that for both types of problem, reasoners can answer according to normative criteria.

Figural difficulty and conclusion preferences were also found to affect both types of task. The believability of the conclusion was another phenomenon that needed to be addressed by any theory of reasoning.

As previously discussed, there have been attempts to investigate how individuals process and store deductive reasoning tasks. There are a wide range of explanations available in the literature suggesting that no one theory can account for all the effects found when individual's solve these types of problem (e.g., Rips, 1994; Johnson-Laird, 1983; Polk & Newell, 1995, Chater & Oaksford, 1999). The theories put forward to explain the phenomenon associated with deductive reasoning tasks were separated into three classes. The first class loosely proposed a verbal, or rule-based, or comprehension account (e.g., Rips, 1994; Polk & Newell, 1995). The second class proposed that the reasoner relied on

spatial or analogical representations (Johnson-Laird, 1983). The third class of theory loosely fitted into one that proposed the reasoner relied on simple strategies (e.g., Chater & Oaksford, 1999). These theories make differing predictions about the role of working memory in the performance of deductive tasks.

For instance, Johnson-Laird (1983) proposed mental models as the mechanism by which people solve deductive tasks, whereas Rips (1994) proposed propositionally based rules as most important. What all these theories have in common (i.e., rules, mental models) is their reliance on the notion of some explicit system that affords the reasoning process. Thus, it might be that working memory predicts performance on deductive tasks. The verbally based theories (i.e., rules) might expect verbal working memory capacity to predict reasoning performance. The verbal reasoning theory (Polk & Newell, 1995) can be considered a special case due to the following explanation. The process of reasoning is suggested to rely upon the same processes as used in verbal comprehension. However, considering the fact that these two processes, reasoning and verbal comprehension, have been found to correlate, the same predictions might be made. as for the other verbally based accounts. The spatially based accounts (i.e., mental models) might expect spatial working memory to predict performance on deductive reasoning tasks. In contrast to these theories, the heuristic based accounts make no claims about underlying explicit processing (e.g., Wetherick & Gilhooly, 1990, Chater & Oaksford, 1999). These theories proposed that reasoners use very low load strategies, which require little in the way of processing. Thus, they might not expect reasoning performance to be highly correlated with working memory capacity.

Many studies have investigated the role of the working memory subsystems in the ability to reason deductively. These studies have been based on dual-task methodology. The major finding was that for every study, except one, working memory was implicated in all the deductive tasks investigated. This included tasks such as syllogisms, spatial inference, conditionals, as well as more simple cognitive tasks. The component of working memory that was most highly implicated as a resource in performing complex deductive tasks was the CE. When the deductive tasks were carried out simultaneously with random number generation, assumed to measure the capacity of the CE, a detriment was seen in primary task performance. Therefore, the two concurrent tasks were assumed to be sharing the CE as a common resource.

The review then went on to criticise the dual-task approach, suggesting problems when two tasks are carried out concurrently. Whenever two tasks are carried out simultaneously, problems will arise with measuring the trade-off between tasks. Problems were also highlighted in the difficulties associated with measuring the performance on tasks such as random number generation.

Individual differences was then introduced as an alternative way of looking at the role of working memory in reasoning, and at the same time would alleviate the problems associated with the dual-task work. Stanovich and colleagues used ability tests, not working memory tasks, as measures of cognitive capacity. They also used thinking disposition as an additional measure of an individual's style of thinking. Both cognitive ability and thinking dispositions were predictive of

performance on a wide range of deductive tasks. As seen in Chapter 1, the use of individual differences in complex cognition has proven fruitful again.

It is generally assumed that working memory capacity is fixed, but that capacity is likely to fluctuate across individuals (e.g., Just & Carpenter, 1992; Shah & Miyake, 1996). The standard method of measuring working memory capacity is exemplified by Daneman and Carpenter's reading-span task, as summarised in Chapter 1. People must carry out the maintenance of a set of last words whilst reading sentences in a set. This task, as previously discussed, requires both the storage and processing of information afforded by an executive resource of working memory. Studies show that cognitive tasks are positively related to a multitude of working memory capacity measures such as these (e.g., Kyllonen & Christal, 1990; Turner & Engle, 1989; Jurden, 1995).

Chapter 1 also summarised research related to the dichotomy between the unitary/separate resources issue of the central executive component of working memory. There is no agreement as to whether differing types of information are dealt with by similar or contrasting central resources of working memory. Some have assumed that all information is processed by the same executive (e.g., Turner & Engle, 1989), whilst others have assumed the need for separate resources for verbal and non verbal information (e.g., Daneman & Tardif, 1987; Jurden, 1995; Shah & Miyake, 1996). This thesis tries to answer this dichotic dilemma by using verbally and spatially based working memory span and deductive reasoning measures.

There has been no specific research investigating working memory capacity measures and complex deductive reasoning tasks, such as syllogisms and spatial

inferences. This thesis attempts to rectify this oversight by using capacity measures together with complex syllogistic and spatial relation tasks. This thesis will examine the pattern of covariance between a range of working memory measures and performance on deductive reasoning tasks, with particular focus on syllogistic and spatial inference.

As we have seen, some deductive tasks may rely more on spatial than verbal processing, and vice versa. The series of experiments in the following chapter uses syllogistic tasks (verbal) and spatial relation tasks (spatial), together with working memory capacity measures, in order to examine the relationship between the two and to potentially differentiate between reasoning theories. The distinction between verbal and spatial deductive tasks will also allow the investigation of the separability of CE resources of working memory (Shah and Miyake, 1996). Thus, the first set of experiments uses a range of working memory capacity measures, proposed to measure both storage and processing for verbal and spatial resources, together with verbal and spatial deductive tasks. Correlational and factor analyses will be performed on the data in order to determine the relationships between the capacity and reasoning measures. An in depth introduction of the rationale behind the first three experiments follows.

CHAPTER 3 - WORKING MEMORY WITH SYLLOGISTIC AND SPATIAL INFERENCE

3.1 GENERAL INTRODUCTION FOR EXPERIMENTS 1-3

Chapter 3 presents three experiments designed to investigate the relationship between working memory and performance on two deductive reasoning tasks - spatial and syllogistic inference. As of yet, there are no studies in the literature that have investigated the relationship between individual differences in deductive reasoning and working memory capacity. The first set of experiments aims to do this by using a range of working memory capacity measures, proposed to measure people's ability to simultaneously store and process spatial and verbal information, together with spatial (five-term series) and verbal (syllogistic) deductive reasoning tasks.

As could be seen from the first two chapters, the role of working memory in reasoning has previously been investigated using dual-task methodology. Studies using syllogistic inferences as a primary task identified differing detriments in performance when carried out with secondary tasks assumed to load the CE, in comparison to secondary tasks proposed to load the two slave systems of working memory (Gilhooly, Logie, Wetherick & Wynn, 1993; Gilhooly, Logie & Wynn, 1999). Gilhooly et al. (1993) investigated verbally and visually presented syllogistic problems, and in a further set of studies (e.g., Gilhooly et al., 1999), syllogistic performance was studied for low and high skilled, as well as trained participants. It was found that allowing the premises to be available on screen at all times (visual condition) leads to higher performance rates. This was presumed to be due to the reduction in verbal working memory load. The results also suggested that, in untrained or low skilled participants, working memory played only a minor role in performance of syllogistic

reasoning tasks. This was interpreted as indicating that the participants relied on low demand strategies that did not place heavy demands on working memory. However, there was a small effect of random number generation for these participants which is suggested to show some role of executive processes in the implementation of the matching heuristic. In the more recent studies, participants who performed poorly were largely unaffected by the secondary tasks, consistent with earlier findings. However, those participants who had produced the highest scores following training were especially vulnerable to the disruptive effects of random generation. This result was consistent with the idea that training results in better performance at the expense of an additional load on executive resources of working memory. This is suggested to support the view that performance on syllogistic inferences, especially for trained participants, relies on executive resources of working memory.

The work investigating spatial inferences as the primary task also identified detriments in performance when carried out with secondary tasks assumed to load the different components of working memory (Klauer, Stegmaier & Meiser, 1997). Klauer et al. (1997) used three-term and five-term series spatial reasoning problems and found that both articulatory suppression (PL) and random number generation (CE) degraded primary task performance. Vandierendonck and De Vooght (1997) investigated spatial reasoning with a concurrent task assumed to load the VSSP of working memory - tapping. It was found that the three concurrent tasks assumed to load the PL, VSSP and the CE degraded accurate performance on the four-term series spatial inferences. These studies suggest that a passive verbal and spatial store as well as an active processing component of working memory play a role in performance on spatial inferences.

The common finding across the experiments investigating complex deductive tasks and working memory is that performance is degraded on these primary tasks by secondary tasks assumed to load the CE component of working memory. The major finding for syllogistic inference was that verbal working memory plays a role in performing them, involving both CE, and to some extent PL, components of working memory. The major finding for spatial inference was that spatial working memory also plays a role in performing them, involving CE, PL, and VSSP components of working memory. These studies demonstrated that performance on these different deductive reasoning tasks depends on different sub-systems of working memory, although the CE is assumed to be the most heavily implicated system across both tasks. Thus, an aim of this thesis is to use the syllogistic and spatial deductive tasks in order to determine the extent to which these tasks draw upon different working memory resources.

It was noted in Chapter 2 that dual-task studies were difficult to interpret. Some of these problems concerned the use of this methodology in specifying the degree of CE involvement in complex cognitive tasks. For instance, Pashler (1994) argued that a bottleneck can occur in dual-task situations at the response selection phase (i.e., two responses can not be selected at the same time) even when other perceptual, cognitive, and motor processes can co-occur. In addition to this, there are problems associated with the strategic trade-off between primary and secondary task performance. This type of strategic trade-off is a source of concern, particularly when the supposedly secondary task is as complex and demanding as random number generation. These considerations suggest that when applied to the CE, the simple dual-task logic might not always hold.

An alternative approach to examining the role of working memory in complex cognitive tasks is from the perspective of individual differences. Many studies have investigated the relationship between measures of working memory capacity and performance on other cognitive tasks (Kyllonen and Christal, 1990; Shah & Miyake, 1996). In the context of this work, many measures of working memory capacity have been developed. These measures can be mapped onto the different subsystems of the working memory model. For instance, reading comprehension was best predicted by complex measures of capacity that involve both a processing and storage component, measures that are assumed to reflect the capacity of the CE (Daneman & Carpenter, 1980; Shah & Miyake, 1996). It might be expected that these complex measures of working memory capacity, assumed to reflect CE processing, will predict performance on both the syllogistic and spatial reasoning tasks. Other working memory capacity tasks, such as the simple word and arrow spans (e.g., Shah & Miyake, 1996), are assumed to reflect PL and VSSP storage, respectively. Given the dual-task work, it might be expected that simple measures of PL storage, such as the word span, will correlate with syllogistic reasoning and spatial reasoning, but measures of VSSP storage, such as the arrow span, may only be expected to predict spatial reasoning.

Carrying on in the individual differences vein, many researchers have investigated the issue of the separation of executive resources in working memory for information processing activities. Some of these researchers proposed that all cognitive information processing is afforded by a singular, general executive resource (Turner & Engle, 1989). Other researchers have identified separate central resources of working memory for verbal and non verbal information (Jurden, 1995; Shah & Miyake, 1996). Therefore, another aim of this thesis is to further investigate the separability claim of working memory resources for deductive reasoning tasks, as yet underrepresented in this field. Many working memory

capacity measures of CE resources have now been developed. Some of these are proposed to rely on verbal processing and storage components (Daneman & Carpenter, 1980). whilst others have been proposed to rely on spatial processing and storage components (Daneman & Tardif, 1987; Shah & Miyake, 1996). By using all of these capacity measures in conjunction with syllogistic and spatial reasoning tasks, it may be possible to identify which reasoning tasks rely on which type of working memory resource.

Another aim of this thesis is to examine the implications of patterns of relationship between working memory and reasoning for theories of deductive reasoning performance. This series of experiments will look at patterns of covariance between the working memory and reasoning measures in relation to the explanations made by different reasoning theory.

Three classes of reasoning theory were introduced earlier, one class relying on verbal processes, another on spatial/analogical processes, and a third on heuristic processes. These reasoning theories all give differing accounts as to the representation of deductive tasks in the mind. Given the Shah and Miyake (1996) framework, these classes of theory might make different predictions about the degree to which different deductive reasoning tasks involve verbal versus spatial working memory capacity. The theories that assume the use of verbal processing and representation (Rips, 1994; Hagert, 1984) might predict that both syllogistic and spatial reasoning performance be predicted by verbal working memory capacity. (It should be noted that Rips (1994) only studied propositional and syllogistic tasks; and Hagert (1984) only studied spatial inferences. Therefore, any predictions about specific deductive reasoning tasks are taken in this context). However, Polk & Newell's (1995) Verbal Reasoning theory, that involves comprehension, suggests that people do not have special procedures for processing deductively, and that this process is not resource

intensive. Therefore, this theory might not expect a relationship between reasoning performance and working memory capacity. However, reading comprehension does correlate with working memory capacity (Daneman & Carpenter, 1980) and therefore, based on the available evidence, it might be expected that these patterns of relationship are observed.

Those theories that assume the use of spatial processing and representation (Johnson-Laird, 1983) might predict that spatial working memory predicts reasoning performance, although Johnson-Laird (1983) acknowledges the possibility that mental models contain information that could be propositional in nature. However, what processing accounts do have in common, for instance rules and mental models, is the notion of a mental workspace in which to carry out computations.

The theories that assume a heuristic account of the reasoning process (Wetherick & Gilhooly, 1990; Chater & Oaksford, 1999) have only been applied to syllogistic reasoning performance. In contrast to the processing theories, these accounts do not draw on the notion of an explicit processing system having a functional role in reasoning. They only claim a minor role for working memory in reasoning of this kind. Therefore, there is no reason to expect that measures of the explicit system (working memory spans) would predict logical performance on the syllogistic task.

Because this series of experiments investigates both verbal (syllogistic) and spatial (five-term series) reasoning and working memory capacity measures, this distinction can be used to identify and differentiate between the reasoning theories. Studies investigating relational reasoning have shown that a spatial representation is invoked by the content of the premises (Byrne & Johnson-Laird, 1989). It is not so clear from the syllogistic reasoning work

whether performance depends upon spatial or verbal representations (Ford, 1994). An aim here is to identify which working memory resources are used on these deductive reasoning problems.

In summary, the primary aim of this set of experiments is to investigate the degree to which individual differences in working memory can explain the variation in performance on syllogistic and spatial inference tasks. The secondary aim is to replicate the finding of a dissociation of spatial and verbal central executive resources reported by Shah and Miyake (1996). The final aim, given this dissociation, is to examine the degree to which the reasoning tasks rely on verbal versus spatial working memory.

Experiment 1 in this series uses measures which include simple and complex verbal and spatial working memory span measures (Shah & Miyake, 1996; Daneman & Carpenter, 1980; Daneman & Tardif, 1987), together with verbal and spatial reasoning tasks (syllogistic and five-term series spatial inferences) and ability (Brown, Bennett & Hanna, 1981; Smith & Whetton, 1988) measures. This allows an assessment of the relationship between deductive reasoning, working memory and a range of spatial and verbal abilities. In Experiment 1 the deductive tasks are presented verbally to participants, requiring them to keep in mind the premise information during the problem solving process. An additional aim in this chapter was to investigate whether modality differences in presentation influenced the patterns of correlation with the working memory measures observed. Presentation modality and working memory capacity were further investigated in Experiments 2 and 3.

3 . 2 EXPERIMENT 1

3 . 2 . 1 Aim

Experiment 1 was designed to assess the relationship between the following information processing activities - the aptitude for verbal and spatial reasoning, together with the capacity of working memory for verbal and spatial material. Five working memory span tasks were used - two complex spatial measures, one complex verbal measure, and a simple spatial and simple verbal measure (Daneman & Tardif, 1987; Shah & Miyake, 1996; Daneman & Carpenter, 1980). The complex span measures involve simultaneous processing and storage (CE), whilst the simple spans involve passive storage (PL and VSSP). The span tasks were given together with two reasoning tasks, syllogistic and five-term series spatial inferences. Tests of verbal and spatial ability were also completed (Brown, Bennett & Hanna, 1981; Smith & Whetton, 1988).

Experiment 1 investigated the degree to which individual differences in working memory capacity can explain variation in performance on syllogistic and five-term series spatial tasks. Experiment 1 also addressed the separability issue of central resources within working memory. Experiment 1 follows the same types of predictions made by Shah and Miyake (1996), but based on deductive reasoning problems. Thus, a dissociation might be expected between the measures of verbal and spatial working memory. Given this dissociation, the degree to which syllogistic and spatial inferences involve verbal versus spatial working memory resources can be investigated.

Another important application of looking at a range of working memory capacity measures and deductive reasoning tasks is in relation to reasoning theory. Given the Shah and

Miyake (1996) separability hypothesis. the three classes of theory, introduced earlier, might make different predictions about the relationships between capacity and reasoning.

Experiment 1 will investigate these working memory and complex cognitive tasks in relation to the verbal, spatial/analogical, and heuristic accounts of reasoning.

3 . 2 . 2 Method

Pilot Study

The working memory span tasks were based on measures reported in the literature (e.g., Shah & Miyake, 1996). A series of computer programs were developed in Visual Basic. A pilot study was run to ensure that these tasks were easy to understand, use, and at a reasonable difficulty level. There were alterations made to certain features of some of the span tasks. These adjustments will be explained within the procedure for each individual measure that follows. The tasks were given to a number of participants, and their verbal accounts of the tasks enabled these adjustments to be made.

Participants

Forty-six participants took part in this study - 18 of whom were men and 28 of whom were women (mean overall age = 25). The participants were undergraduate students and postgraduate students at the University of Plymouth and they received course credit or cash payment for participating. None of the sample had prior training in logic and all were native English speakers.

Procedure

Experiment 1 was carried out in two testing sessions per participant, the sessions being of approximately equal length. Participants signed up in a time slot on an experimenter sheet

for session one. Session two was arranged individually with each participant at the end of session one. Session one consisted of five tasks. All participants carried out these tasks in the following order (where an asterix (*) indicates a computerised task):-

spatial ability measure,

simple verbal word span*,

simple spatial arrow span*,

complex verbal sentence span*,

complex spatial letter span*.

Session two consisted of four tasks. All participants carried out these tasks in the following order:-

spatial inference measure (verbal presentation),

syllogistic inference measure (verbal presentation),

complex spatial tic-tac-toe span,

verbal ability measure.

What follows is a summary of each task used in Experiment 1.

Simple Verbal Word Span (word)

Procedure

All four computer based span tasks were developed in Visual Basic 4.0. The simple word span task was based on one used by Shah and Miyake (1996). The span task was designed to measure a participant's relatively passive storage capacity for verbal information. There was no explicit processing requirement. The words were chosen from the Oxford Psycholinguistic Database (Quinlan, 1992). The words were all two syllables long and were concrete nouns. All the words were checked for frequency in Kucera and Francis's (1970) corpus of over one million words. The frequency of words ranged from 5 to 216 in their

corpus. None of the words in a particular set began with the same letter (see Appendix A1.1 for a full list of the words used). Participants were instructed to remember a set of these words that appeared on a computer screen. A single trial consisted of the individual presentation of a set (two to seven) of words on the computer screen. Each item appeared on the screen for 800 ms, with an interstimulus interval of 50 ms. At the end of each trial, a light bulb appeared on the screen, indicating that the participant was to recall the to-be-remembered words. The participants gave written responses to the word sets. There were five sets at each set size, from two to seven, for a total of 30 sets, presented in the order of increasing set size. The participants also received practice sets of words at the two-word set level. An example of a set of two words given to the participants is 'wallet', followed by 'coffee' (see Appendix A1.2 for a full set of instructions).

Scoring

The participant's simple word span score was defined as the maximum level at which a participant correctly recalled all the words in the correct order for at least three of five trials. If the participant was accurate in two of the five trials at the next level, the score was incremented by half a point. In relatively rare cases in which a participant successfully recalled less than three of the five sets at a particular level but was able to recall two or more sets at a higher level, the average of the lower and the upper limits was used as the span score. The maximum score obtainable for the span measure was 7. A global score was also calculated for this task as follows. If a participant was accurate on a set, regardless of level, they received a point for each correctly recalled word (e.g., a participant gets four of the 2-sets correct and three of the 3-sets, global score = $((4*2) + (3*3) = 17)$). The maximum score obtainable for the global measure was 135.

Simple Spatial Arrow Span (arrow)

Procedure

This task was based on one used by Shah and Miyake (1996). The span task was designed to measure a participant's relatively passive storage capacity for spatial information. There was no explicit processing requirement. A set of arrows (two to six) was presented on the screen with each arrow in one of eight possible orientations. Each arrow remained on the screen for 1000 ms, with a short delay (250 ms) between the presentation of arrows. Following the presentation of a set of arrows, a participant's task was to identify the directions that the arrows were pointing in on a diamond shaped grid that appeared on the screen after every set. The grid appeared on the screen with eight square "buttons" indicating the eight possible orientations. The task of the participant was to use a mouse to click in the direction that each arrow was pointing in the order of appearance. There were three trials at each level, from two to six arrows, for a total of 15 trials. There were no practice sets before commencing the task. Following is an example of set of arrows at the two arrows level. The first arrow appears on the screen as shown in Figure 3.1. Then a second arrow appears on the screen as shown in Figure 3.2. Following the presentation of the two arrows, a recall grid appears on screen onto which the participant mouse-clicks the correct orientation of each arrow in the set in the order in which they were presented, as shown in Figure 3.3.

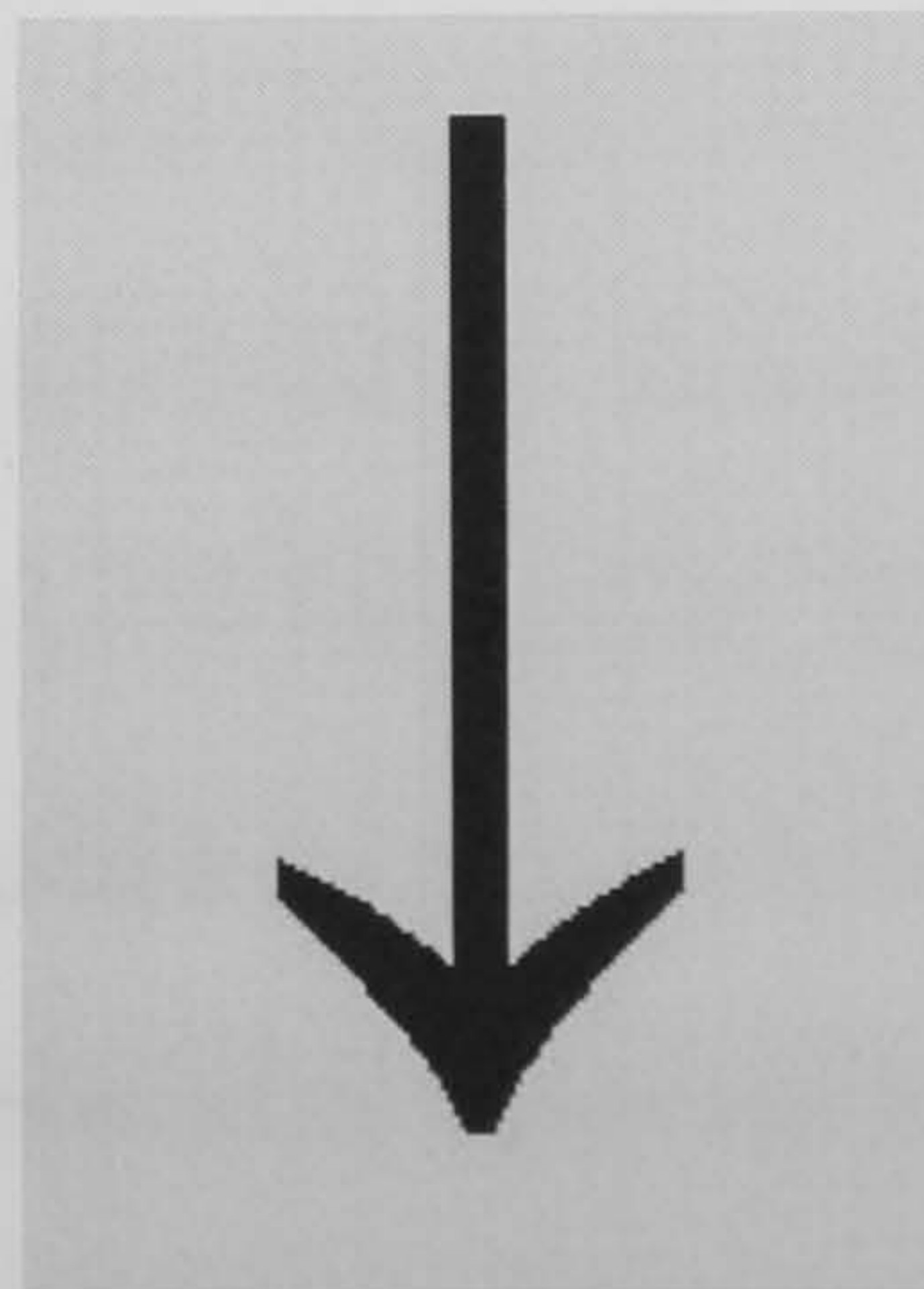


Figure 3 . 1 - Arrow span, representation of arrow 1



Figure 3 . 2 - Arrow span, representation of arrow 2

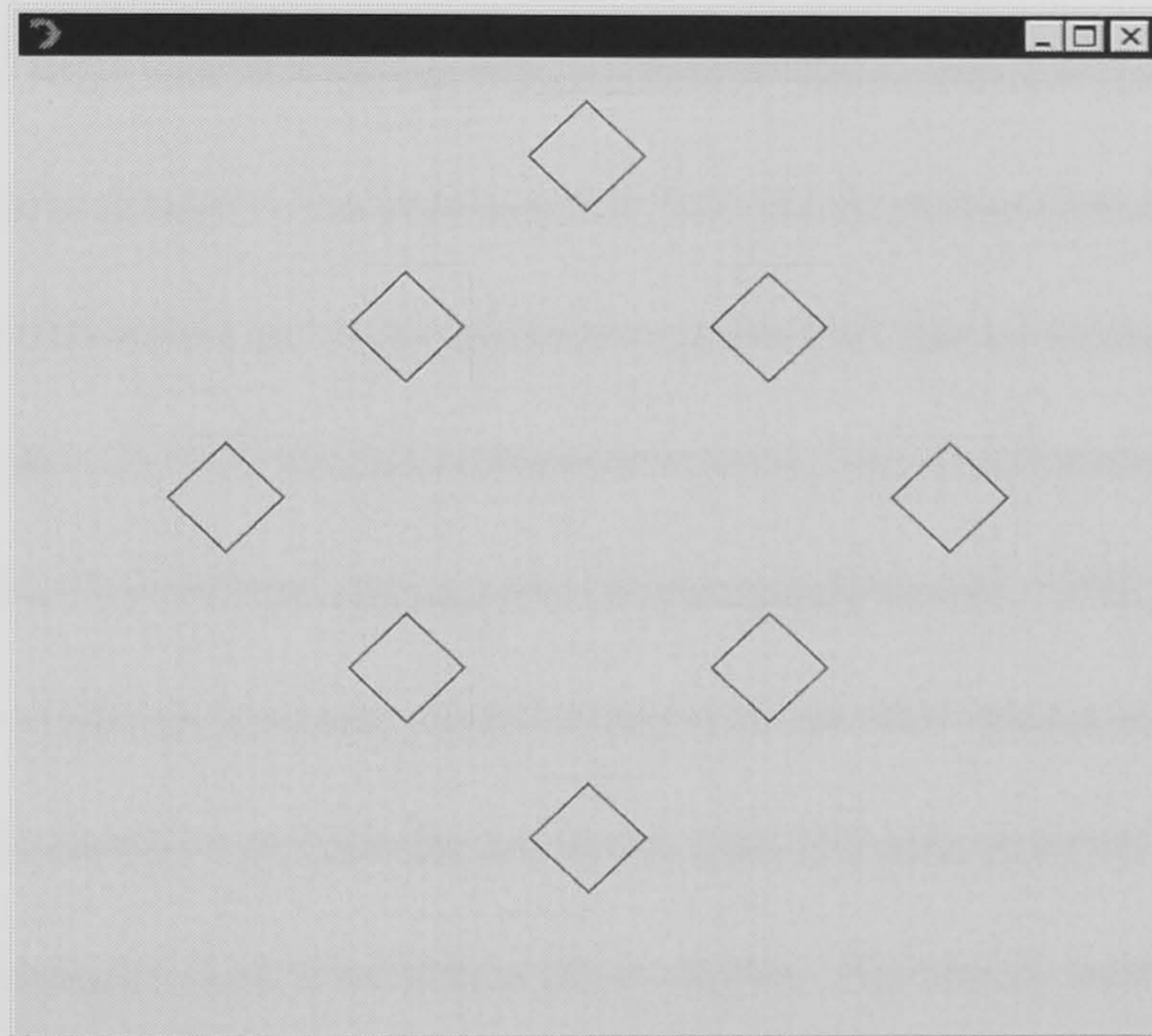


Figure 3 . 3 - Arrow span, recall grid for arrows

Scoring

The participant's simple arrow span score was defined as the maximum level at which a participant correctly recalled the spatial orientation of all the arrows in the correct order for at least two of three trials. If the participant was accurate in one of the three trials at the next level, the score was incremented by half a point. If a participant did not follow this pattern of scoring, the average was taken of their lowest and highest correctly recalled set. In other words, any discrepancy between the upper and lower limits was solved by taking the average of them. The maximum score obtainable for the span measure was 6. A global

score was also calculated for this task. The maximum score obtainable for the global measure was 60. If a participant was accurate on a set, regardless of level, they received a point for each correct arrow (e.g., a participant gets two 2-sets correct and one 3-set, global score = $((2*2) + (1*3) = 7)$).

Complex Verbal Sentence Span (verbal)

Procedure

This task was based on the reading span task developed by Daneman and Carpenter (1980). The verbal sentence span task was designed as a span measure of functional working memory capacity for language. The task requires the simultaneous maintenance and processing of verbal information. A single trial consisted of the individual presentation of a small set of true or false sentences on a computer screen. The end words of each sentence were chosen from the Oxford Psycholinguistic Database (Quinlan, 1992). The end words were all one or two syllables long and were concrete nouns (for instance, 'morning', 'line'). Word frequency was checked as seen for the word span (Kucera & Francis, 1970). The frequency of words ranged from 5 to 216 in their corpus. Sentences were allocated randomly as true or false prior to the start of the experiment and remained in the same order throughout. Except for the two set size, all sets contained a mixture of true and false sentences. An example of a set of two sentences given to the participants is 'breakfast is eaten in the morning', followed by 'a ruler draws a straight line'. In this instance, both sentences were true and the words to be remembered were 'morning' and 'line'. Overall, there were as many true as false sentences in the sets. None of the end words in a particular set began with the same letter (see Appendix A1.3 for a full list of the sentences used). The sentence remained on the screen for 800ms, after which two boxes appeared marked 'true' and 'false'. The participants task was to respond to the sentence by pressing either the true

or false button using a click of the mouse. The buttons appeared on the screen for a maximum of 2000ms. The next sentence in the set appeared on the screen 250ms after the response boxes disappeared. At the end of each set of sentences a light bulb appeared on the screen indicating that the participant must now recall the final words of that particular set of sentences. A response sheet was provided for their answers. The participants were instructed to click the light bulb with their mouse to move on to the next set of sentences. After an entire set of sentences was presented, the participants task was to recall the final word in each sentence, in the order in which they had seen them. The verbal span task included a total of 25 sentence sets, 5 sets at each level ranging from two-sentence to six-sentence sets, and participants were presented with increasingly longer sets of sentences. The participants also received practice sets of sentences at the two-sentence set level (see Appendix A1.4 for a full set of instructions).

Scoring

The participant's complex verbal span score was defined as the highest set size for which all of the final words of the sentences were recalled in the correct sequence, for at least three of the five sets. Half a point was added to the score if the participant's recall was accurate on two of the five sets at the next set size. The maximum score obtainable for the verbal span measure was 6. A global score was also calculated for this task as follows. If a participant was accurate on a set, regardless of level, they received a point for each correctly recalled final word (e.g., participant gets four 2-sets correct and three 3-sets, global score = $((4*2) + (3*3) = 17)$). The maximum score obtainable for the global measure was 100. The number of errors made in the verification component of the task (true/false judgement) was also collected (verbal error).

Complex Spatial Letter Span (letter)

Procedure

This task was based on one used by Shah and Miyake (1996). The spatial task was designed as a span measure of functional working memory capacity for spatial information. The task requires the simultaneous maintenance and processing of spatial information. A single trial consisted of the individual presentation of a small set of normal or mirror-imaged letters (F, J, L, P, or R) on a computer screen. The same letter was used throughout a set, and the participant was told which letter would appear on the screen in that particular trial. Each letter was presented in one of seven orientations (in 45° increments, excluding the upright orientation), and each of the 70 possible combinations (letters x orientations x normal/mirror-image status) appeared once in the task. The presentation of letters was constrained such that opposing orientations were not presented successively within a set and that the same orientation could appear only once in the same set. Shah and Miyake's (1996) design was altered slightly due to the problems found in the pilot study. The task differs from Shah and Miyake's (1996) task where stated below. The participant was asked to respond to whether the letter was 'normal' or 'mirror-imaged' by clicking the mouse in the appropriate box on the screen - Shah and Miyake's (1996) participants responded aloud and a trained experimenter then recorded their response. The letter remained on the screen for a maximum of 1000ms after the participant had pressed one of the boxes indicating 'normal' or 'mirror-image'. Shah and Miyake (1996) used a maximum of 200ms at this stage. If no response was made the letter remained on the screen for 5000ms. Shah and Miyake (1996) used a maximum of 2200ms at this stage. The next letter appeared on the screen following a delay of 250ms. After the entire set of letters was presented, a diamond-shaped grid appeared on the screen with eight "buttons" indicating the seven possible orientations (plus upright). The task of the participant was to use a mouse to click in the direction of the top

of each letter in the order of appearance. After the participant clicked on the appropriate number of buttons, the grid disappeared from the screen and the first letter of the next trial appeared. The spatial span task included a total of 20 letter sets, 5 sets at each level ranging from two-letter to five-letter sets, and participants were presented with increasingly longer sets of letters. The participants also received practice sets of letters at the two-letter set level. Following is an example of set of letters at the two letter level. The first letter appears on the screen as shown in Figure 3.4. For each letter, participants must remember the orientation of the letter as well as processing it as normal or mirror-imaged. Then a second letter appears on the screen as shown in Figure 3.5. Following the presentation of the two letters, a recall grid appears on screen onto which the participant mouse-clicks the correct orientation of each letter in the set, in the order they were presented, as shown in Figure 3.6.

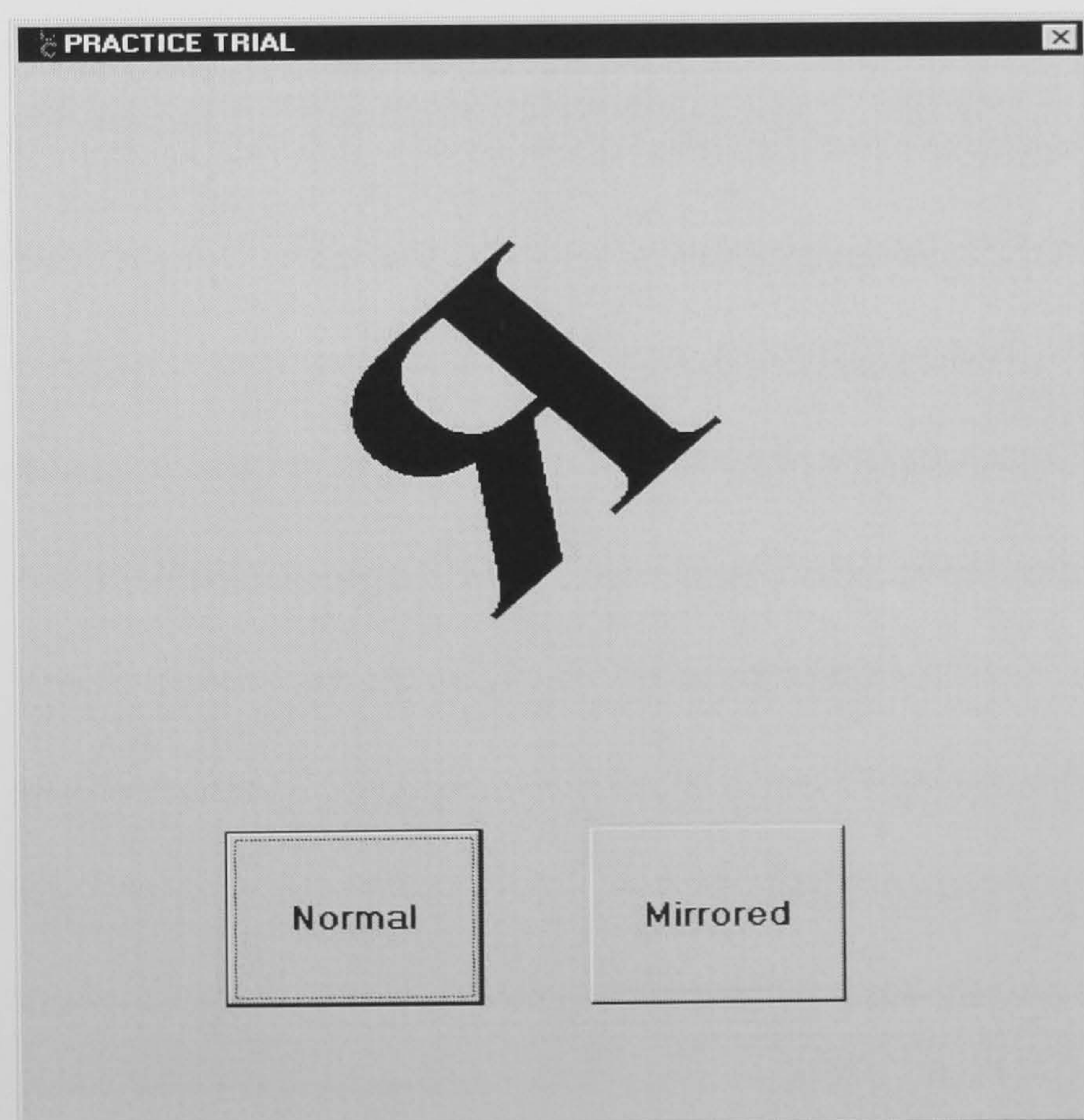


Figure 3 . 4 - Letter span, representation of letter 1

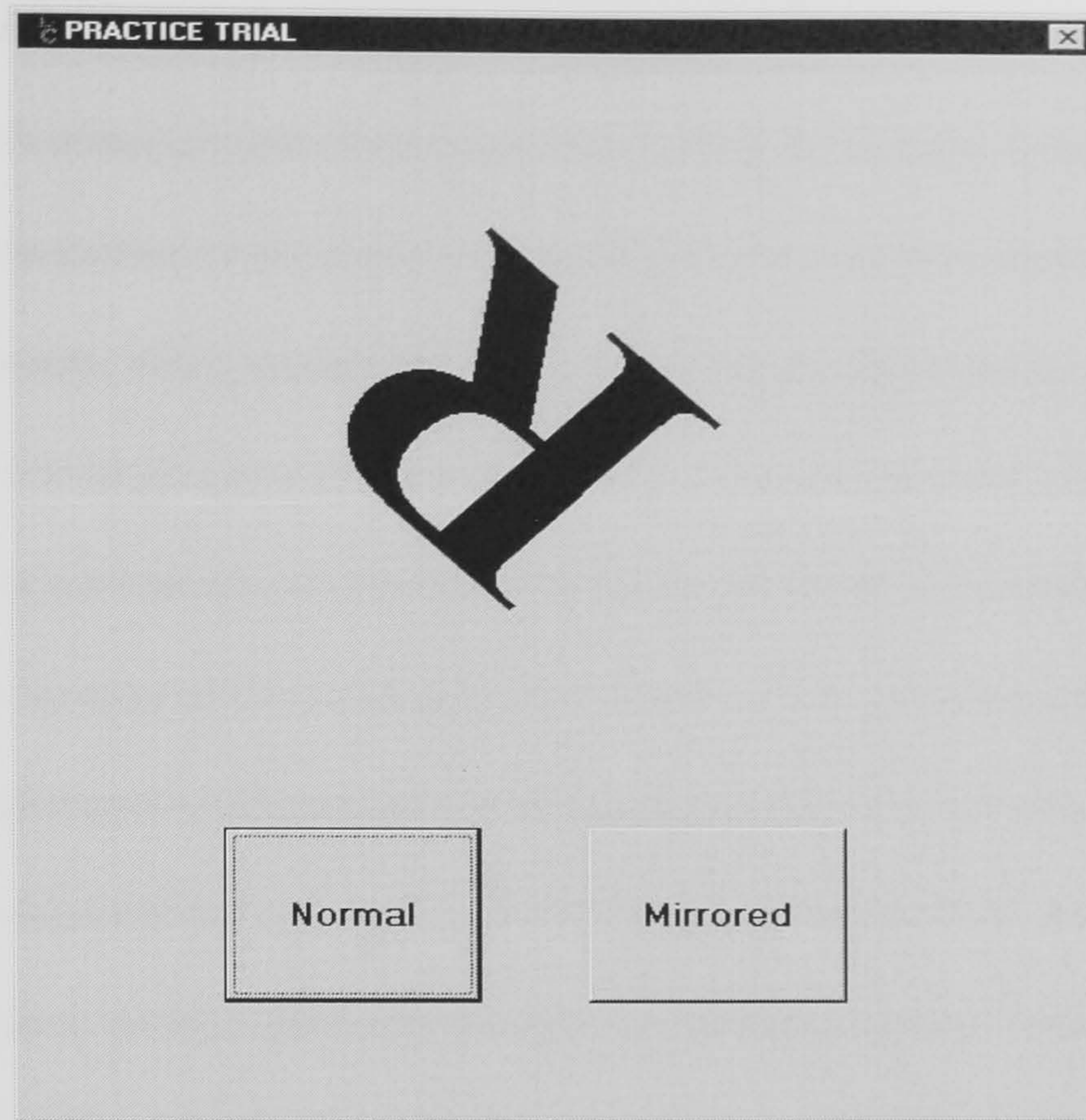


Figure 3 . 5 - Letter span, representation of letter 2

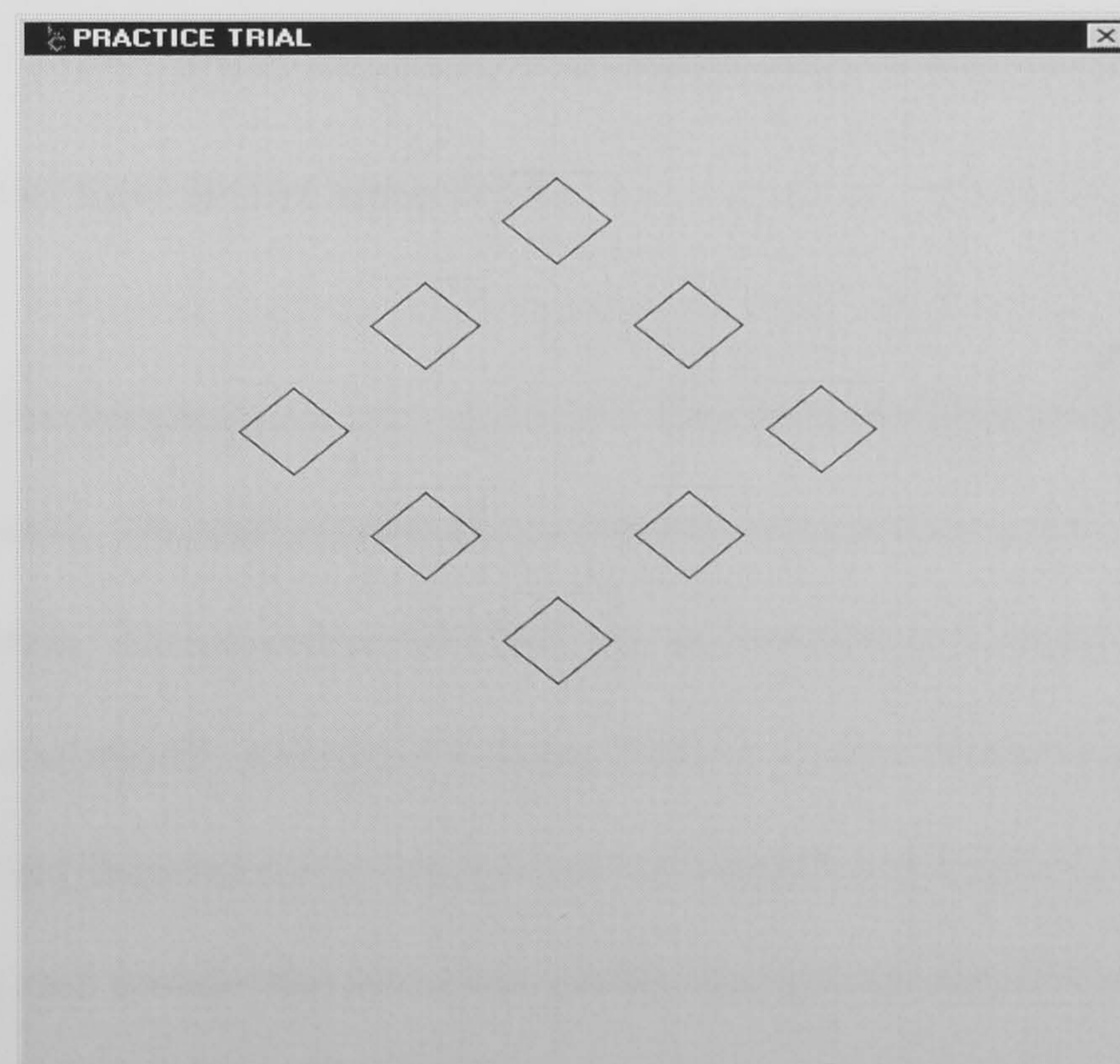


Figure 3 . 6 - Letter span, recall grid for letters

Scoring

The participant's complex spatial span score was defined as the highest set size for which all of the spatial orientations of the letters were recalled in the correct sequence, for at least three of the five sets. Half a point was added to the score if the participant's recall was accurate on two of the five sets at the next set size. When a participant successfully recalled less than three of the five sets at a particular level but was able to recall two or more sets at a higher level, the average of the lower and the upper limits was used as the span score (limits score). The maximum score obtainable for the span or limits measure was 5. A global score was also calculated for this task. When a participant was correct on a full set of letters, regardless of set-size, they were awarded one point per correctly recalled letter. The maximum score obtainable for the global measure was 70. The number of errors made in the verification component of the task (normal/mirror-image judgement) was also recorded (letter error).

Complex Spatial Tic-Tac-Toe Span (TTT)

Procedure

This task was also designed as a span measure of functional working memory capacity for spatial information. The task requires the simultaneous maintenance and processing of spatial information. The procedure followed was as Daneman and Tardif's (1987) study. Participants had to identify a series of winning lines in a game of three-dimensional tic-tac-toe. Each A4 card depicted a two-dimensional representation of a three-dimensional tic-tac-toe game. The card was divided into a top, middle, and bottom panel, each containing a 3x3 cell grid which participants were to imagine as the top, middle, and bottom platforms on a three-dimensional tic-tac-toe board. Some of the cells were occupied by red and blue tokens representing the pieces of the two players in the game. Embedded in this configuration of

tokens was a winning sequence, that is, three tokens of the same colour that formed a straight line in conformity with the rules of three-dimensional tic-tac-toe (a line that was horizontal, vertical, or diagonal in a two-dimensional or three-dimensional plane). The participant's task in each case was to locate the winning line. Participants saw one card at a time and identified the winning line by touching the three tokens with their index finger. At the end of a set of two, three, or four cards, participants had to recall the locations of each winning line in the set by pointing to the correct positions on an actual three-dimensional tic-tac-toe game board. Participants were given several practice items at the two winning lines level before the test began. They were warned to expect the amount of winning lines per set to increase during the course of the test. The span test consisted of five sets each of two, three, and four winning lines. Participants were presented with increasingly long sets of cards. Participants continued through all of the fifteen sets (five sets each of two, three, and four cards). Following is an example of a set of lines at the two line level. Please note that the representations shown in Figures 3.7 to 3.9 are from the computerised version of the task used in Experiment 2 onwards. The first line appears on a card as shown in Figure 3.7. For each winning line, participants must remember its orientation whilst processing it by pointing out its position. Then a second line appears on a card as shown in Figure 3.8. Following the presentation of the two lines, a real 3-dimensional recall grid is used onto which the participant points out the correct winning lines in the set, in the order they had been presented, as shown in Figure 3.9.

Scoring

The number of sets correctly recalled, per line, was taken as a measure of a participant's complex spatial TTI span. For example, if a participant was correct on all five of the two-card sets, and only one of the five three-card sets, he/she was assigned a span of 13 ($2+2+2+2+2+3=13$). If a participant was correct on four out of the five two-card sets, two

out of the five three-card sets, and one of the five four-card sets, he/she was assigned a span of 18 ($2+2+2+2+3+3+4=18$). The maximum score obtainable on this task was 45.

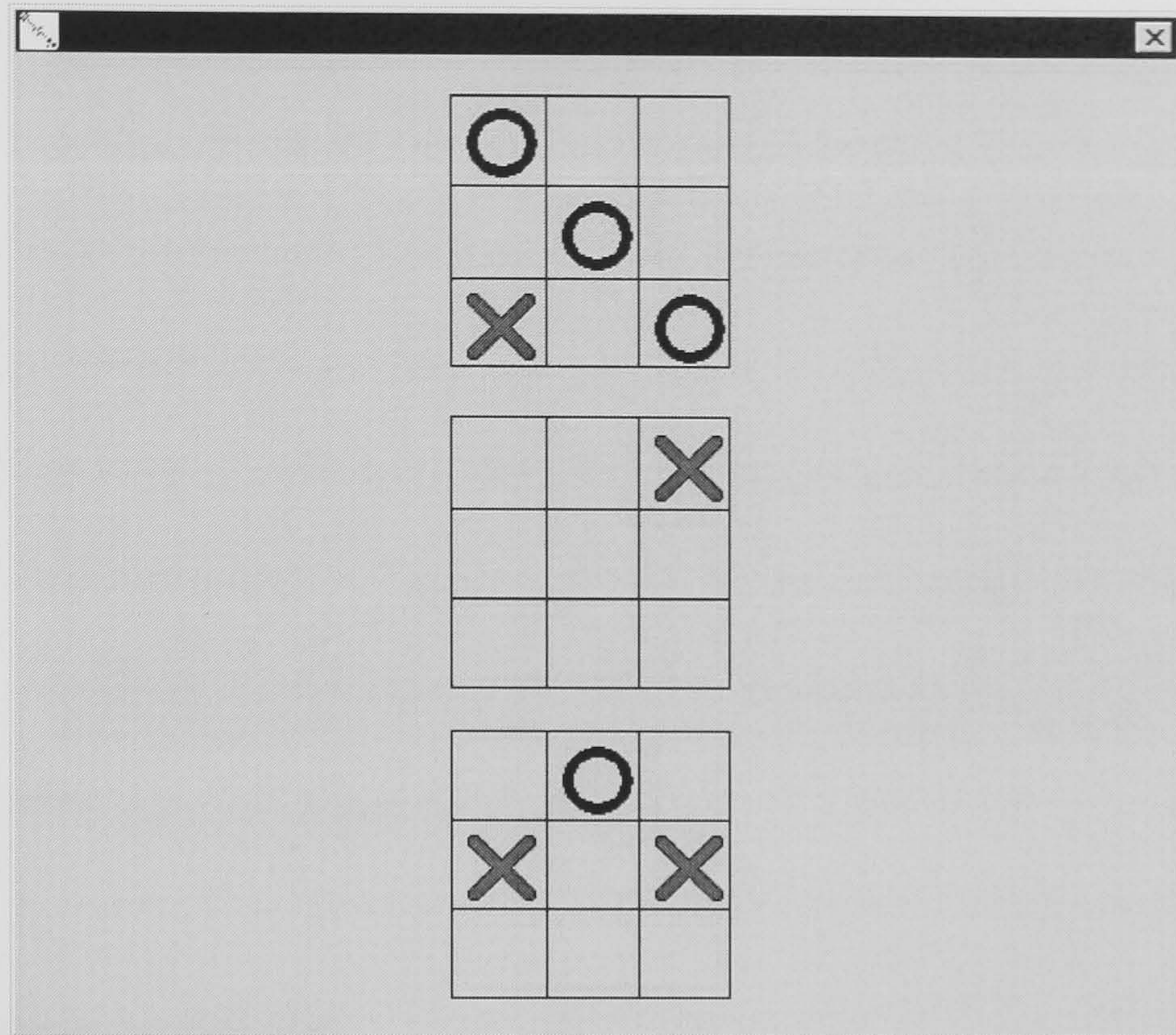


Figure 3 . 7 - TTT span, representation of winning line 1

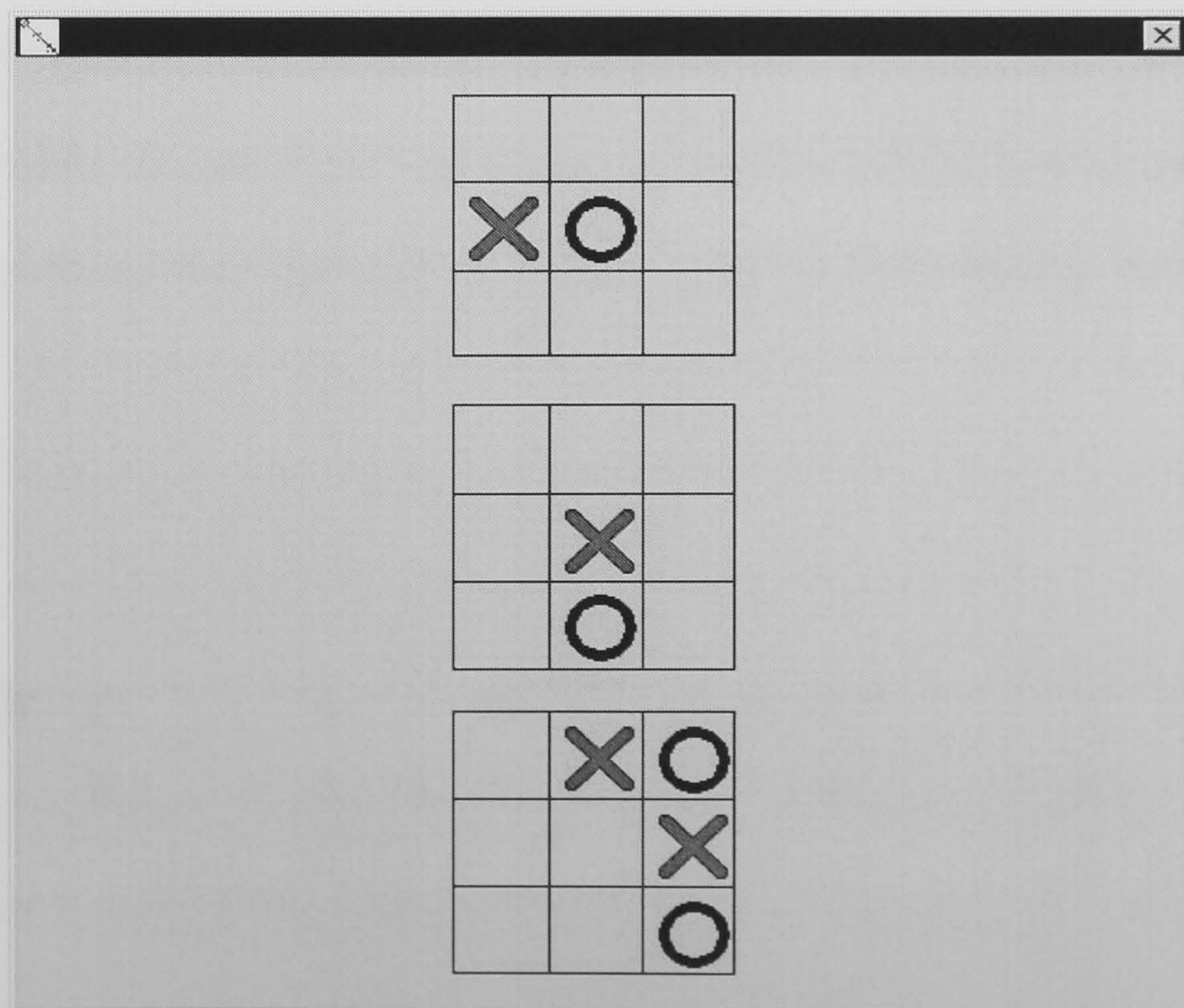


Figure 3 . 8 - TTT span, representation of winning line 2

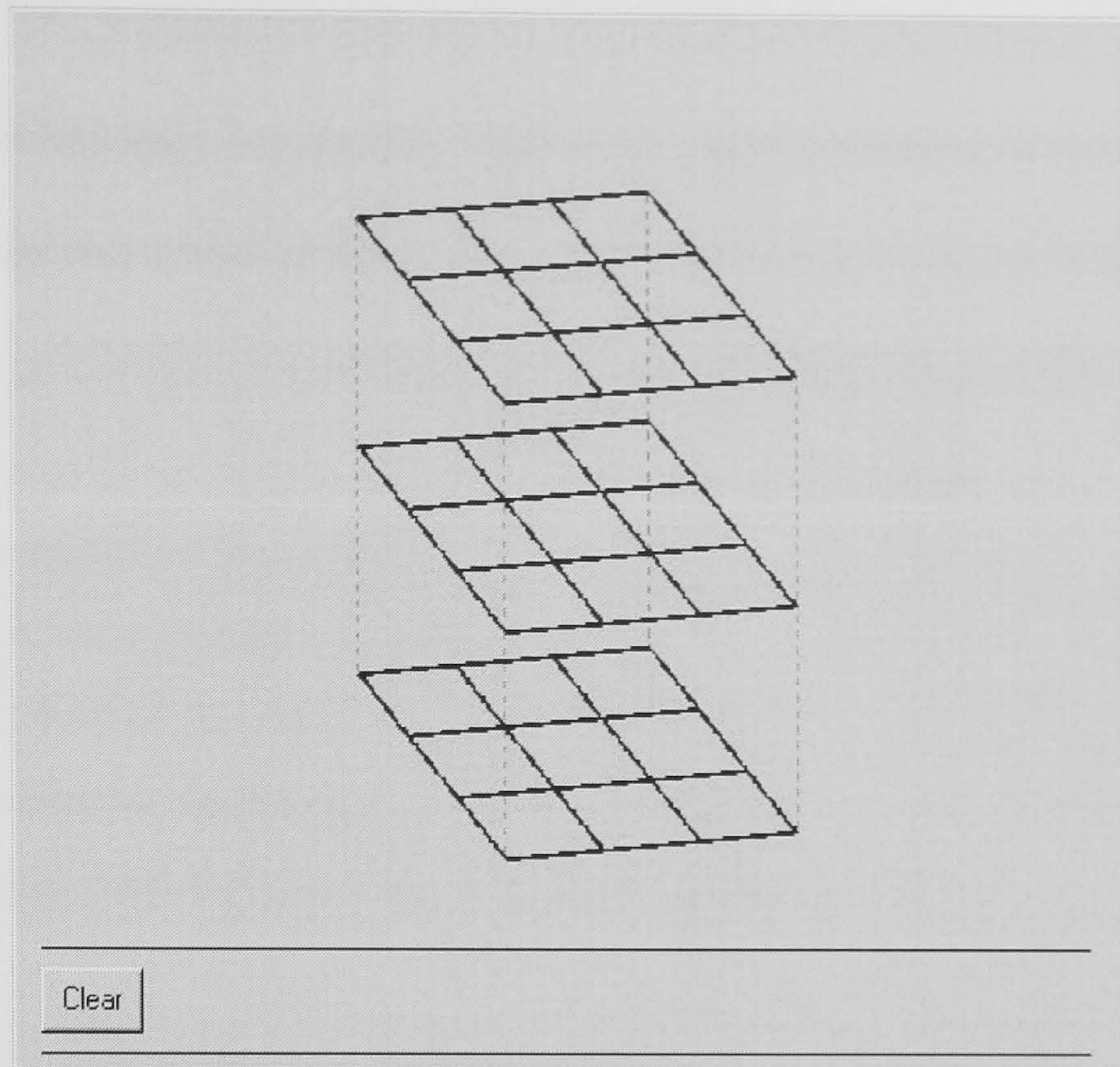


Figure 3 . 9 - TTT span, recall grid for winning lines

Spatial Inference (spatial ver)

Procedure

The procedure followed was based on Byrne and Johnson-Laird's (1989) study. Each participant received 24 two-dimensional spatial problems. The order of presentation was randomised before the start of the experiment but remained in this order for each participant. The spatial problems were used from each of four possible orientations as shown in Figure 3.10.

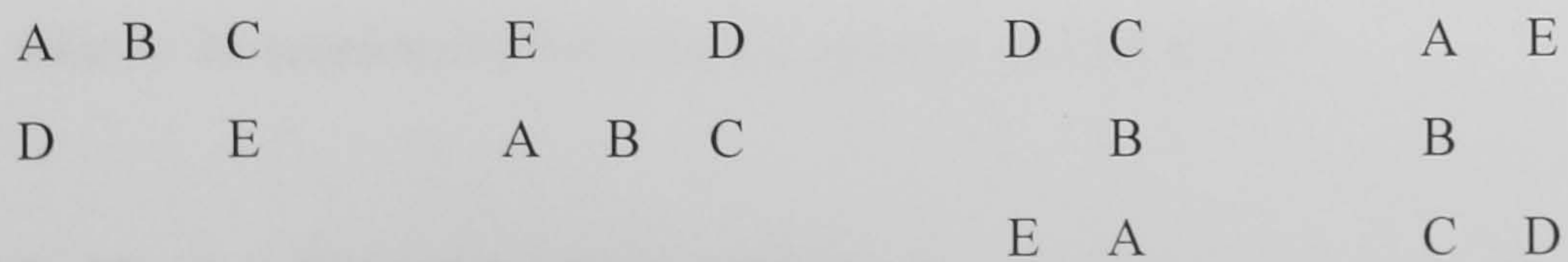


Figure 3 . 10 - Four possible orientations for five-term series spatial problems

All 24 problems had valid conclusions. Half of the spatial problems, according to model-based theory, had one model answers. Figure 3.11 presents an example of this.

<i>Premises</i>	<i>Models</i>
The apple is on the left of the orange	A O B
The banana is on the right of the orange	M P
The melon is in front of the apple	
The peach is in front of the banana	
Hence, what is the relationship between the melon and the peach?	

Figure 3 . 11 - A one model spatial problem

Only one model or spatial array of the premises can be used here to describe the layout of the objects within this set. The correct answer would be that the melon is to the left of the peach. The other half of the spatial problems, according to model-based theory, were consistent with two models or spatial arrangements, but were valid. Figure 3.12 shows an example of this.

<i>Premises</i>	<i>Models</i>
The torch is on the right of the hammer	H D T D H T
The drill is on the left of the torch	S P S P
The spanner is in front of the drill	
The pliers are in front of the torch	
Hence, what is the relationship between the spanner and the pliers?	

Figure 3 . 12 - A multi model spatial problem

More than one model of the premises can be used here to describe the layout of the objects within this set. In both of these models it holds that the spanner is to the left of the pliers. A full list of all the problems used can be found in Appendix A1.5. The participants were given instructions by way of an example. They were told that the experimenter would read aloud a description of the layout of some objects, which they could imagine arranged on the floor in front of them. They were encouraged to listen attentively. The description was read to them twice at a reasonable pace. They were then asked about the location of two of the objects, e.g., "What is the relation between the dish and the spoon?" The participants were told that their answers must include one of the following: to the left of, to the right of, behind, in front, and not enough information to tell. The two items requiring the participant to infer a relationship were always the pair of items (the D and the E in the examples) that were not explicitly interrelated in any premise of the four-premise problems. The participants were provided with an answer sheet for their responses to the 24 spatial problems. The participants were given a practice problem before they began the test items to ensure they understood the process of forming a conclusion and the format that each problem was in (see Appendix A1.6 for a full set of instructions).

Scoring

The number of spatial problems correctly solved (out of a maximum of 24) was taken as a participant's score for the spatial reasoning task. This figure was then used in the main analysis. In addition, it was noted how many of each type of problem (one-model or two-model) was solved. This was used in further analysis to test the assumptions made by the mental model theory of reasoning.

Syllogistic Inference (syls ver)

Procedure

A series of 20 syllogisms was presented to each participant. The order of presentation was randomised at the start of the experiment and then remained in that order for each participant. The syllogisms were used from all of the four possible figures: AB-BC, BA-CB, AB-CB, and BA-BC. All of the syllogisms used had valid conclusions. Twelve syllogisms, according to the model-based theory, had one model answers. Figure 3.13 shows an example of this.

Premises (AB-BC)

Some of the Welders are Builders

All of the Builders are Carpenters

Models

W [[B] C]

W [[B] C] ...

Figure 3 . 13 - A one model syllogistic problem

Only one model of the premises can be used to describe the relationships between the classes. The correct answer here would be that some of the welders are carpenters. Eight of the syllogisms, according to model-based theory, had more than one model answers, but were valid. Figure 3.14 presents an example of this.

Premises (AB-BC)

Some of the Electricians are Bakers

None of the Bakers are Geologists

Models

E [B] E [B] E [B]

E [B] E [B] E [B]

[G] E [G] E [G]

[G] [G] E [G] ...

Figure 3 . 14 - A multi model syllogistic problem

More than one model of the premises can be used to describe the relationships between the classes. The correct answer here would be that some of the electricians are not geologists as this is the only conclusion that holds across all three models. A full list of all the problems used can be found in Appendix A1.7. Instructions that were given to the participants included a full explanation of what the task required. The participants were told to provide a conclusion to the information given in the two statements. The participants were given an example of a syllogistic argument, and were shown all the quantifiers that were possible in the conclusion: all, some, none, some....not, and not enough information to tell. The participants were told that the written instructions could be kept in front of them to assist in the answering of the set of problems. The participants were given a practice syllogism before they began the test items to ensure they understood the process of forming a conclusion and the format that each problem was in. Each syllogism was read aloud to the participant twice, at a reasonable pace. They were always told which two terms in the premises to link together in the conclusion. The participants were instructed to write their conclusion to the syllogism on the response sheet provided (see Appendix A1.8 for a full set of instructions).

Scoring

The number of syllogisms correctly solved (out of a possible maximum of 20) was taken as the participant's score on the syllogistic reasoning task. This figure was then used in the main analysis. In addition, it was noted how many of each type of problem (e.g. one-model, or multi-model) was solved. This was used in further analysis to test the assumptions made by the mental model theory of reasoning.

Spatial Ability (spat ability)

Procedure

This task was taken from the National Foundation for Educational Research (Smith & Whetton, 1988). The General Ability Tests provide four separate assessments. The present study used one of those assessments, the Spatial Ability Test. The test required the encoding, rotation and comparison of complex visual images. The participants task was to imagine what a flat pattern would look like if it were cut out and folded into a solid object. The participants were given a booklet of 20 flat patterns, each of which had four solid objects shown below the flat pattern. They were also given a response sheet. Their task was to answer 'no' if one of the four solid objects definitely could not be made and 'yes' if one of the objects definitely could be made from the relative flat pattern. The participants were given fully standardised instructions.

Scoring

If the participant accurately responded to a yes/no question, they received a point. The maximum score obtainable on this task was 80.

Verbal Ability and Reading Rate (ver ability, RR)

Procedure

The task used to measure participants verbal ability was part of the Nelson-Denny Reading Test (Brown, Bennett and Hanna, 1981). The Nelson-Denny Test provides an assessment of the skills involved in the reading process - reading comprehension, vocabulary development, and reading rate. The present study used the reading comprehension and reading rate measures. The participants task was to read completely through a passage of text and then answer the questions following that passage. The questions had multiple choice answers (five choices per question). The participants were given a booklet of eight

passages. They were also given a response sheet. The participants reading rate was assessed in the first minute of the test. This was achieved by timing how far each participant read within that minute on a standardised scoring sheet per line of text. The participants were given fully standardised instructions.

Scoring

If the participant accurately responded to a multiple choice question, they received a point. The maximum score obtainable on this task was 36. The standard scale present in the Nelson-Denny test was used to assess a participant's reading rate.

3 . 2 . 3 Results

In addition to correlational analyses and t-tests, confirmatory factor analysis was also performed on the data from this experiment. Analyses based on correlations and comparison of means will be reported here, but the factor analysis for all of the first three experiments will be discussed in a factor analytic section after Experiment 3.

Please note that from now on all the task names will be simplified for use in the results section. The labelling will be as marked in brackets in the procedure section - for instance, the complex spatial letter span will be expressed as letter in the results section. There are a number of options for scoring many of the span tasks. However, only one method per task will be used throughout the results section. Reasons for use of particular scoring methods will be explained below. The cross-correlations of the global, normal and limits span scores were between $r = .76^{**}$ and $r = .95^{***}$. This suggests that the measures reflect the same underlying processing capacity. The global score was chosen for use due to it being a continuous variable with a wider range. The global score will be used throughout the following six experiments for the arrow, word and verbal spans. For the letter span a floor

effect in the normal spatial letter span score was found. The following analysis accounts for this by using the spatial letter limits score, as described in the method section. This limits score was calculated by taking the average score between a participant's lowest and highest set size (the intercorrelation between the spatial letter span and spatial letter limits score was $r = .76^{**}$).

3.2.3.1 Descriptive Statistics

One participant was dropped following an examination of the descriptive statistics and scatterplots for the arrow span. There was an outlying score which was skewing this measure (a low score). The participant also voiced concern during the experiment in relation to their ability to continue with other tasks. The participant generally underperformed on all tasks, and was considered to be unrepresentative of scores obtained by all others within the sample. Without this participant, the distribution of scores was roughly normal for all the tasks. Thus, the participant who scored 0.0 on the arrow span was dropped and the rest of the analysis was run on 45, not 46, participants. Table 3.1 shows descriptive statistics and reliabilities for the nine measures used in Experiment 1.

Due to different (but related) scoring methods being used for the span measures, as opposed to Shah and Miyake's (1996) data, it is hard to make meaningful comparisons of average scores and standard deviations with those they obtained. Experiment 1 uses the global score where as Shah and Miyake (1996) used the strict span score. Regardless of these related scoring methods the data from the span measures, using the scoring method described previously, do indicate a good overall range. There are no heavily skewed distributions, indicating the absence of floor and ceiling effects.

It is of interest here that on the verbal span task, participants made on average almost 8 errors (8%) in total when making the true/false judgement in processing. Also, participants made on average almost 9 errors (13%) in total when making the normal/mirror-imaged judgement in processing. These error levels are similar for both the complex verbal and complex letter span tasks. These levels are well below that expected if a guessing strategy (50% error rate) had been used by participants. This finding suggests that participants are attempting both the processing and storage components of the tasks.

	Mean	SD	Min	Max	Skewness	Reliability
<i>Span Tasks</i>						
arrow	25.38	10.26	10	54	.54	.77
word	49.31	15.43	26	94	.96	.76
verbal	34.09	17.00	8	71	.42	.92
verbal error	7.76	5.67	0	27	1.33	-
letter	3.08	1.10	0	5	-1.26	.69
letter error	8.78	5.97	1	26	1.17	-
TTT	17.71	8.26	6	45	1.05	.51
<i>Reasoning Tasks</i>						
spatial ver	14.18	4.25	7	23	.15	.77
spatver1M	8.20	2.56	3	12	-.28	-
spatverMM	5.98	2.57	1	11	.13	-
sylls ver	9.98	2.35	4	18	.94	.45
syllsver1M	8.78	1.36	4	10	-1.32	-
syllsverMM	1.20	2.07	0	8	2.04	-
<i>Ability Tasks</i>						
spat ability	61.11	10.87	33	80	-.73	-
ver ability	22.36	6.56	5	33	-.94	-
RR	253.58	60.05	162	413	.60	-

Table 3 . 1 - Descriptive statistics for Experiment 1 tasks (n=45)

Table 3.1 also shows data for the reasoning and ability measures. The spatial and verbal ability measures show a good overall range of scores indicating the varied performance levels for the participants.

Each span task was inspected for each participant to ensure internal reliability of the measures. Their reliability was assessed by performing a split-half test calculation on each of the five span measures used. The Spearman-Brown formula was then used to provide an estimate of reliability for each task. If a corrected correlation of over $r = .70$ is obtained, the result is highly satisfactory for group measurement (Rust & Golombok, 1999). If a corrected correlation of over $r = .50$ is obtained, the result is adequate for group measurement (Rust & Golombok, 1999). Any corrected correlation below $r = .50$ is considered inadequate for use (Rust & Golombok, 1999). The corrected correlations for the span tasks were as in Table 3.1. Table 3.1 shows that the arrow, word and verbal span tasks produced a reliable odd-even corrected correlation of over $r = .75$. The letter span produced an odd-even corrected correlation at an adequate level. The non-computerised TTT span produced the lowest corrected correlation of $r = .51$, but this is still of some value in group measurement. This lower reliability for the non-computerised TTT span may have been due to an audience effect - the experimenter was involved at every stage of the task and may have inadvertently influenced responses made by the participants. It might also be due to the non-computerised nature of the task. Procedural differences between participants are more likely when the Experimenter is implementing a non-computerised version of the task, rather than the standardised computer-based ones.

Each reasoning measure was also assessed for internal reliability using Cronbach's alpha.

This method assesses reliability from the consistency of all items in the sum scales and

thereby reduces the arbitrariness of the split-half method which just divides the scale in a random manner. As seen in Table 3.1, the reliability estimate for the spatial reasoning task was adequate for group measurement ($r = .77$). The reliability estimate for the syllogistic reasoning task was just below an adequate level for group measurement ($r = .45$).

3.2.3.2 The Span Measures

Correlational analysis was carried out on the data in order to assess the relationships between the nine measures used in Experiment 1 (see Appendix B1.7 for a full matrix).

Significant relationships would suggest some commonalities between the tasks. Due to the size of the whole matrix, it has been split into sections in order to answer specific questions.

The correlational analysis begins by examining the prediction that a dissociation would be expected between the complex verbal and spatial spans, as in Shah and Miyake (1996). The correlation matrix for Experiment 1 span measures is presented in Table 3.2.

Firstly, Table 3.2 shows the highest correlation is between the complex verbal and word span measures at ($r = .59, p < .001$). This is unsurprising given the similarity between material to be remembered, in this instance words. It is of note that the word span is assumed to measure passive resources of working memory, whilst the verbal span is assumed to measure active and passive resources of working memory. The word span also correlated significantly with the two complex spatial measures, letter and TTT ($r = .42, r = .39, p < .01$), but not with the simple arrow span ($r = .23$). The word span correlated significantly with the complex spatial spans which might indicate that an element of both spatial measures involved similar processes to that of the word span.

	arrow	word	verbal	verbal err	letter	letter err
arrow	1					
word	.23	1				
verbal	.11	.59***	1			
verbal err	.05	-.17	-.29 ⁻	1		
letter	.31*	.42**	.37*	-.28 ⁺	1	
letter err	-.12	-.24	-.41**	.49***	-.27 ⁺	1
TTT	.08	.39**	.31*	-.25 ⁺	.55***	-.28 ⁻

(two-tailed sig. levels: ⁺ = p < 0.1, * = p < 0.05, ** = p < 0.01, *** = p < 0.001)

Table 3 . 2 - Experiment 1 correlations between the five span tasks

An absence of a correlation between the word and arrow spans would be expected if the two tasks were in fact tapping into the two sub-systems of working memory which are thought to be separate resources. The second highest correlation was between the two complex spatial spans (at $r = .55$, $p < .001$). This would be expected if they rely on similar information and are tapping into common properties afforded by the same resource. The only significant correlation in relation to the simple arrow span was with the complex letter span ($r = .31$, $p < .05$). This finding was also seen in Shah and Miyake's (1996) data. Of note here is the fact that the arrow span did not correlate with the complex TTT span ($r = .08$). The complex verbal span also correlated significantly with both the complex spatial spans. This indicates that there may be similar processes common to all, and is not consistent with the dissociation between verbal and spatial resources over and above the sub-systems of working memory.

Lastly, the processing error associated with the complex verbal and letter spans was looked at. There was a significant cross correlation between these two types of error ($r = .49$, $p < .05$). These two tasks rely upon very different information (true/false versus normal/mirror-image judgements) but seem to be related in some way. If there was a dissociation between

the way we process verbal and spatial information, this correlation would not be expected. The error associated with the complex verbal and complex letter spans correlated negatively with all the complex span measures ($r = -.25$ to $r = -.29$, $p < .10$ for verbal error: $r = -.27$ to $r = -.41$, $p < .10$ to $p < .01$ for letter error). This suggests that as an individual's performance improves on the complex verbal and complex letter spans, their processing error on the complex spans decreases. It also suggests that people making errors on the processing component are also making errors on the memory component of the tasks.

There are two major findings here. The first is that no clear dissociation was seen between the complex span measures, a finding inconsistent with Shah and Miyake (1996). The data indicates there are significant correlations between all the complex span measures. This may lead to the conclusion that the complex spans are tapping a common resource of working memory. The second finding is that there is no significant correlation, although they are still positively related, between the simple spans. This may be expected if these measures reflect the passive storage capacity of the PL and VSSP systems of working memory.

3.2.3.3 The Span and Reasoning Measures

The next set of correlations refer to the relationships between the span measures and the reasoning measures. The data in Table 3.3 refers to whether individual differences in working memory predict individual differences in both spatial and syllogistic reasoning, and spatial and verbal abilities. Past dual-task literature has shown that spatial and syllogistic reasoning involve the CE component of working memory, whilst it is less clear of the role played by the two sub-systems. The reasoning tasks shown in Table 3.3 would be expected to show relationships with the complex measures of working memory. The extent to which

spatial reasoning and syllogistic reasoning tasks relate to particular complex span tasks may lead to a better understanding of the role of spatial and verbal working memory resources in deduction.

	arrow span	word span	verbal span	letter span	TTT span	spatial ver	syllogisms ver	spatial ability	verbal ability
spatial ver	.10	.50***	.36*	.24	.48***	1			
sylls ver	.43**	.40**	.30*	.37*	.10	.27 ⁺	1		
spatial ability	.26 ⁺	.06	.19	-.05	-.00	-.12	.28 ⁺	1	
verbal ability	.12	.12	.10	.19	.22	-.09	-.04	.11	1
RR	-.10	.20	.32*	.05	.03	-.12	-.13	.07	.52***

(two-tailed sig. levels: ⁺ = $p < 0.1$, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$)

Table 3 . 3 - Experiment 1 correlations between the span, reasoning and ability tasks

Considering first the verbally presented spatial reasoning task (spatial ver), Table 3.3 shows the best predictors are the word ($r = .50$, $p < .001$) and TTT ($r = .48$, $p < .001$) spans, and to a lesser extent the verbal span ($r = .36$, $p < .05$). These results are consistent with past results investigating the relationship between spatial reasoning and working memory. For instance, Vandierendonck and De Vooght (1997) found that dual-tasks assumed to load both the verbal and spatial stores, as well as those secondary tasks assumed to load the CE, interfered with spatial inference as a primary task. The correlation between the spatial task and the TTT span may suggest that both tasks rely on common processes that are spatial in nature. This correlation might also suggest that both these tasks rely on general working memory resources. This is because all of the complex span measures shown in Table 3.2 correlate significantly, perhaps suggesting a common resource for both verbal and spatial material. The correlation between the spatial task and the word span may suggest that both tasks rely on common processes - the necessity of holding in mind premise information.

Perhaps the verbal presentation mode increased the load on verbal working memory. This finding will be followed up in Experiments 2 and 3 by controlling for modality differences (presenting problems both verbally and visually). The last significant correlation was between the spatial reasoning task and the verbal span ($r = .36, p < .05$), suggesting the need, in both tasks, to store and process verbal information. So for the verbally presented spatial reasoning task, the results shown in Table 3.3 are consistent with the view that both spatial and verbal resources of working memory are required in order to solve five-term series spatial inferences. In addition, it may be assumed from this data that some sort of passive verbal storage is also required (link with word span) in order to perform the task.

Turning now to the verbally presented syllogistic reasoning task (syllogisms ver), Table 3.3 shows the syllogistic task is predicted by all the complex and simple span measures, apart from the TTT span (correlations ranging between $r = .30$ and $r = .43$). One possible interpretation of this is that both verbal and spatial strategies are used in order to solve a verbally presented syllogistic task. These results are compatible with the notion that both spatial and verbal resources of working memory are required in order to solve verbally presented syllogistic reasoning tasks. Again, however, due to the fact that the complex span measures correlated significantly, a general resource explanation of these correlations may provide the best account.

Table 3.3 presents a generally positive matrix where both reasoning tasks are concerned. The data in Table 3.3 indicates that the word and verbal spans consistently predict performance on both reasoning tasks, spatial and syllogistic. The pattern becomes more complicated in relation to the spatial spans, although they do predict the reasoning tasks overall. The data in Table 3.3 also shows that performance on spatial reasoning moderately

correlated with performance on syllogistic reasoning ($r = .27, p < .10$), although this is a borderline effect.

Finally, the ability measures will be discussed in relation to the span measures used. Table 3.3 shows a significant correlation between reading rate and the verbal span. One possible explanation for this is that efficiency in the processing of sentences is aided by quicker rates of reading. Table 3.3 also shows a highly significant correlation between verbal ability and reading rate ($r = .52, p < .001$). This would be expected due to them being related components of the same standardised task. Table 3.3 indicates no further significant relationships between the span measures and spatial or verbal ability tasks. Shah and Miyake (1996) did find significant correlations between spatial and verbal abilities with spatial and verbal spans, respectively. The results shown in Table 3.3 differ from Shah and Miyake's (1996) results which is quite surprising. An explanation for this could be specific to the particular tasks used in this experiment. Experiment 1 used different ability measures to those used by Shah and Miyake (1996), although it is not clear why the ability tasks here are not correlating with working memory capacity. This discrepancy remains an unsolved issue with respect to the ability measures that merits further investigation. However, the main focus of this thesis is on deductive reasoning, and it should be noted that there was no further inclusion of ability measures in the following studies.

3.2.3.4 Further Analysis

What follows is other analysis carried out on the data relating to the easy/hard divide in both the spatial and syllogistic task. Past research suggests that problems that only need one model to be constructed in order to reach a conclusion should be easier than problems where multiple models are needed.

	% correct
spatial1M	68
spatialMM	50
sylogisms1M	73
sylogismsMM	15

Table 3 . 4 - Experiment 1 % of correct responses to 1M and MM problems

Problem difficulty percentages for the spatial and syllogistic task are presented in Table 3.4. The 1M problems, both spatial and syllogistic, are solved more often than the MM problems. An average of 14 out of 24 (58%) verbally presented spatial inferences were solved, 8 of which were single model problems. An average of 10 out of 20 (50%) verbally presented syllogistic inferences were solved, 9 of which were single model problems. To confirm the significance of these differences, t-tests were performed on the data (see Appendix B1.1 for full results). Participants performed at a significantly higher rate on both the syllogistic and spatial one model problems, than the corresponding multi-model problems (d.f. = 44, $t = 19.5$, $p = .001$; d.f. = 44, $t = 5.2$, $p = .005$ respectively).

3 . 2 . 4 Discussion

The results from Experiment 1 suggest that individual differences in working memory capacity can explain variation in performance on syllogistic and five-term series spatial tasks. Experiment 1 also addressed the separability issue of central resources within working memory. Experiment 1 followed the same types of predictions made by Shah and Miyake (1996), but based on deductive reasoning problems. There was no indication that one reasoning task could be better predicted by verbal than spatial span tasks, and vice versa. Both the verbal and spatial spans predicted individual differences on the reasoning

tasks, a similar pattern of correlations being observed for both the syllogistic and spatial inference measures, when verbally presented.

Positive relationships were shown between the span measures used in Experiment 1. These results do not replicate the dissociation between verbal and spatial working memory shown by Shah and Miyake (1996). There was no indication that five-term series problems were best predicted by spatial working memory and syllogisms by verbal working memory, even at the level of the simple span measures, that are assumed to measure the relatively passive PL and VSSP resources.

The complex working memory spans were designed in order to assess individual differences in centrally processed information. The verbal and spatial span tasks consisted of different processing and storage components, indicating the verbal or spatial nature of the task. For instance, the verbal span combined the active task of verifying sentences as true or false with the passive task of remembering the last words of each sentence. Where as, the letter span combined the active task of verifying letters as normal or mirror-imaged with the passive task of remembering the orientation of each letter in a set. These complex measures of working memory were shown to cross correlate significantly. In addition, the processing components of the tasks also cross-correlated significantly. This suggests that both span tasks draw on common resources - this resource may be identified as a domain general working memory resource. This interpretation suggests that there may be a common central resource of working memory which is tapped by both the working memory span measures, the processing components of these tasks, and the reasoning tasks. This will be further addressed in the factor analysis section at the end of this experimental series.

The aims outlined at the start of this experiment relied upon the assumption that a dissociation would be observed between verbal and spatial complex working memory spans. This was not found and so no clear evidence can be forwarded for a separation of CE resources of working memory, as measured by the complex spans. Consequently, due to this lack of dissociation, no clear interpretation can be made of the verbal versus spatial resources that different reasoning tasks rely upon. In addition, another expectation was that the simple word and arrow spans would be dissociated for verbal and spatial information, respectively. Again, there was no clear indication of this, although they showed a low positive correlation, perhaps questioning the assumption that the simple spans map onto the PL and VSSP systems of working memory specifically. What can be said when looking at this correlational data is that there is clear evidence that working memory capacity did predict deductive reasoning performance.

As well as the deductive tasks correlating with the capacity measures, they also correlated marginally well together. Although a small sample was used here, these results are not inconsistent with Stanovich and West (1998). They showed that a range of deductive measures correlated significantly together and with measures of cognitive ability. Their result is consistent with the findings here in that the spatial and syllogistic task showed a borderline relationship with each other and with the capacity measures.

The extent to which these findings are consistent with the theoretical accounts of syllogistic and spatial reasoning, presented at the beginning of the chapter, is also of interest.

However, any interpretation must be considered in light of the failure to find a clear dissociation between the complex span measures, discussed above. Nevertheless, these findings do have implications for the three classes of reasoning theory discussed in Chapter

2. The findings clearly show that working memory correlates with reasoning performance, which is consistent with the processing theories. However, it is difficult to distinguish between the processing accounts in relation to verbal or spatial representations. Nevertheless, the finding that reasoning draws on working memory resources is a challenge to accounts that suggest that reasoning involves a low processing load, for instance the heuristic accounts. This will be summarised in more detail in the general discussion at the end of this chapter.

The reasoning tasks in Experiment 1 were presented verbally which entails keeping the premise information in mind. A possible interpretation of the results found is that performance on both types of problems first requires the interpretation of the linguistic information given in each premise. This link with verbal comprehension or processing may explain the relationships of both syllogistic and spatial inferences with the simple and complex verbal working memory resources. The verbal information may then be processed by the manipulation of some form of mental model. If the manipulation of such a mental model is spatial in nature then spatial working memory capacity should predict performance of both syllogistic and spatial inferences. The spatial inferences correlated highest with the word span, which at first glance seemed to disagree with the many authors maintaining a spatial model explanation. Assuming that the word span measures the capacity of the PL, a possibility might be that people are relying on their passive phonological store to hold in mind premise information due to the verbal presentation modality. They may need to translate this verbal information before constructing a mental model.

Further investigation was needed to investigate the influence of modality of each of the reasoning tasks (verbal versus visual presentation). The pattern of correlations between the

word span and the spatial inference problems indicated that individuals may be holding premise information in mind before being able to translate and interpret it. By presenting them visually, the need to keep the premise information in mind may be reduced. Hence, there may be less involvement of the PL, as measured by the word span. Reductions might be expected in the correlations between verbal working memory and spatial inferences, but especially with the simple word span. Thus, Experiment 2 investigated the interrelationships between the span measures and performance on five-term series spatial reasoning problems using both verbal and visual presentations.

3 . 3 EXPERIMENT 2

3 . 3 . 1 Aim

The pattern of findings in Experiment 1 suggested that both syllogistic and spatial inferences are predicted by a sub-set of both verbal and spatial working memory capacity measures. This may have been because the verbal presentation involved keeping in mind premise information. Therefore, the correlations with verbal working memory are perhaps not unexpected. However, this does not necessarily show that the deductive processes involved in the tasks involve similar processes to those involved in the verbal span tasks. Thus, the main aim of Experiments 2 and 3 is to examine this by controlling for modality of presentation.

Gilhooly, Logie, Wetherick and Wynn (1993) investigated syllogisms, presented both verbally and visually, together with dual tasks proposed to load the working memory sub-components. They proposed that presenting the syllogisms verbally would cause a higher memory load due to holding in mind the premise information and that conversely,

presenting the syllogisms visually, so that the premises were continuously available for inspection, would lower the memory load. Indeed, a significant effect of memory load on accuracy of syllogistic performance was obtained. Experiments 2 and 3 investigate the influence of presentation mode - both spatial and syllogistic tasks will be presented in both modalities, together with a range of working memory measures.

Experiment 2 used four working memory span tasks: one complex spatial measure, one complex verbal measure (Daneman & Tardif, 1987; Daneman & Carpenter, 1980), and a simple spatial and verbal measure (Shah & Miyake, 1996). The span tasks were given together with two reasoning tasks, a spatial inference task presented both verbally and visually. It might be expected that when the spatial inferences are presented visually, a higher accuracy rate is indicated in performance than when they are presented verbally. The tasks were as Experiment 1 unless otherwise stated.

The complex spatial letter span task was not used in Experiment 2 due to the following reasons. Firstly the two complex spatial span measures from Experiment 1, letter and TTT spans, correlated significantly ($r = .55$). This would suggest some similarity between the types of information being remembered and the resources required to respond to those tasks. Of all the computerised span tasks, the letter span measure showed the lowest reliability ($r = .69$). The non-computerised version of the complex TTT span also showed a low reliability, although in this study a computerised version of the task was developed. Computerising the TTT span would alleviate the effect of the Experimenter (an audience) that may have contributed to its low reliability in Experiment 1. Due to major time constraints in running individual differences studies, it was decided that only one complex spatial span could viably fit into the time scheme. Therefore, a version of the complex TTT span, compatible

with the other span task methodologies, was used instead of both the letter and non-computerised TTT tasks.

Experiment 2 in this series examines the extent to which the patterns of correlation in Experiment 1 resulted from the modality of presentation of the five-term series relational reasoning problems. The additional visual reasoning task will allow the assessment of resource differences dependent on task modality. A reduction of the involvement of the PL component, as measured by the word span, might be expected when presenting the spatial task visually. Experiment 2 again addressed the separability issue of central resources within working memory. Experiment 2 allows a further test of the separability hypothesis (Shah and Miyake, 1996).

3 . 3 . 2 Method

Pilot Study

In Experiment 2, it was decided to computerise the complex spatial TTT span used in Experiment 1. This was in order to standardise the presentation of all the span tasks in the same format. A pilot study was run to ensure that the new Visual Basic computer span task (TTT span) and the visually presented spatial inferences were easy to understand, use, and at a reasonable difficulty level. The new tasks were run on several participants who highlighted any difficulties they had with responding to the tasks.

Participants

Forty-nine participants took part in this study - 12 of whom were men and 37 of whom were women. The participants were undergraduate students and postgraduate students at the University of Plymouth and they received course credit or cash payment for participating. None of the sample had prior training in logic, or prior experience on any of the tasks used in Experiment 1. They were all native English speakers.

Procedure

Experiment 2 was carried out in one testing session per participant, the session being of approximately an hour and a half long. Participants signed up in a time slot on an experimenter sheet for the session. The session consisted of six tasks. All participants carried out these tasks in the following order (where an asterix (*) indicates a computerised task):-

spatial inference measure (verbal presentation),

complex spatial tic-tac-toe span*,

complex verbal span*,

simple arrow span*,

simple word span*,

spatial inference measure (visual presentation)*.

What follows is a summary of those tasks used in Experiment 2 which have not already been introduced in Experiment 1.

Simple Verbal Word Span (word)

The procedure and scoring was identical to Experiment 1.

Simple Spatial Arrow Span (arrow)

The procedure and scoring was identical to Experiment 1.

Complex Verbal Sentence Span (verbal)

The procedure and scoring was identical to Experiment 1.

Complex Spatial Tic-Tac-Toe Span (TTT)

Procedure

This task was designed as a span measure of functional working memory capacity for spatial information, as explained in Experiment 1. The procedure followed was based on Daneman and Tardif's (1987) study, but what follows are the parameters used to computerise this task. Each screen displayed a card depicting a two-dimensional representation of a three-dimensional tic-tac-toe game. The card was divided into a top, middle, and bottom panel, each containing a 3x3 cell grid which participants were to imagine as the top, middle, and bottom platforms on a three-dimensional tic-tac-toe board. Some of the cells were occupied by red (X) and blue (O) tokens representing the pieces of the two players in the game. Embedded in this configuration of tokens was winning sequence, that is, three tokens of the same colour that formed a straight line in conformity with the rules of three-dimensional tic-tac-toe. The participant's task in each case was to locate the winning line. Participants saw one card at a time on the screen and identified the winning line by clicking the three tokens with their computer mouse (in any order). The computer emitted a beep if the participant clicked on an incorrect line. At the end of a set of two, three, or four cards, participants had to recall the locations of each winning line in the set (in the order in which they had seen them) on a three-dimensional tic-tac-toe game board representation on screen. They did this by clicking their computer mouse on the screen in

the correct positions on the 3-D board representation. Participants were given several practice items at the two winning lines level before the test began. They were warned to expect the amount of winning lines per set to increase during the course of the test. The span test consisted of five sets of two, three, and four winning lines. Participants were presented with increasingly longer sets of cards. Participants continued through all of the fifteen sets (five sets each of two, three, and four cards). Please see Figures 3.7 to 3.9 in Experiment 1 for a depiction of a two line set. The only difference between the two tasks was that in this version all cards and recall grids were computerised and participants made choices by mouse clicks rather than by pointing out the winning lines to the Experimenter.

Scoring

The scoring was identical to Experiment 1.

Spatial Inference (spatial ver and spatial vis)

Procedure

The procedure followed was based on Byrne and Johnson-Laird's (1989) study, and as Experiment 1. Each participant received 16 two-dimensional spatial problems in both modalities, verbal and visual. The same 16 problems, a sub-set of the 24 used in Experiment 1, were used for both the verbal and visual task, but the content and order of presentation was changed. The problem content can be found in Appendix A1.5. The order of presentation of problem in both tasks was randomised before the start of the experiment but remained in this order for each participant.

Instructions for the verbal problems: The procedure was identical to Experiment 1.

Instructions for the visual problems: The screen displayed a set of instructions, similar to those for the verbal problems. The instructions explained that in this part of the experiment the participants would be asked about the relations between objects in a set. A problem

would appear on the screen describing the layout of a set of five objects. Shortly afterwards, a question would also appear on the screen asking them to relate two of the objects within a set. They were told that there would be 16 descriptions using the following terms to describe the layout of the objects: in front, behind, left and right. The participants were told that their task was to link two of the objects in the set using the above relations. They were told to imagine the five objects as if they were in front of them and then apply these relations to them. They were also told that for some of the descriptions they may feel there was not enough information to determine the relation between the objects, in which case they should answer “not enough information”. The instructions continued with a summary on how to respond to the descriptions. The participants were told that the screen would also display a blank box in which they were to type their responses to each problem. For example, if they thought that the answer was “fork left of plate” then this is what they should type. There was a button to press to move on to the next problem. They were given one practice problem.

Scoring

The scoring was identical to Experiment 1.

3 . 3 . 3 Results

In addition to correlational analyses and ANOVAs, confirmatory factor analysis was also performed on the data from this experiment. The correlational, significance analysis and discussion of findings for this experiment will be reported in this experimental section, but the factor analysis for all of the first three experiments will be explored in a factor analytic section that follows Experiment 3.

The cross-correlations of the global and normal span scores were again highly significant. Due to this, all scoring methods used in this results section are as previously explained in Experiment 1, at the start of Section 3.2.3.

3.3.3.1 Descriptive Statistics

Two participants were dropped due to reasons consistent with those concerning the dropped participant in Experiment 1. These participants expressed verbal concern throughout the experiment on their performance, but were asked to continue regardless of this. When looking at the results, both these participant's scores on all tasks was extremely low and no other participants failed the tests in such a manner. Thus, the participants who scored 0.0 on the arrow span were dropped and the rest of the analysis was run on 47, not 49, participants. A summary of descriptive statistics and reliabilities for the measures used in Experiment 2 can be found in Table 3.5.

	Mean	SD	Min	Max	Skewness	Reliability
<i>Span Tasks</i>						
arrow	24.94	9.53	5	51	.56	.80
word	52.06	16.56	26	88	.49	.92
verbal	38.47	17.96	10	88	.36	.90
TTT	8.40	5.19	0	24	.47	.80
<i>Reasoning Tasks</i>						
spatial ver	9.34	3.75	2	16	.14	.80
spatver1M	5.64	1.90	1	8	-.56	-
spatverMM	3.70	2.33	1	8	.23	-
spatial vis	11.21	3.13	6	16	-.33	.75
spatvis1M	6.30	1.69	3	8	-.63	-
spatvisMM	4.91	1.86	2	8	.11	-

Table 3.5 - Descriptive statistics for Experiment 2 tasks (n=47)

Table 3.5 shows that the span tasks were responded to with a good overall range and distribution of scores and are consistent with the results obtained in Experiment 1. The data of all tasks show distributions that were not distorted by skewed distributions. There was no evidence of floor or ceiling effects. The only discrepancy is the TTT span, which shows a lower mean score in Experiment 2. Note that The TTT span is computerised in Experiment 2. It may be that the removal of experimenter intervention after every set of lines by computerising the task enabled participants to answer without embarrassment of incorrect responding. Table 3.5 also shows data for the reasoning measures. There was again a good range of scores overall, and the ranges of data were normally distributed.

The reliability of the span tasks was assessed again by performing a split-half test calculation on each of the four span measures used. The Spearman-Brown formula was then used to provide an estimate of reliability for each task. The corrected correlations for the span tasks were as in Table 3.5. Table 3.5 shows that all the span tasks used produced a reliable odd-even corrected correlation of over $r = .75$. A corrected correlation of over $r = .75$ is deemed satisfactory for group measurement (Rust & Golombok, 1999). The non-computerised TTT span produced a low, but adequate, corrected correlation of $r = .51$ in Experiment 1. The now computerised version in Experiment 2 produced a much higher corrected correlation of $r = .80$, deemed very satisfactory in group measurement (Rust & Golombok, 1999).

Each reasoning measure was also assessed for internal reliability using Cronbach's alpha. As seen in Table 3.5, the reliability estimate for the spatial reasoning task, in both modalities, was satisfactory for group measurement ($r = .80$ for verbal and $r = .75$ for visual). This is consistent with the results found in Experiment 1 (see Table 3.1).

3.3.3.2 The Span Measures

Correlational analysis was carried out on the data in order to investigate whether the tasks were associated together (see Appendix B1.8 for a full matrix). The results for Experiment 2 span measures are presented in Table 3.6.

	arrow	word	verbal	TTT
arrow	1			
word	.28 ⁺	1		
verbal	.10	.64***	1	
TTT	.19	.40**	.41**	1

(two-tailed sig. levels: ⁺ = $p < 0.1$, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$)

Table 3.6 - Experiment 2 correlations between the four span tasks

Firstly, the highest correlation is between the word and verbal span measures, as in Experiment 1 ($r = .64$, $p < .001$). The word span also correlated significantly with the complex spatial measure ($r = .40$, $p < .01$), and there was borderline significance with the simple arrow span ($r = .28$, $p < .10$). The word and arrow span correlation is consistent with the result found in Experiment 1, although significance there was not reached ($r = .23$). It may be that the arrow and word spans tap separate storage systems afforded by the VSSP and PL respectively, although they may also be measuring some common resource. The arrow span did not correlate significantly with the other span measures, another result consistent with Experiment 1. In Experiment 1 the arrow span only showed a significant relationship with the letter span, which was not used here. Lastly, the verbal span correlated with the TTT span ($r = .41$, $p < .01$), a reasonably moderate relationship. Experiment 1 and 2 showed no clear dissociation between complex measures of spatial and verbal working memory. No evidence was found for a dissociation between tasks designed to measure centrally allocated working memory resources. Daneman and Tardif (1987) used the TTT

span in conjunction with a math and a verbal span task. Daneman and Tardif (1987) showed that only the math and verbal span tasks, not the TTT span task, predicted verbal ability. They concluded that there are independent processors in memory for language and non-language based information. The correlations between complex span measures in Experiments 1 and 2 are incompatible with this finding.

In conclusion for the data shown in Table 3.6, the results are consistent with those obtained in Experiment 1. A positive matrix can be seen between all the span measures in Table 3.6. The data indicates there are significant correlations between the two complex span measures used. This result would not be expected if there was a total separation of verbal and spatial resources at the level of the CE. The results from this, and Experiment 1, support the notion that all information, be it verbal or spatial in nature, is dealt with similarly by a domain-general resource. Experiment 1 and 2 both show low correlational relationships between the word and arrow spans. Thus, the data from both experiments also lends support to the notion of two separate sub-systems of working memory that deal with the passive storage of verbal versus spatial information. However, the low, but positive correlation between these measures suggests that they share some, albeit small, common processing requirement.

3.3.3.3 The Span and Reasoning Measures

The next set of correlations refer to the relationships between the span measures and the reasoning measures. This section examines whether individual differences in working memory predict individual differences in both verbally and visually presented spatial inferences. From past research, it is suggested that spatial reasoning requires CE resources, as well as a reliance on a spatial storage component (e.g., Klauer, Stegmaier & Meiser,

1997; Vandierendonck & De Vooght, 1997). Hence, the reasoning tasks shown in Table 3.7 would be expected to show relationships with the complex measures of working memory.

	arrow	word	verbal	TTT
spatial ver	.52***	.30*	.36*	.46**
spatial vis	.25 ⁺	.14	.24	.41**

(two-tailed sig. levels: ⁺ = $p < 0.1$, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$)

Table 3 . 7 - Experiment 2 correlations between the span and reasoning tasks

The verbally presented spatial reasoning task (spatial ver), will be considered first, in relation to the span measures used. Table 3.7 shows the best predictors of this spatial task are the arrow and TTT spans, and to a lesser extent the verbal and word spans. These results are consistent with those from Experiment 1, except for the arrow span correlations. There was no significant correlation between arrow span and this task in Experiment 1 ($r = .10$), but in Experiment 2 the correlation was significant ($r = .52$, $p < .001$). The results here are consistent with predictions made about the relationships between spatial reasoning and working memory. Stronger relationships would be expected with the spatial spans, over and above the verbal spans, due to the spatial nature of the reasoning task. The verbally presented spatial task correlated moderately with all the span tasks indicating that both verbal and spatial resources are required for correct responding. So for the verbally presented spatial reasoning task, the results shown in Table 3.7 are consistent with the view that both spatial and verbal resources of working memory are required in order to interpret and make inferences from spatial descriptions. Due to the complex span measures being strongly correlated, it may be the case that those central resources are domain dependent. In addition, it may be assumed from this data that some sort of passive verbal storage is required (as measured by the word span) in order to perform the task.

Considered next is the visually presented spatial reasoning task (spatial vis), in relation to the span measures used. Table 3.7 shows the best predictor of this spatial task is the complex TTT span ($r = .41, p < .01$). A drop in significant correlations with the other span measures can be seen for this task. This finding may provide support for the suggestion that presenting the problems visually lowers the load put on verbal working memory, explaining the drop in cross-correlations with the verbal spans. The participants did not need to keep in mind the premises as they were presented on screen for the duration of their responding time. However, the correlation between the two spatial reasoning tasks was significant ($r = .69, p < .001$), suggesting that both tasks draw to some extent on common resources. Furthermore, the data reflects a general reduction in the magnitude of correlations between the visually presented spatial task and both spatial and verbal spans, perhaps indicating a general reduction in working memory load. This, together with the fact that there are significant correlations between the complex span measures makes it difficult to clearly interpret these findings. Therefore, these results may be consistent with the view that central resources of working memory are important in accurate responding, but spatial resources are more pertinent when working memory load for the verbal aspect of the task are controlled for.

3 . 3 . 3 . 4 Further Analysis

This section assesses the reasoning tasks based on whether they were easy or difficult in nature. Past research suggests that problems based on one model of the premises are easier than problems based on multiple models of the premises.

	% correct
spatver1M	71
spatverMM	46
spatvis1M	79
spatvisMM	61

Table 3 . 8 - Experiment 2 % of correct responses to 1M and MM problems

Problem difficulty percentages for the spatial tasks are presented in Table 3.8. The 1M problems, both verbally and visually presented, were solved more often than the MM problems. An average of 9 out of 16 (56%) verbally presented spatial inferences were solved, 5 of which were single model problems. This is highly consistent with Experiment 1 (58% spatial ver solved correctly). An average of 11 out of 16 (69%) visually presented spatial inferences were solved, 6 of which were single model problems.

A two by two repeated ANOVA was used for testing the differences among means for model type (1M or MM) and modality type (verbal or visual). The prediction would be that the number of models and type of presentation would significantly effect the spatial inference scores (see Appendix B1.5 for full results). From Table 3.8 it might be expected that one model visual presentation would be the easiest to perform. The analysis revealed a main effect of modality, $F(1,46) = 21.34, p < .001$; a main effect of number of models, $F(1,46) = 49.81, p < .001$; and an interaction between the two, $F(1,46) = 4.20, p < .046$. This finding suggests that the effect of model number depends upon the modality of presentation. A participant's performance on the visual modality task was generally better than on the verbal modality task. This would be expected given that in the visual version of the task, verbal working memory load is reduced. This is consistent with the findings of

Gilhooly, Logie, Wetherick and Wynn (1993) who found similar results with syllogistic inferences. It can be seen that in both the spatial inference tasks, verbal and visual presentations, participants performed at a higher rate on the one model problems, than the multi-model problems. The interaction suggests that when the spatial descriptions are presented verbally, there is a greater memory load, which causes poorer performance on the difficult multiple model problems.

3 . 3 . 4 Discussion

The results from Experiment 2 suggest that individual differences in working memory capacity can explain variation in performance on five-term series spatial tasks, both verbally and visually presented. Experiment 2 also provided a further test of the separability hypothesis. The results showed no clear dissociation between the different span measures and their relationships with different reasoning tasks. Both the verbal and spatial spans predicted individual differences on the reasoning tasks, a similar pattern of correlations was observed, consistent with Experiment 1. Again, correlations between the complex verbal and spatial spans were significant, indicating no clear dissociation. This was a second failure in replicating Shah and Miyake's (1996) results. The positive correlation matrix between the complex span measures in Experiment 2 may imply that a general working memory resource explains variation on the working memory capacity measures. This result is entirely consistent with Experiment 1. This will be further addressed in the factor analysis section at the end of this series of experiments.

There was a replication of the predictors of the verbal modality spatial task, as in Experiment 1, although there is a discrepancy with the arrow span between experiments. The visual modality spatial task was again, generally correlated with all the spans, but the

only significant predictor was the complex TTT span. These findings could be interpreted as follows. There was a reduction in correlations with the verbal span measures when the spatial reasoning task premises were presented on screen. This could have lowered the verbal working memory load, and perhaps participants therefore showed a greater reliance on spatial working memory. The ANOVA analysis supports the notion that visual presentation reduces working memory load due to the fact that in the verbal presentation mode, a reduction in performance on the multiple model problems was seen. However, there was a general reduction in the correlations with both verbal and spatial spans, when the problems were presented visually. This might suggest that a general reduction in working memory load accounts for the change in patterns of correlation. However, it is important to note that these interpretations should all be seen in the context of the correlation between the complex spans, TTT and verbal.

In Experiment 2, and in Byrne and Johnson-Laird's (1989) study, these spatial inferences were investigated and in both the number of models to be constructed was controlled for. The results here confirm Byrne et al.'s (1989) study where it was found that one model problems were reliably easier to solve than multiple model problems. An interaction was also found between the modality of presentation and the number of models to be constructed. This finding could be interpreted as the participants greater reliance on verbal working memory load when the problems were presented verbally, causing poorer performance on the multiple model descriptions. This finding is supported by the work of Gilhooly, Logie, Wetherick and Wynn (1993) who found that presenting syllogistic problems visually improved logical performance. Due to the generally positive correlational patterns between verbal and spatial span measures, there is difficulty in clearly interpreting these findings in the light of reasoning theory. Further discussion of these

theories will follow the presentation of a multi-group factor analysis section, when all three experiments in this series can be related together.

In Experiment 1, the syllogistic inferences (verbally presented) correlated generally with all the span measures used, although not significantly with the complex TTT span. An interpretation of this might be that both verbal and spatial resources are required in order to perform the verbal modality syllogistic task. Verbal presentation requires the reasoner to remember the premise information in advance of processing. The correlations with verbal working memory capacity may in part be explained by the requirement to temporarily store the passive information. As with the spatial inferences, the correlations with verbal working memory are perhaps not unexpected. However, this does not necessarily show that the deductive processes involved in the syllogistic tasks involve similar processes to those involved in the verbal span tasks. Thus, the main aim of Experiment 3, as for Experiment 2, is to examine this by controlling for modality of presentation.

Further investigation was needed to investigate the influence of modality of the syllogistic reasoning task (verbal versus visual presentation). It might be expected that presenting the syllogisms visually will lead to higher accuracy rates in performance. In addition, reductions might be expected in the correlations between verbal working memory and syllogistic inferences, but especially with the simple word span (passive verbal storage).

3 . 4 EXPERIMENT 3

3 . 4 . 1 Aim

The motivation for Experiment 3 is the same as for Experiment 2, but in relation to syllogistic inferences. The analysis from Experiment 1 showed that the syllogistic inferences, when verbally presented, were best predicted by a sub-set of both verbal and spatial working memory spans. By presenting the problems verbally, a high load may have been put on verbal working memory, requiring participants to hold the premise information in mind. Therefore, the correlations between verbally presented syllogistic inferences and verbal working memory are not inconsistent with this interpretation. Experiment 3 aims to examine the patterns of correlational change as a function of modality.

As mentioned in the aim for Experiment 2, Gilhooly, Logie, Wetherick and Wynn (1993) investigated syllogisms, presented both verbally and visually, together with dual tasks proposed to load the working memory sub-components. A significant effect of memory load on accuracy of syllogistic performance was obtained, when presentation modality was manipulated. Presenting the syllogistic task visually increased accurate performance on the problems, indicating that a reduction in memory load may indeed influence responses. Experiment 2 of this series also showed that visually presented spatial inferences were solved more often than verbally presented problems. Additionally, Experiment 2 found a clear model type by modality type interaction, suggesting that verbal presentation mode places a high demand on verbal working memory which, in turn, increases the difficulty of constructing multiple model representations. Experiment 3, in line with Experiment 2, extends this work by presenting syllogistic tasks in both modalities, together with a range of working memory spans.

Experiment 3 was designed as a follow-up to Experiment 1. The syllogistic inferences were investigated both verbally and visually in order to account for presentation differences in performing the task. Experiment 3 also allowed a further test of the separability issue of central resources within working memory. Five working memory span tasks were used: two complex spatial measures, one complex verbal measure, and a simple spatial and verbal measure. The span tasks were given together with two reasoning tasks, a syllogistic inference task presented both verbally and visually. The tasks were as Experiment 1 unless otherwise stated.

As stated previously, the complex spatial letter span task was not used in Experiment 2. There was a longer time scale for carrying out Experiment 3 in relation to Experiment 2. The addition of another task in Experiment 3 therefore would not interfere in the overall time scheme. Thus the letter span was used again here, as well as the TTT span.

3 . 4 . 2 Method

Pilot Study

A pilot study was run to ensure that the new Visual Basic computerised syllogistic task (visual task) was easy to understand, and use.

Participants

Forty-seven participants were used - 15 of whom were men and 32 of whom were women.

The participants were undergraduate students and postgraduate students at the University of Plymouth, and they received course credit or cash payment for participating. None of the sample had prior training in logic, or prior experience on any of the tasks used in Experiment 1 or Experiment 2. They were all native English speakers.

Procedure

Experiment 3 was carried out in two testing sessions per participant, the sessions being of approximately equal length. Participants signed up in a time slot on an experimenter sheet for session one. Session two was arranged individually with each participant at the end of session one. Session one consisted of three tasks. All participants carried out these tasks in the following order (where an asterix (*) indicates a computerised task):-

sylogistic inference measure (verbal presentation),

complex spatial TTT span*,

complex verbal span*.

Session two consisted of four tasks. All participants carried out these tasks in the same following order:-

simple arrow span*,

simple word span*,

complex spatial letter rotation span*,

sylogistic inference measure (visual presentation)*.

What follows is a summary of those tasks used in Experiment 3 which have not already been introduced in prior experiments.

Simple Verbal Word Span (word)

The procedure and scoring was identical to Experiment 1.

Simple Spatial Arrow Span (arrow)

The procedure and scoring was identical to Experiment 1.

Complex Verbal Sentence Span (verbal)

The procedure and scoring was identical to Experiment 1.

Complex Spatial Letter Span (letter)

The procedure and scoring was identical to Experiment 1.

Complex Spatial Tic-Tac-Toe Span (TTT)

The procedure and scoring was identical to Experiment 2.

Syllogistic Inference (Verbal and Visual Presentation) (sylls ver, sylls vis)

Procedure

A series of 16 syllogisms was presented to each participant in both modalities, verbal and visual, and the procedure followed is as Experiment 1. The same 16 problems, a sub-set of the 20 used in Experiment 1, were used for both the verbal and visual task, but the content and order of presentation was changed. The problem content can be found in Appendix A1.7. The order of presentation of problem in both tasks was randomised before the start of the experiment but remained in that order for each participant.

Instructions for the verbal problems: The procedure was identical to Experiment 1.

Instructions for the visual problems: The screen displayed a set of instructions, similar to those for the verbal problems. The instructions explained that in this part of the experiment the participants would be asked to solve a set of syllogisms. A problem would appear on the screen describing the relationship between three classes. Shortly afterwards, a question would also appear on the screen asking them to relate two of the classes within the two statements. The participants were shown by way of an example how to solve a syllogism, and were shown all the quantifiers that were possible in the conclusion: all, some, none.

some....not, and not enough information. The instructions continued with a summary on how to respond to the syllogisms. The participants were told that the screen would also display a blank box in which they were to type their responses to each problem. For example, if they thought that the answer was “All of the bakers are carpenters” then this is what they should type. There was a button to press to move on to the next problem. They were given one practice problem.

Scoring

The scoring was identical to Experiment 1.

3 . 4 . 3 Results

In addition to correlational analyses and ANOVAs, confirmatory factor analysis was also performed on the data from this experiment. The correlational, significance analysis and discussion of findings for this experiment will be reported in this experimental section, but the factor analysis for all of the first three experiments will be explored in a section after Experiment 3.

As in Experiments 1 and 2 the cross-correlations of the global and normal span scores were highly significant. All scoring methods used in the results section are exactly the same as those used in previous experiments.

3 . 4 . 3 . 1 Descriptive Statistics

One participant was dropped due to the reasons outlined in Experiment 1 and 2. Thus, the participant who performed at base levels on all the tasks, and scored 0.0 on the arrow span was dropped and the rest of the analysis was run on 46, not 47, participants. A summary of

descriptive statistics for the measures used in Experiment 3, together with reliability estimates, can be found in Table 3.9.

	Mean	SD	Min	Max	Skewness	Reliability
<i>Span Tasks</i>						
arrow	19.83	11.08	3	42	.38	.89
word	45.63	18.68	20	96	.81	.87
verbal	34.33	19.11	10	94	1.03	.93
letter	2.83	1.25	0	5	-.79	.50
letter err	9.72	7.21	1	41	2.02	-
TTT	7.85	5.07	2	21	.70	.82
<i>Reasoning Tasks</i>						
sylls ver	7.11	2.73	1	13	.37	.66
syllsver1M	5.67	1.71	0	8	-.93	-
syllsverMM	1.43	1.80	0	6	1.08	-
sylls vis	7.96	3.74	0	16	.54	.82
syllsvis1M	5.98	1.82	0	8	-.98	-
syllsvisMM	1.98	2.65	0	8	1.13	-

Table 3.9 - Descriptive statistics for Experiment 3 tasks (n=46)

Again, the span tasks in Table 3.9 show a good overall range of scores and distributions, and are consistent with the results obtained in Experiment 1. The TTT span now shows a similar mean score to that obtained in Experiment 2. The processing component of the letter span was again recorded in this experiment. Participants made on average 9 errors (13% error) in total when making the normal/mirror-imaged judgement in processing, the same as in Experiment 1. This is well below the percentage of 50% that would be expected if a guessing strategy had been used. This result suggests that participants are attempting both the processing and storage components of the complex span measure.

The reliability of the span tasks was assessed by performing a split-half test calculation on each of the five span measures used. The Spearman-Brown formula was then used to provide an estimate of reliability for each task. The corrected correlations for the span tasks are shown in Table 3.9. As Table 3.9 shows, all the span tasks used produced a reliable odd-even corrected correlation of over $r = .75$, except for the letter span. This is an adequate value for group measurement (Rust & Golombok, 1999). The letter span produced an odd-even corrected correlation of $r = .50$. As in Experiment 1, this is the only span task below the adequate reliability level.

Each reasoning measure was also assessed for internal reliability using Cronbach's alpha. As seen in Table 3.9, the reliability estimate for the syllogistic reasoning task, in the verbal modality, was of some value for group measurement ($r = .66$). This is a higher reliability than that seen in Experiment 1 (where $r = .45$). The reliability estimate for the visually presented syllogistic task was highly satisfactory ($r = .82$).

3.4.3.2 The Span Measures

Correlational analysis was carried out on the data in order to investigate whether the tasks were associated together (see Appendix B1.9 for the full matrix). As in Experiment 1 and 2, the first analysis reports the correlations between the span measures. Given the findings of Experiments 1 and 2, the complex span measures might be expected to cross correlate, whilst the simple measures might be expected to show some independence. The results for Experiment 3 span measures can be seen in Table 3.10.

	arrow	word	verbal	letter	letter err
arrow	1				
word	.26 ⁺	1			
verbal	.31*	.66***	1		
letter	.47***	.41**	.48***	1	
letter err	-.24 ⁺	-.26 ⁺	-.32*	-.33*	1
TTT	.40**	.30*	.35*	.49***	-.17

(two-tailed sig. levels: ⁺ = $p < 0.1$, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$)

Table 3 . 10 - Experiment 3 correlations between the five span tasks

As Table 3.10 shows, the highest correlation is between the word and verbal span measures, as in Experiment 1 and 2 ($r = .66$, $p < .001$). The word span also correlated significantly with the complex spatial measures, letter and TTT ($r = .41$, $p < .01$, $r = .30$, $p < .05$), and there was borderline significance with the simple arrow span ($r = .26$, $p < .10$). This is consistent with the result found in Experiments 1 and 2. It may be that the arrow and word spans tap separate storage systems afforded by the VSSP and PL respectively, although they may have some common processing requirement. The arrow span did correlate significantly with the other span measures, this was not found in the first two experiments. In Experiment 1 the arrow span only showed a significant relationship with the letter span, which is consistent with the correlation here ($r = .47$, $p < .001$). What is inconsistent are the relationships between the arrow span and the verbal and TTT spans. The correlations here are significant, whereas this was not found in Experiments 1 and 2. The verbal span correlated with the letter and TTT spans at levels of $r = .48$ and $r = .35$ respectively, reasonably moderate relationships. This was also seen in Experiments 1 and 2 and is inconsistent with a dissociation between verbal and spatial information at the level of a central executive of working memory. Lastly, the two complex spatial spans correlated significantly ($r = .49$, $p < .001$). This result is in line with Experiment 1 and would be

expected if the two are tapping similar resources. The processing error associated with the letter span shows similar results to those obtained in Experiment 1 - there are negative relationships between this error and the accuracy of responding on all the spans. There are significant relationships again between the complex verbal and letter spans with the letter error ($r = -.32$, $r = -.33$, $p < .05$). This suggests that there was no trade off between the processing and storage components on the complex letter span task. There were also borderline significances between the letter error and the simple span measures ($r = -.24$, $r = -.26$, $p < .10$). This is consistent with Experiment 1, although the correlations were not significant, but in the same negative direction.

For the data shown in Table 3.10, the results are generally consistent with those obtained in Experiments 1 and 2. A positive matrix can be seen between all the span measures in Table 3.10. The data indicates there are significant correlations between all the complex span measures used. This result would not be expected if there was a separation of verbal and spatial resources at the level of the CE. The results from this, taken together with Experiments 1 and 2, support the notion that a general working memory resource may underlie variation in the working memory measures. Experiments 1, 2 and 3 all show low relationships between the word and arrow spans. Thus, the data from these experiments also lends some support to the notion of two separate sub-systems of working memory that deal with the relatively passive storage of verbal versus spatial information. However, some common processing requirement on these tasks may also be suggested by the positive correlation between the two.

3.4.3.3 The Span and Reasoning Measures

The next set of analyses examined the relationships between the span measures and the reasoning measures. The reasoning tasks shown in Table 3.11 would be expected to show relationships with the complex measures of working memory. The extent to which syllogistic reasoning tasks relate to particular complex span tasks may lead to a better understanding of the role of verbal and spatial working memory in syllogistic reasoning.

	arrow	word	verbal	letter	TTT
sylls ver	.50***	.57***	.45**	.46**	.45**
sylls vis	.45**	.44**	.50***	.47***	.21

(two-tailed sig. levels: + = $p < 0.1$, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$)

Table 3.11 - Experiment 3 correlations between the span and reasoning tasks

Consider first the verbally presented syllogistic reasoning task (sylls ver), in relation to the span measures used. Table 3.11 shows that this syllogistic task correlated significantly with all the span measures, but the best predictor was the word span ($r = .57$, $p < .001$). These results are consistent with those from Experiment 1, except in Experiment 3 there is a positive correlation with the TTT span. There was no significant correlation between the TTT span and this task in Experiment 1 (.10), but in Experiment 3 the correlation was significant ($r = .45$, $p < .01$). These results, as in Experiment 1, are consistent with findings from previous studies that investigated syllogistic reasoning and working memory.

Gilhooly, Logie, Wetherick and Wynn (1993) found that secondary tasks assumed to load the CE component of working memory, affected syllogistic performance. Relationships with both verbal and spatial measures of working memory would be expected if the syllogistic task were solved using both verbal and spatial strategies (Ford, 1991). The verbally presented syllogistic task correlated moderately with all the span tasks indicating

that both verbal and spatial resources are required for correct responding. However, due to the lack of a dissociation between the complex measures, it is difficult to conclude that there are separate central resources for verbal and spatial information. So for the verbally presented syllogistic reasoning task, the results shown in Table 3.11 are consistent with the view that generally allocated CE resources of working memory, as measured by the complex span tasks, are required in order to interpret and process their meaning. In addition, it may be assumed from this data that some sort of passive verbal storage is required (link with word span) in order to perform the task.

Secondly, the visually presented syllogistic reasoning task (sylls vis), will be discussed in relation to the span measures used. Table 3.11 shows very similar results to the task when verbally presented, although there is a drop in significance with the TTT span ($r = .21$). There is a difference here which is inconsistent with that found for the five-term series spatial task. When presented visually, the spatial task did not correlate as highly with the verbal spans which may suggest that presenting the problems visually may lower the load put on verbal working memory. For the syllogisms, there was no drop off in correlations with the verbal spans when it was presented visually. This might be due to the differences in interpreting the quantifiers (such as some) against interpreting orientations (such as left). It might have to be considered that participants understood the spatial terms better and thus a reduction in the use of verbal resources is shown when they are presented visually. For the syllogisms, even though the participants had the premises on screen in the visual variant of the task, this may not have helped because of the complexity of understanding the quantifiers. Therefore, these results are consistent with the view that both verbal and spatial central resources of working memory are important in accurate responding, when modality is controlled for. However, the complex verbal and spatial span tasks, assumed to measure

CE capacity, correlated significantly, and so any interpretation made from this must account for the commonality between the complex measures. The syllogisms may rely upon a general or domain independent CE, as measured by both the complex verbal and spatial span tasks.

Table 3.11 shows a generally positive matrix where both reasoning tasks are predicted by both the spatial and verbal working memory tasks. The data in Table 3.11 indicates almost identical patterns of results on both syllogistic reasoning tasks, verbal and visual. The cross correlation between the two was also highly significant ($r = .66, p < .001$).

3 . 4 . 3 . 4 Further Analysis

Presentation mode and model type will be investigated in this section. Past research suggests that problems that only need one model to be constructed in order to reach a conclusion should be easier than problems where multiple models are needed.

	% correct
syllsver1M	71
syllsverMM	18
syllsvis1M	75
syllsvisMM	25

Table 3 . 12 - Experiment 3 % of correct responses to 1M and MM problems

Problem difficulty percentages for the syllogistic tasks are presented in Table 3.12. The 1M problems, in both the verbal and visual variant of the syllogistic task, appear to be solved

more often than the MM problems. Table 3.12 shows an average of 7 out of 16 (44%) verbally presented syllogistic inferences were solved, 6 of which were single model problems. This is consistent with Experiment 1 (50% syls ver solved correctly). An average of 8 out of 16 (50%) visually presented syllogistic inferences were solved, 6 of which were single model problems.

A two by two repeated ANOVA was used for testing the differences among means for model type (1M or MM) and modality type (verbal or visual). The prediction would be that the number of models and type of presentation would significantly effect the syllogistic inference scores (see Appendix B1.6 for full results). From Table 3.12 it might be expected that one model visual presentation would be the easiest to perform. The analysis revealed a main effect of modality, $F(1,45) = 4.13, p < .048$; a main effect of number of models, $F(1,45) = 166.81, p < .001$; but no interaction between the two, $F(1,45) = .58, p < .448$.

This finding suggests that the effect of model number does not depend upon the modality of presentation. A participant's performance on the visual modality task was slightly better than on the verbal modality task. This would be expected given that in the visual version of the task, verbal working memory load is reduced. This finding is consistent with that of Gilhooly, Logie, Wetherick and Wynn (1993). It can be seen that in both the syllogistic inference tasks, verbal and visual presentations, participants performed at a higher rate on the one model problems, than the multi-model problems. The interaction showed no effect of the two factors, the effect of modality is the same for both one and multi model problems.

3 . 4 . 4 Discussion

The results from Experiment 3 again suggest that individual differences in working memory capacity can explain variation in performance on syllogistic tasks, both verbally and

visually presented. Experiment 3 also provided a further test of the separability issue of central resources within working memory. Experiment 3 showed no clear dissociation between the different span measures and their relationships with different reasoning tasks. Both the verbal and spatial spans predicted individual differences on the reasoning tasks, a similar pattern of correlations was observed, consistent with Experiment 1.

There were cross-correlations between the complex verbal and spatial spans, indicating a positive relationship, with little dissociation. Shah and Miyake (1996) found a clear dissociation between the spatial spans and spatial information versus the verbal spans and verbal information. Experiment 3 results again failed to replicate this finding. The positive correlation matrix between the complex span measures in Experiment 3 may imply that a domain general central resource of working memory accounts for the relationships found. The results could imply that a common factor underlies individual differences on the measures used.

There was a clear replication of the predictors of the verbal modality syllogistic task, as in Experiment 1, although there is a discrepancy with the complex TTT span. An explanation for this might be that the TTT span used in Experiment 1 had a lower reliability due to it being a non-computerised version. These correlations suggest that modality differences did not give differing results. The ANOVA carried out on this data also suggested no interaction between the modality of presentation and the number of models.

In Experiment 1, both the spatial and syllogistic inferences correlated with both spatial and verbal working memory span tasks (verbal modality). Experiment 2 also supports the fact that spatial inferences in both modalities, were correlated with both verbal and spatial

working memory capacity measures, although the only consistently significant predictor was spatial working memory capacity. Experiment 3 results suggest that there are similar relationships between the syllogistic problems (both verbally and visually presented) and working memory resources, both spatial and verbal. However, any interpretation must be considered in light of the significant relationship between the complex span measures. The findings from these three experiments have implications for the three classes of reasoning theory outlined earlier. Further interpretations of this will take place in the general discussion (Section 3.6) after the factor analysis.

The findings from the first three experiments in this thesis lend support to the notion that working memory capacity predicts variation in performance on both spatial and syllogistic inference. This is an important and very clear finding, and is consistent with processing theories of reasoning (i.e., verbal and analogical) which suggest that reasoning depends upon an explicit system capable of maintaining and processing representations involved in making deductive inferences. At the outset, this thesis was interested in the following question. To what extent do verbal and spatial working memory, as measured by the complex and simple verbal and spatial spans, predict reasoning performance? In particular, given the dual-task work, it might have been expected that verbal working memory correlated most highly with syllogistic reasoning performance, and spatial working memory with spatial reasoning performance. Because there was no clear dissociation between verbal and spatial working memory, it is difficult to make interpretations from the correlational data about the representation in these deductive tasks. Therefore, there is a clear need for a more sophisticated analysis.

The first three experiments in this thesis lends some support to the notion of a central executive of working memory, served by two slave sub-systems for storage of verbal and spatial information. However, for both the simple verbal and spatial spans, as for both complex verbal and spatial spans, there are low positive correlations between them, perhaps suggesting that they also have something in common. The model proposed by Shah and Miyake (1996) consisted of two separate systems for verbal and spatial information at the level of the executive. The results here do not conform to this pattern, where the five-term series and syllogistic tasks were predicted by both complex verbal and spatial span measures showing common influences used by all the tasks. What follows is an investigation of working memory models using a multi-group confirmatory factor analysis. In this way, the data here can be tested against the types of model proposed in the literature (e.g., Baddeley & Hitch, 1974; Shah & Miyake, 1996).

3 . 5 CONFIRMATORY FACTOR ANALYSIS

EXPERIMENTS 1, 2 AND 3

3 . 5 . 1 Rationale for the Factor Analysis

The syllogistic inference data from both Experiments 1 and 3 suggested that individual differences in both the verbally and visually presented tasks were predicted by both complex and simple verbal and spatial working memory spans. The spatial inference data from both Experiments 1 and 2 suggested that individual differences in both the visually and verbally presented tasks were predicted by both complex and simple verbal and spatial spans, although the visual mode problems were only significantly predicted by the complex TTT span. One potential explanation for these findings for both the syllogistic and spatial

reasoning tasks is that they draw on central resources of working memory and that the working memory capacity tasks are also measuring this resource.

Evidence for this is suggested by intercorrelations between the span measures across all three experiments. In all experiments, the complex verbal and spatial spans intercorrelated significantly. Thus, what might underlie this finding is variations in central executive function. There were marginal correlations between the simple word and arrow spans throughout the experimental series (passive stores). This suggests that the simple spans may measure some common resource, but may also show that the simple spans measure the separate storage systems. There was a complete failure to show a dissociation between the complex spatial and verbal spans (as Shah & Miyake, 1996). Therefore, the findings so far are not inconsistent with Baddeley's (1986) model of a central executive with two peripheral subsystems.

The following confirmatory factor analytic (CFA) models were run in order to further test the findings of the correlational data obtained from the first three experiments. A number of different models were tested, including one, two and three factor models. The following summary contains the best fit model found from those that were originally investigated. A three-factor multiple group analysis was run on Experiments 1, 2 and 3. This involved pooling the data from the different studies. Using EQS it is possible to do this even though the measures from the three experiments only partially overlap. For the tasks that were common across experiments, the parameters of the model were constrained for equality. This is an important point given the sample sizes used in relation to individual differences work. By pooling the data across experiments, the sample size was increased overall. Another important point to note is that the use of CFA allows for unreliability in those

measures and methods and for measure specific variance. The following summary also contains a two-factor model, based on that of Shah and Miyake (1996), in order to compare the nine measures used here with an already accepted model of these tasks. The following investigation also covers a re-analysis of Shah and Miyake's (1996) data (see factor analysis discussion).

3 . 5 . 2 Method and Results

A model based on two underlying factors was run before the three-factor model, in line with Shah and Miyake (1996). They investigated an orthogonal two-factor model, where the spatial tasks loaded on one factor and the verbal ones on another. The difference between the Shah and Miyake (1996) model and the one used here is that the reasoning tasks from Experiments 1-3 were loaded on both factors. In this way, the reasoning tasks could be tested to see which resource or factor was most involved in their performance. The set up of the model loadings are shown in Table 3.13. The spatial span tasks were loaded on a spatial factor, whilst the verbal span tasks were loaded on a verbal factor. The reasoning tasks were loaded on both factors. The two factors were investigated both uncorrelated and correlated.

The factors are represented as F1 - verbal resource, and F2 - spatial resource. The loadings in Table 3.13 were constrained to equality across the experiments. The error variance associated with each of the tasks was also constrained across experiments. The appropriate EQS methods were set up to handle the data. Table 3.14 presents the goodness of fit summary for the uncorrelated two-factor model for Experiments 1, 2 and 3.

	F1	F2
arrow		✓
word	✓	
verbal	✓	
verbal process error	✓	
letter		✓
letter process error		✓
TTT		✓
syllogisms - verbal mode	✓	✓
syllogisms - visual mode	✓	✓
spatial inference - verbal mode	✓	✓
spatial inference - verbal mode	✓	✓

Table 3 . 13 - Set up of two-factor model for Experiments 1-3

The goodness of fit summary tables contain multiple fit indices which allow the evaluation of the fit of the model. The chi-square statistic measures the degree to which the covariances predicted by the specified model differ from the observed covariances. A small value for the chi-square statistic indicates no statistically meaningful difference between the covariance matrix generated by the model and the observed matrix, suggesting a satisfactory fit. Akaike's Information Criterion (AIC) is a modified version of the chi-square statistic that takes into consideration the 'complexity' of the evaluated model (in terms of degrees of freedom) and penalises more complex models (models with fewer degrees of freedom). Lower values of AIC, including negative ones, indicate a better fit. In contrast, for the other fit indices, Bentler-Bonett Normed Fit Index, Bentler-Bonett Nonnormed Fit Index, and the Comparative Fit Index, higher values indicate a better fit. These fit indices quantify the extent to which the tested model is better than a baseline model. Typically values that exceed .9 are considered good fits.

chi-square	138.02
degrees of freedom	75
probability of chi-square	0.001
akaike`s information criterion	-11.98
bentler-bonett normed fit index	0.621
bentler-bonett nonnormed fit index	0.767
comparative fit index	0.779

Table 3 . 14 - Goodness of fit summary (two-factor model, uncorrelated) - multiple group analysis of Experiments 1-3

It can be seen from Table 3.14 that the uncorrelated two factor model was a poor fit to the data, giving low fit indices and a significant chi-square value. The Chi-square for testing the goodness of fit of the model is 138.02, based on 75 degrees of freedom. Thus, there was a significant discrepancy between the model and the data. The fit indices were below an acceptable level (Dunn, Everitt & Pickles, 1993). Recall that this was the model used by Shah and Miyake (1996) to account for their data set. Even though there are measures they used that overlap with the current set, this model does not hold for this data. Table 3.15 presents the goodness of fit summary for the correlated two-factor model for Experiments 1, 2 and 3.

The correlated two factor model in Table 3.15 faired better, giving higher fit indices but still with a significant chi-square value. There are significant discrepancies between the two factor models and the data from Experiments 1-3. so both were rejected as incompatible.

chi-square	97.016
degrees of freedom	72
probability of chi-square	0.026
akaike's information criterion	-46.98
bentler-bonett normed fit index	0.733
bentler-bonett nonnormed fit index	0.904
comparative fit index	0.912
correlation between factors	0.77

Table 3 . 15 - Goodness of fit summary (two-factor model, correlated) - multiple group analysis of Experiments 1-3

Due to the significant chi-square values for the two-factor models, both with and without correlated factors, a three-factor model was further investigated. The general assumption underlying the three-factor model was to test the data based on the Baddeley model. The span measures were loaded onto particular factors to identify them as different stores or resources. For instance the word span was loaded onto factor 1, together with the verbal span (passive verbal resource). The verbal and word spans were additionally loaded onto factor 3 (general resource). In contrast, the arrow span was loaded onto factor 2, together with the TTT and letter spans (passive spatial resource). The arrow, TTT and letter spans were additionally loaded onto factor 3 (general resource). The reasoning measures were loaded onto all three factors in order to identify which resources were most associated with them. A multiple group confirmatory three-factor analyses was run on all three experiments. The set up of the model is shown in Table 3.16.

The factors are represented as F1 - verbal resource, F2 - spatial resource, and F3 - general resource. All the above loadings were constrained to equality across the experiments. The error variance associated with each of the tasks was also constrained to equality across

experiments. The three factors were not allowed to correlate. Note that the complex TTT span was treated as the same task in the computerised and non-computerised versions. It needs to be acknowledged that this model is less parsimonious than the previous two-factor models. There are more parameters in this model and therefore it will fit the set of data easier. The appropriate EQS methods were set up to handle the data. Table 3.17 presents the goodness of fit summary for this model.

	F1	F2	F3
arrow		✓	✓
word	✓		✓
verbal	✓		✓
verbal process error			✓
letter		✓	✓
letter process error			✓
TTT		✓	✓
syllogisms - verbal mode	✓	✓	✓
syllogisms - visual mode	✓	✓	✓
spatial inference - verbal mode	✓	✓	✓
spatial inference - verbal mode	✓	✓	✓

Table 3 . 16 - Set up of orthogonal three-factor model for Experiments 1-3

Table 3.17 gives the results of fitting the hypothetical three-factor model outlined above. The three factors were orthogonal in nature. The Chi-square for testing the goodness of fit of the model is 66.269, based on 66 degrees of freedom. This result may be likely ($p < .47$) to occur with a sample of this size if the model truly holds in the population. A fit index of above .9 suggests that data fit the model well (Dunn, Everitt & Pickles, 1993). This was the

best set of fit indices out of the three models tested here, and was the only one to achieve a non-significant chi-square probability, which further strengthens the result.

chi-square	66.269
degrees of freedom	66
probability of chi-square	0.47
akaike's information criterion	-65.73
bentler-bonett normed fit index	0.818
bentler-bonett nonnormed fit index	0.999
comparative fit index	0.999

Table 3 . 17 - Goodness of fit summary (three-factor model, uncorrelated) - multiple group analysis of Experiments 1-3

Tables 3.17A-C present the factor loadings of each of the nine tasks on the three factors.

MEASURE	LOADINGS			R-sq
	Factor 1 verbal store	Factor 2 spatial store	Factor 3 general store	
arrow		.57***	.33***	.44
word	.48***		.64***	.63
verbal	.47***		.64***	.63
verbal error			-.46***	.21
letter		.38***	.59***	.49
letter error			-.44***	.19
TTT		.10	.60***	.37
syllogisms - verbal mode	.31**	.52***	.42***	.54
spatial inference - verbal mode	-.18	.14	.72***	.57

(significant loadings * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.005$)

Table 3 . 17A - Experiment 1 factor loadings for the uncorrelated three-factor model

MEASURE	LOADINGS			R-sq
	Factor 1 verbal store	Factor 2 spatial store	Factor 3 general store	
arrow		.57***	.33***	.44
word	.48***		.64***	.63
verbal	.47***		.64***	.63
TTT		.10	.60***	.37
spatial inference - verbal mode	-.18	.14	.72***	.57
spatial inference - visual mode	-.60***	-.17	.78***	1.00

(significant loadings * = p < 0.05, ** = p < 0.01, *** = p < 0.005)

Table 3 . 17B - Experiment 2 factor loadings for the uncorrelated three-factor model

MEASURE	LOADINGS			R-sq
	Factor 1 verbal store	Factor 2 spatial store	Factor 3 general store	
arrow		.57***	.33***	.44
word	.48***		.64***	.63
verbal	.47***		.64***	.63
letter		.38***	.59***	.49
letter error			-.44***	.19
TTT		.10	.60***	.37
sylogisms - verbal mode	.31**	.52***	.42***	.54
sylogisms - visual mode	.64***	.75***	.14	1.00

(significant loadings * = p < 0.05, ** = p < 0.01, *** = p < 0.005)

Table 3 . 17C - Experiment 3 factor loadings for the uncorrelated three-factor model

- **The span measure loadings**

The following summary relates to Tables 3.17A-C. The tables represent the data from Experiment 1, 2 and 3 data which was constrained as a multiple group analysis. explained previously (see Appendix B1.13 for the full analysis). The loadings show that the word and verbal spans significantly load factor 1 and factor 3 (verbal and general resources). The

loadings for the word and verbal spans on factor 3 are identical suggesting that the complex task is no more effective in tapping factor 3 than is the simple task. This may mean that the word span, mostly thought to rely on simple passive resources, may also have a processing component, and vice versa for the verbal span. Experiments 1-3 correlational analyses also revealed that the arrow and word span had commonalities as represented by their positive, sometimes significant, relationship. Due to the issue of processing, it might be the case that the word span does not only measure a passive resource, but is part of the active resource component of working memory. It may be that factor 1 and factor 2 are tapping into domain specific parts of the CE, rather than necessarily mapping onto the PL and the VSSP. A consideration of how to interpret factor 1 and factor 2 follows in the discussion. The arrow and letter spans significantly load factor 2 and factor 3 (spatial and general resources). The TTT span only significantly loads factor 3.

The three-factor CFA for Experiments 1-3 successfully identified a spatial and verbal factor. The simple and complex verbal spans loaded the verbal factor. The simple and complex spatial spans loaded the spatial factor (except TTT span). However, all the verbal and spatial span measures also loaded a general factor.

- **The reasoning task loadings**

The syllogisms, verbally presented, significantly load all three factors. The syllogisms, visually presented, lose significance when loading factor 3. The spatial inferences, in the verbal mode, show significance only on the factor 3 loading. The spatial inferences, visually presented, significantly load factor 1 and factor 3, although the loading on factor 1 is negative. Interestingly, eight of the nine tasks significantly load factor 3 (the general

store). Also of interest, when processing error is taken into consideration, verbal and letter error load significantly and negatively on factor 3.

The R-squared values account for how much of the variance in the measures are explained by the factors. These values are generally lower than the split half reliabilities from Tables 3.1, 3.5 and 3.9. This may suggest that there is systematic variance in the measures which is not down to the variance of working memory factors, but may be task specific. This could indicate that one participant may have a better strategy over and above another participants choice of strategy. Due to the fact that this was a multiple group CFA, the problem of task specific variance may have been overcome. Using multiple measures will alleviate variance specific to one measure.

The only R-squared values that were a problem were in relation to the visual variants of the spatial and syllogistic tasks (see Table 3.17B and C). These values imply that the error variance was fixed at 0.00. This zero error variance estimate has been called the Heywood case and has been found to be induced from sampling fluctuations (Harman, 1971). The visual variants of the spatial and syllogistic task were the only reasoning measures to appear only once in the data set, and thus the sample size was small. This might be the reason for this discrepancy (Dillon, Kumar & Mulani, 1987).

An important point to note is that the R-squared values for the reasoning measures are similar to those for the working memory measures. The reasoning measures R-squared values ranged between .54 and .57 (except the Heywood cases outlined above). The R-squared values for the span measures ranged between .37 and .63. This suggests that reasoning is as good a way to measure working memory capacity as using the working

memory measures themselves. Thus, individual differences in reasoning performance are not much more than individual differences in working memory capacity measures. This emphasises the strength of the relationship between working memory and reasoning.

3 . 5 . 3 Factor Analysis Discussion

The confirmatory multiple group three-factor analysis carried out for this series of experiments suggests that some form of general factor of working memory capacity is implicated in all higher level thinking, be it verbal or spatial in nature. The data analysis and discussion below support the notion of a centrally controlled executive, but the interpretation of factor 1 and factor 2 as the two relatively passive storage systems is questionable.

Firstly, a comparison needs to be made between the two-factor and three-factor models investigated above. It can be seen from the goodness of fit summaries that both the two-factor models, either correlated or uncorrelated, did not fit the data significantly. Shah and Miyake (1996) performed an exploratory factor analysis on their nine measures from their Experiment 1. They found that a two-factor orthogonal model fitted their data best. The first factor was loaded highly on the spatial ability tasks and spatial spans. The second factor was loaded highly on the verbal ability tasks and verbal span. They interpret these results in terms of two separate pools of domain-specific resources that support both the processing and maintenance of spatial and language information. The results of fitting the two-factor model (orthogonal) to the data from the current series does not conform to this picture.

In an attempt to resolve this discrepancy, a CFA was run using Shah and Miyake's (1996) Experiment 1 data. This was in order to assess their results in detail by looking at two different models. Firstly, their data was re-run as a CFA two-factor uncorrelated model with the set up of the model corresponding to the reported EFA in their paper. Secondly, their data was assessed as a CFA three-factor model as in Experiments 1, 2 and 3. This would enable comparisons between best fit model types for their own data, as well as comparisons with the present data set.

Shah and Miyake's (1996) original two-factor EFA was re-run using the methods of EQS in order to ensure that the resultant two-factor orthogonal CFA model was consistent with the methods they used previously (BMDP 4M program). Tables 3.18A and 3.18B give the results of fitting the hypothetical two-factor orthogonal model outlined in their Experiment 1 to EQS methodology.

chi-square	26.563
degrees of freedom	26
probability of chi-square	0.43
akaike's information criterion	-25.44
bentler-bonett normed fit index	0.861
bentler-bonett nonnormed fit index	0.995
comparative fit index	0.996

Table 3 . 18A - Goodness of fit summary (two-factor model, uncorrelated) - analysis of Shah and Miyake's (1996) data

MEASURE	LOADINGS	
	Factor 1 verbal	Factor 2 spatial
Simple arrow span		.69***
Complex verbal span	.54	
Complex spatial letter span		.74***
Paper form board test		.60***
Spatial relations test		.90***
Clocks test		.71***
Quantitative SAT		.75***
Verbal SAT	.83	
Identical pictures	.20	.32*

(significant loadings * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.005$)

Table 3 . 18B - Shah and Miyake`s (1996) data factor loadings, a two-factor orthogonal model

Table 3.18B gives the factor loadings of the two factor CFA carried out. Note that none of the loadings of factor 1 are significant. Table 3.18A and 3.18B are consistent with the original findings of Shah and Miyake`s (1996) model. Table 3.18B shows a good model, but nothing loaded significantly on factor 1, indicating that they may not have needed this extra factor. Due to similar results being obtained through EQS methodology in comparison with Shah and Miyake`s (1996) methodology, it can be assumed that the two programs are compatible in nature.

Table 3.19A gives the results of fitting the hypothetical three-factor model outlined above to the Shah and Miyake (1996) data, where all the tasks additionally loaded a third factor (general). The three factors were orthogonal in nature. The Chi-square for testing the goodness of fit of the model was 10.730, based on 17 degrees of freedom. This result may

be likely ($p < .87$) to occur with a sample of this size if the model truly holds in the population. Thus, considering the high fit indices, the model was accepted.

The conclusion that the three-factor model (Table 3.19) fitted the data better than the two-factor model (Table 3.18) for Shah and Miyake's (1996) data was further supported by the direct statistical comparison of alternative models. The chi-square difference test produced a significant chi-square, $\chi^2(9) = 15.83$, $.05 < p < .10$, suggesting that the two-factor model fits the data significantly worse than the three-factor model did and hence must be rejected. It has to be noted that the three-factor model is less parsimonious than the two-factor model, but with a significant chi-square difference result, the finding is strengthened.

chi-square	10.730
degrees of freedom	17
probability of chi-square	0.87
akaike's information criterion	-23.27
bentler-bonett normed fit index	0.944
bentler-bonett nonnormed fit index	1.085
comparative fit index	1.000

Table 3 . 19A - Goodness of fit summary (three-factor model, uncorrelated) - re-analysis of Shah and Miyake's (1996) data

Table 3.19B gives the results of the factor loadings of each of the nine tasks on the three factors. The simple arrow and complex letter spans significantly load factor 2 (spatial store). The complex verbal span significantly loads factor 3 (general store). All of the spatial ability tests significantly load factor 2. The verbal ability tests significantly load factor 1 (verbal store), where previously they did not. A proportion of the spatial and verbal ability tests also load factor 3. Identical pictures significantly load factors 1 and 2.

MEASURE	LOADINGS		
	Factor 1 verbal store	Factor 2 spatial store	Factor 3 general store
Simple arrow span		.73***	.01
Complex verbal span	.28		.35*
Complex spatial letter span		.73***	.14
Paper form board test		.59***	.12
Spatial relations test		.84***	.32
Clocks test		.63***	.38*
Quantitative SAT		.67***	.37*
Verbal SAT	.59**		.81***
Identical pictures	.82***	.49**	-.30

(significant loadings * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.005$)

Table 3 . 19B - Shah and Miyake's (1996) data factor loadings, a three-factor orthogonal model

The factor loadings for the span tasks in this three-factor model of Shah and Miyake's (1996) data are consistent with those of Experiments 1, 2 and 3. Both the simple arrow and the complex letter spans significantly loaded the spatial factor in both models. They were also positively loaded on the general factor in both models. The complex verbal span was positively loaded on both the verbal and general factors in both models.

The other tasks used in the models were not comparable in the same way. Shah and Miyake (1996) did not use many of the tasks seen in the present series, including some of the span and all of the reasoning measures. But all of the ability measures used by Shah and Miyake (1996) positively loaded the general factor (except identical pictures) in the three-factor model. If factor 3 is to be described as a general resource then these ability tasks seem to be reliant on it as well as a specific resource. Deductive tasks would be expected to require active processing and specific storage resources for accurate performance. It seems that

even ability measures may require active processing over and above the specific resource proposed by Shah and Miyake (1996).

Turning now to the present data set (Experiments 1, 2 and 3), the uncorrelated two-factor model was not able to fully explain the data from the nine tasks used here. Thus, a correlated two-factor model was run to investigate whether there was a relationship between the two factors. There was a highly significant correlation between what were called the verbal and spatial resources ($r = .77$) in this version of the model. This may imply some common processes underlying both the verbal and spatial resource. This finding disagrees with that of Shah and Miyake (1996) and will be further discussed in the general discussion at the end of the full set of experiments. The high correlation between the verbal and spatial resources could reflect a common resource underlying them both. Therefore, it was decided that a three-factor model should be tested.

The three-factor model was indeed the only model that produced a non-significant chi-square in explaining the data sets from Experiments 1, 2 and 3. It was also the best model in explaining the Shah and Miyake (1996) data set. In this model the correlation represented in the two-factor model was replaced by a general resource factor. This three-factor model is partially compatible with the working memory model put forward by Baddeley and Hitch (1974). The verbal resource might represent the storage of information in the PL. The spatial resource might represent the storage of spatial information in the VSSP. The general resource might represent the storage and processing of verbal and spatial information in the CE.

The only problem with this explanation are the loadings of the word span and verbal span tasks on the factors assumed to measure the PL and CE components of working memory. The traditional view of working memory would expect the word span to load more on factor 1 (PL) than factor 3 (CE), and vice versa for the verbal span. This traditional view is more supported when looking at the spatial spans, where the arrow span loads factor 2 more than the complex spatial spans, and vice versa for factor 3. The fact that there are similar loadings on the specific versus general factors, especially for the verbal span tasks, might lead to the conclusion that what was originally thought of as passive storage systems (F1 and F2) are actually domain specific components of the CE resource. This is an alternative position to the one put forward in the last paragraph. It might be the case that the three-factor model represents a totally active processing and storage system that is more likely to be domain specific parts of the CE. Even Shah and Miyake (1996) admit that other models of working memory with different architectural assumptions might also be compatible with their results. They acknowledge that Baddeley's (1986) model could account for their findings by attributing more active functions to the spatial component of the peripheral subsystems than they assume.

As summarised in Chapter 1, La Pointe and Engle (1990) investigated the relationship between a simple (word) and complex (reading) span and reading comprehension. This study can be related to the finding that the word span loads the general resource as well as the verbal resource. La Pointe et al. (1990) showed that the simple word span significantly predicted reading comprehension, and that the correlations were as high as those obtained for the complex reading span. Previous experiments have not found a relationship between simple word span and comprehension (Daneman & Carpenter, 1980; Turner & Engle, 1989). La Pointe et al. (1990) suggested that the complex and simple tasks may not be

greatly different in what they measure, but it is unclear what they do measure that is important to higher level cognition. It is suggested that the crucial similarity is in relation to some form of articulatory coding. This interpretation is that the phonological input store plays a role in comprehension, but possibly only for demanding material. This finding could be mapped onto the data here. It could be that the complex verbal span has a PL component, and this would explain its loading on factor 1. What this does not explain, however, is the loading of the simple word span on factor 3. It could be the case that the passive word span contains elements of processing more in line with an active CE component. Thus, an interpretation of this data is that both the simple word and complex verbal spans consist of a PL and CE component.

The explanation of the reasoning task loadings on the three factors will be considered next. From the verbally presented syllogistic reasoning loadings, it can be seen that performance relied on the verbal, spatial and general working memory resources. The visually presented syllogisms also loaded the verbal and spatial factors, but not significantly on the general factor. The ANOVA identified that a main effect of modality did influence syllogistic reasoning performance. Therefore, a reduction in the loadings on the verbal factor might have been expected for the visually presented task. An explanation stated previously for this was that participants may have to keep in mind the premise information, which logically relies on a verbal working memory resource, and then formulate a mental model that holds spatial (or propositional) information from the problem, which may rely on a spatial working memory resource. Due to the verbal and visual variants of the task loading both verbal and spatial factors, an alternative explanation might be more appropriate. This finding is most consistent with recent work into strategies, that suggests that syllogisms may be solved by using spatial and/or propositional representations (Ford, 1994).

The spatial reasoning data suggests that performance relied most heavily on the general resource. It does not seem to rely on the spatial resource as would be expected. This can also be seen for the TTT span, which also only loaded significantly on the general resource. This is consistent with the notion that both require the maintenance and processing of information, but that this is achieved over and above the level of peripheral subsystems. This is consistent with the notion that a high proportion of individual differences in spatial inference is accountable to centrally located working memory resources.

It has already been noted that the R-squared values for the reasoning measures are similar to those for the working memory measures. Therefore, the same predictions could be made from the reasoning measures, as could be made from the working memory capacity measures. The reasoning tasks would predict performance on working memory span tasks, and vice versa. This result highlights the strength of the relationship between working memory and reasoning ability. Individual differences in working memory capacity may be little more than individual differences in reasoning ability.

The modality differences were not clear for either the syllogistic or spatial inference data. There were no major differences in the patterns of loadings as a function of the mode of presentation, except for the spatial inference verbal/visual tasks loadings on the verbal resource (-.18 vs. -.60*, respectively). The visual modality inferences were more heavily influenced by the verbal resource factor than were the verbal modality problems. The direction of this influence was in the opposite direction to the verbal span measures association with this resource. An explanation for the negative loading on the verbal factor of the visually presented spatial inferences might be that people with good verbal stores makes them worse at this task. It may tempt those people into using a verbal strategy that

doesn't help in the long run. However, this negative loading is most marked for the visually presented spatial inferences, in contrast to the verbally presented task. An explanation for this could be because, in the verbal presentation mode, participants use verbal resources to keep the premise information in mind. Therefore, there might not be any verbal resources left for using an inappropriate verbal strategy. In the visual presentation mode, there may be more resources available for using an inappropriate verbal strategy. Related findings have been reported elsewhere. For example, Phillips, Wynn, Gilhooly, Della Sala and Logie (1999) found that articulatory suppression enhanced performance on the Tower of London task. Their results suggested that preventing verbal rehearsal in the PL during planning encouraged the use of an optimal spatial strategy involving the VSSP. The negative loading found here similarly suggests that reasoners using a verbal strategy may have performed worse overall, when an alternative strategy would have improved their performance.

Lastly, the results obtained from entering the processing error into the analyses are informative. Both errors, verifying sentences as true or false and verifying letters as mirror-imaged or normal, loaded negatively on the general, or central resource. This lends support to the notion that the general resource is a processing factor. Also, error associated with the span tasks should load the factor in a contrasting manner to the accuracy scores of the nine measures used. All the tasks loaded positively on the general factor. Therefore this result shows that as performance levels get better on the tasks, the error associated with them actually decreases. This strengthens the three-factor model for the tasks due to the opposite loadings for error on the central resource. These negative loadings of error on factor 3 identifies the need to look at both storage and processing components of the tasks. The verbal span and letter span errors correlated highly ($r = .49$) in Experiment 1 and this says something about the generality of factor 3. The processing component of these tasks

involves accessing background knowledge, for instance knowledge of orientations, and then integrating this with a representation of the problem at hand. This is assumed to be a core function associated with the CE component of working memory and is consistent with the traditional type working memory model (e.g., Baddeley, 1986). The traditional model of the CE is that it deals with the processing of all information in the same way (domain-general viewpoint). This finding also emphasises the need to investigate both storage and processing components when using either the reading span or letter span tasks.

3 . 6 GENERAL DISCUSSION FOR EXPERIMENTS 1-3

A central goal of the current set of studies was to demonstrate that individual differences in working memory capacity can explain variation in performance on syllogistic and spatial reasoning tasks. The pattern of correlations in Experiments 1-3 support this notion. The correlational patterns demonstrated that the deductive tasks were best predicted by both simple and complex capacity measures (requiring either storage or additional processing components for accurate responding). The factor analysis also yielded results suggesting that a general factor of working memory capacity underlies both these deductive tasks, regardless of the presentation modality.

Another aim of this set of experiments was to investigate the separability hypothesis for central resources of working memory (e.g., Shah & Miyake, 1996). The patterns of correlations in Experiments 1-3 are inconsistent with this view. The correlations between the complex capacity measures ranged significantly between .31 and .48. The correlations between the simple capacity measures were of lower magnitude, and ranged between .23 and .28. These results are consistent with the notion of a general central executive, as measured by the complex spans, with two relatively independent passive systems for verbal

and spatial information, as measured by the simple spans (Baddeley, 1986). The interpretation of the factor analysis also suggests that the deductive tasks require a centrally controlled resource.

As briefly reviewed earlier, the dual-task work identified that performance on simple deductive tasks was only affected by similar secondary tasks assumed to load the relatively passive storage systems of working memory (e.g., Farmer, Berman & Fletcher, 1986).

Work involving performance on more complex deductive tasks, such as syllogisms and spatial inferences, demonstrated that they were most affected by secondary tasks assumed to load the central executive of working memory (e.g., Gilhooly, Logie, Wetherick & Wynn, 1993; Vandierendonck & De Vooght, 1997).

The results from Experiments 1-3 are consistent with these findings. The syllogistic tasks were associated with both complex verbal and spatial working memory span tasks. In addition, the factor analysis demonstrated that, regardless of modality, syllogistic inference performance relied upon both the verbal and spatial factors, and in addition the verbal modality problems loaded the general factor. This is consistent with studies that suggested a link between syllogistic reasoning and the PL and CE components of working memory (e.g., Gilhooly, Logie, Wetherick & Wynn, 1993). However, the syllogisms also loaded the spatial factor in the CFA. This result may be consistent with studies that identified a sub-set of participants who use verbal strategies against a sub-set of participants who use spatial strategies (e.g., Ford, 1991). Experiment 3 also found a difference between the verbally and visually presented syllogisms consistent with Gilhooly et al. (1993) who found better performance rates when the premises were available for inspection. However, some of the findings here do not sit as comfortably with Gilhooly et al.'s (1993) results. Their

explanation of dual-task data was that participants were using heuristic strategies (matching or atmosphere) which placed low demands on the slave systems but enough of a load on the CE to be disrupted by random generation. The present results suggest that syllogistic reasoning performance requires more than CE processing and supports the view that they require verbal and spatial components plus a general processing component of working memory in order for accurate responding.

The spatial tasks were associated with both complex verbal and spatial working memory span tasks. The factor analysis demonstrated that, regardless of modality, spatial inference performance relied upon the general factor, and to some extent negatively on the verbal factor (for the visual modality task). This is consistent with studies that suggest a link between spatial reasoning and the CE component of working memory (e.g., Vandierendonck & De Vooght, 1997). Experiments 1 and 2 showed differences between the verbally and visually presented spatial inferences in line with Gilhooly, Logie, Wetherick & Wynn's (1993) syllogistic study. An interpretation of this might be that presenting the spatial problems visually reduced the need for verbal resources of working memory. The factor analysis also identified a negative loading of the spatial problems on the verbal factor, when they were visually presented. An explanation of this might be that, in the visual modality, participants have more verbal resources available for using an inappropriate verbal strategy. The findings for the spatial inference task support the view that generally they require an active processing component of working memory in order for accurate responding.

Past research also shows that complex tasks such as reading comprehension are best predicted by span tasks assumed to measure the CE resource (e.g., Daneman & Carpenter,

1980). The findings here are consistent with this, but also show that complex reasoning tasks require both the storage and maintenance of information, over and above the peripheral subsystems of working memory. This is highlighted in the factor analysis where both the syllogistic and spatial tasks loaded most heavily on a general factor.

The findings from Experiments 1-3 support a unitary view of working memory resources. There was no clear dissociation between the complex verbal and spatial span measures. Thus, the results are in line with accounts suggesting that information is stored and maintained in an all purpose, not domain specific processing system (as Turner & Engle, 1989). The correlational analysis showed no clear evidence of a separability of resources at a centrally controlled level (as Shah & Miyake, 1996). Having said this, there may be an indication of a separation of central resources depending on the interpretation of the factor analysis. The two factors originally assumed to represent the PL and VSSP may actually relate to domain specific areas of the CE. This is because the complex verbal and simple word span loaded similarly on both the verbal and general stores. If the two factors were measuring the capacity for passive storage of information, the verbal span should only relate to factor 3 (general) and the word span to factor 1 (verbal). Thus, there may be some indication of the separation of information within the CE, as well as there being a domain general component.

The brief review at the start of these experiments also explored the relation between these tasks and the interpretations that might be made by three classes of reasoning theorists. The first and second class of theory were introduced as processing theories. The verbal accounts (e.g., Rips, 1994; Polk & Newell, 1995; Hagert, 1984) all rely on the notion of linguistic or verbally based processes for performance on syllogistic and spatial inferences. For instance,

Rips (1994) proposed that reasoning performance depends on the participant's ability to generate internal abstract rules of proof. A special case was also highlighted in relation to the Verbal Reasoning model of Polk and Newell (1995). They proposed that conclusions are derived from encoding and reencoding the premises, processes that are also required for verbal comprehension, and that this occurs without any separate or special process of reasoning. The spatial or analogical accounts (e.g., Johnson-Laird, 1983) all rely on the notion that reasoners form mental models representing possible states of affairs consistent with the premises. The third class of theory, the heuristic accounts of reasoning, propose that syllogistic performance relies entirely on noninferential mechanisms (e.g., Woodworth & Sells, 1935; Wetherick & Gilhooly, 1990; Chater & Oaksford, 1999).

Byrne and Johnson-Laird (1989) suggested a spatial representation explanation when performing spatial reasoning tasks. Experiments 1 and 2 are consistent with this view because of the high association between the spatial spans and especially the visual variant of the task. However, the CFA results suggested that spatial memory, specifically, is not a limiting factor in these kinds of task. This issue will be discussed more fully in the following chapter, where the relationship between spatial reasoning and working memory is further examined. Ford (1991) assumed the need to look at both verbal and spatial strategies in relation to solving syllogistic tasks. Experiments 1 and 3 are consistent with this view where the syllogistic task, in both modalities, loaded the verbal and spatial factors.

All three experiments are inconsistent with the verbal class of theory, where single deterministic verbal processes are proposed. Spatial working memory tasks were predictive of performance on both the spatial and syllogistic tasks, a finding that would not have been expected by this class of theory. These findings are also inconsistent with the Verbal

Reasoning theory that relies on the notion that reasoning is not resource intensive and does not require any special processing over and above the type used in tasks such as verbal comprehension. Working memory span tasks, assumed to measure the capacity of an explicit processing system, did predict individual differences on the reasoning tasks. However, all these interpretations must be made in light of the correlations between the complex span measures, and the emergence of a general factor in the CFA. These results do not clearly differentiate between process theories of reasoning. They do, however, have implications for heuristic accounts.

The heuristic accounts rely upon the notion of an implicit processing system, and do not assume the use of conscious deductive reasoning processes (e.g., Wetherick & Gilhooly, 1990; Chater & Oaksford, 1999). Wetherick and Gilhooly (1990) proposed that people use heuristic strategies, such as matching, rather than actually reasoning syllogistically. Chater and Oaksford (1999) provided a set of heuristics that operate under a general principle of information gain to generate candidate conclusions which are then evaluated for, and chosen on the basis of, their informativeness. If participants were following low load strategies, such as matching, then there would be no reason to expect a significant correlational relationship between tasks proposed to measure the capacity of an explicit processing system and deductive reasoning tasks. This correlation would not necessarily be expected by the heuristic theories because they assume there is no special requirement for this kind of resource intensive processing. The results of Experiments 1-3 showed that measures of the efficiency of an explicit system (working memory spans) can predict logical performance on the reasoning tasks, both syllogistic and spatial. The heuristic accounts must be able to explain why these measures of working memory span predict individual differences in performance on deductive reasoning tasks.

From the further analysis sections of Experiment 1-3 results, it can be seen that the one model problem types were significantly easier to solve than the problems requiring more than one model to be constructed. This is consistent with most theories which rely on the notion of a limited capacity working memory as a constraint on processing and maintaining information (e.g., Johnson-Laird, 1983; Rips, 1994).

Experiments 1, 2 and 3 investigated the degree to which individual differences in working memory can explain the variation in performance on five-term series spatial and syllogistic inference tasks. Thus, Experiments 1-3 were limited to only two deductive tasks. As was outlined in Chapter 2, Stanovich and West (1998) have recently studied a range of deductive and problem-solving tasks and found that they were all cross-correlated, and could be predicted from the combined influence of cognitive ability and a reasoner's thinking style. Experiments 4, 5 and 6 will examine the same types of working memory span measures, but extend their relationships to additional reasoning tasks such as temporal, conditional and abstract reasoning measures. The working memory span measures are valuable here because they have been shown to differentiate the reasoning tasks at the level of the sub-systems (or domain specific CEs). The complex span measures would be expected to cross correlate again and give some indication of CE processing in a wider range of reasoning tasks.

CHAPTER 4 - WORKING MEMORY AND DEDUCTIVE REASONING: COGNITIVE CAPACITY AND THINKING STYLE

4.1 GENERAL INTRODUCTION FOR EXPERIMENTS 4-6

Chapter 4 presents three experiments which extend the findings from Experiments 1-3 to a wider range of reasoning tasks. The degree to which individual differences in working memory can explain the variation in performance on five-term series spatial and syllogistic inference tasks was investigated in Chapter 3. Chapter 4 will examine the same working memory measures and their relationships with additional reasoning tasks such as temporal, conditional and abstract/propositional deductive measures. The reasoning and capacity measures will be investigated in the context of the work of Stanovich and colleagues. Stanovich and West (1998) used cognitive ability tests and thinking disposition questionnaires as predictors of accurate reasoning performance. Experiments 4-6 will use working memory capacity tests as an alternative way of measuring cognitive ability in order to predict correct reasoning responses.

From Experiments 1-3 of this thesis, it was concluded that individual differences in five-term series spatial and syllogistic reasoning were associated with individual differences in a range of working memory measures. The pattern of findings were consistent with a unitary view of executive function. The factor analysis showed a dissociation between the verbal and spatial resources, but a general factor also emerged. Two explanations were put forward as interpretations of this finding. The first was based on the Baddeley perspective, where the general factor related to the CE and the verbal and spatial resources to the PL and VSSP, respectively. The second perspective accounted for the fact that the complex and passive span measures loaded similarly on what had originally been thought of as the

subsystem factors, the PL and VSSP. Consequently, this interpretation related all three factors to the CE, with general and specific components of executive processing. This explanation contrasts with that put forward by Shah and Miyake (1996). They presented a factor analytic model compatible with separate central resources of working memory, over and above the level of the subsystems. A general factor did not emerge from their analysis. These separate systems accounted for the maintenance and processing of spatial and verbal information as autonomous systems. This contrasts with the model found here where spatial and verbal information showed no clear independence, and additionally a general factor also emerged. The primary aim of this set of experiments is to extend the findings from the three-factor working memory model from Experiments 1-3 to a wider range of reasoning tasks. In this way, further justification of this working memory model may be obtained. Simpler and more complex reasoning tasks will be investigated, in addition to those already looked at, to try and differentiate the factors found in the model outlined above.

Stanovich and West (1998) attempted to demonstrate that patterns of individual differences across cognitive tasks can have important implications for current debates about what it means to be rational. They investigated a wide range of reasoning tasks and their relationships with measures of cognitive ability and a person's thinking disposition.

Stanovich and colleagues have studied a multitude of tasks from the heuristics and biases literature, and investigated whether, at least for certain classes of task, there are significant cross-task correlations. The direction of these correlations could give useful results - participants giving the normative response on one task may be more likely to give it on another.

In four studies, Stanovich and West (1998) explored the extent to which measures of cognitive ability and thinking dispositions could predict discrepancies from normative responding on a variety of tasks from the deductive literature. The deductive tasks were syllogisms, the selection task, statistical reasoning, and argument evaluation. They measured cognitive capacity by administering well-known cognitive ability and academic aptitude tasks. Stanovich and colleagues suggest that cognitive ability reflects cognitive capacity, specifically that measures of cognitive ability reflect computational limitations. They used a cognitive ability composite score composed of the Scholastic Aptitude Test (SAT), the Raven Matrices, and reading comprehension. All are known to load highly on psychometric *g* (Carpenter, Just & Shell, 1990), and such measures have been linked to neuropsychological and information-processing indicators of efficient cognitive computation (Deary & Stough, 1996; Vernon, 1991, 1993). In Baron's (1985, 1988) conceptualisation, measures of cognitive capacity include perceptual speed, working memory capacity and the efficiency of the retrieval of information stored in long-term memory. These capacities are thought to underlie traditional psychometric intelligence. Thus a link can be seen here between Stanovich and West's (1998) cognitive ability tests, and the working memory tests used in this series of experiments, as measures of cognitive capacity. Stanovich and West (1998) also measured thinking dispositions by administering a questionnaire consisting of intermixed items from a number of subscales (including counterfactual thinking scale and others). This measure was proposed to tap processes such as the motivation to search for counter-examples on the deductive tasks.

Stanovich and West (1998) looked at the relationships between performance on the reasoning tasks, composite scores of cognitive ability and measures of thinking dispositions. Each of the rational thinking (or reasoning) tasks displayed individual differences that

tended to be reliably correlated with the individual differences displayed on other reasoning tasks. It was also found that the cognitive ability composite score was significantly correlated with performance on all four rational thinking tasks. Due to the cross-correlations between the range of deductive tasks and the significant correlations with ability (assumed to reflect capacity), Stanovich and West (1998) suggested that people with higher ability do better overall (when their performance is judged by a normative model). Secondly, it was found that the thinking dispositions composite score was significantly correlated with performance on all four rational thinking tasks. The authors suggested that the reasoner's thinking disposition could explain performance, partly, on the basis of non-normative thought patterns. For instance, one interpretation of thinking dispositions was that the reasoner's performance is affected by their motivation to search for alternative models where the conclusion does not hold. Stanovich and colleagues suggest that jointly, cognitive ability and thinking dispositions account for a moderate amount of variance in overall rational thinking performance, although cognitive ability was a stronger unique predictor.

This account suggested that cross-correlations should be observed between deductive tasks, and that these tasks should also correlate with cognitive ability, assumed to reflect cognitive capacity. There should also be an additional link between reasoning tasks and those suggested to measure a thinking style or disposition. If Stanovich and West's (1998) cognitive ability composite score, assumed to measure capacity, correlated with the deductive tasks, it could be assumed that working memory capacity tasks should also correlate with those reasoning tasks. For instance, Kyllonen and Christal (1990) found that individual differences in reasoning ability reflected differences in working memory capacity. In this way, they suggested that general ability reflected the availability of

attentional resources (Ackerman, 1988). Thus, research has shown a link between reasoning or general fluid ability and working memory capacity.

Stanovich and West (1998) used cognitive ability tests as a measure of capacity. In Experiments 4-6 working memory capacity measures will replace these. In Experiments 1-3 of this thesis, observations were made in agreement with the findings of Stanovich and colleagues, as well as Kyllonen and Christal (1990). There were cross-correlations between syllogistic and spatial reasoning, and both these tasks correlated highly with measures of working memory capacity. Given Stanovich's account, and the results of Experiments 1-3, it would be expected that a wider range of reasoning tasks may correlate together, and with measures of working memory capacity. Experiments 4-6 aim to investigate these findings, using a selection of deductive measures and working memory capacity measures.

The implications of this research for reasoning theory was also examined in Experiments 1-3. The syllogistic reasoning tasks were most associated with the specific verbal and spatial factors, although most consistently with the spatial factor, whilst the five-term series spatial reasoning tasks were most strongly associated with the general factor. This set of findings has implications for verbal and heuristic accounts of reasoning performance. The syllogistic reasoning tasks cross-correlated with span tasks requiring the maintenance and storage of spatial information. The verbal theorists assume that verbal comprehension and rule-based processes can account for performance on syllogistic problems (e.g., Polk & Newell, 1995; Rips, 1994). These theories must be able to explain why syllogistic performance was significantly influenced by the spatial factor in the CFA. The heuristic theorists suggest that much of the data from these types of experiments can be explained by recourse to matching or probabilistic strategies (e.g., Wetherick & Gilhooly, 1990; Chater & Oaksford, 1999).

These theories must be able to explain why the reasoning measures in Experiments 1-3 were predicted by measures of the explicit system (capacity tasks). Experiments 4-6 provide additional reasoning measures that allow the further examination and implications for reasoning theory.

In Experiments 1-3, the factor analysis pointed to a three-factor model of working memory to account for the processing and representation of syllogistic and spatial inferences. All the tasks loaded on the general factor to some extent. What was less clear were the loadings of the reasoning tasks on the specific verbal and spatial factors. The syllogisms, when verbally and visually presented, consistently loaded both the verbal and spatial factors. The five-term series spatial inferences, when verbally presented, only loaded on the general factor. The spatial inferences, when visually presented, loaded negatively on the verbal factor. The complex span measures used in Experiments 1-3 showed relationships with the reasoning tasks in a very consistent way - both complex verbal and spatial spans predicted performance on the reasoning tasks. The simple passive measures used showed variable relationships with the reasoning measures, thus more evidence is needed to establish what these tasks are measuring.

Previous work in this area has looked at simple versus complex tasks and working memory. This allowed the investigation of verbal and spatial resources of working memory, at the level of the subsystems. Farmer, Berman and Fletcher (1986) investigated simple tasks and their relation with working memory through dual task studies. Articulatory suppression disrupted concurrent performance of a verbal reasoning task (A is not followed by B - BA), but had no effect upon performance of a spatial reasoning task (manikin task). In contrast, spatial suppression (continuous sequential tapping) produced reliable interference only with

spatial reasoning. The results implied that the slave systems of working memory are involved in these simple reasoning tasks, and were taken as consistent with Baddeley's notion of two passive subsystems.

Whilst this distinction between verbal and spatial resources at the level of the PL and VSSP is well documented, the role that these subsystems play in more complex reasoning tasks appears to be minimal (e.g., Gilhooly, Logie, Wetherick & Wynn, 1993). Dual-task studies that have used complex reasoning tasks, such as syllogistic tasks, have found strong relationships with CE tasks, and weaker relationships with tasks assumed to load the PL and VSSP. This could also be seen in Experiments 1-3 where the complex deductive tasks showed smaller loadings on factor 1 and factor 2, the domain specific memory factors. Experiments 4-6 will take simple and complex measures of reasoning and examine them together with the working memory capacity measures. Given the dual-task work, it might be expected that the simpler reasoning tasks show stronger relationships with simple working memory spans, assumed to reflect relatively passive storage (as Farmer, Berman & Fletcher, 1986), and that the complex reasoning tasks show stronger relationships with the complex working memory spans, assumed to reflect executive function (as Gilhooly et al., 1993).

Experiment 4 will examine spatial and temporal inferences, together with the working memory capacity measures seen in previous experiments. Both the spatial and temporal problems are complex in nature and would be expected to rely upon resources that reflect both an active (processing) and passive (storage) component. Experiment 5 and 6 in this series use measures which include simple and complex verbal and spatial working memory span measures (Shah & Miyake, 1996; Daneman & Carpenter, 1980; Daneman & Tardif,

1980) together with a range of reasoning measures. The reasoning measures used in Experiment 5 and 6 are simple and complex tasks, which superficially seem verbal or spatial in nature. The factor analysis from Experiments 1-3 produced a model where two of the factors (the verbal and spatial factor) were open to alternative explanations. By using simple and complex reasoning tasks, Experiments 5 and 6 might be able to distinguish whether these factors are passive or active in nature. All the reasoning tasks were presented visually and were evaluative in nature. The tasks include the following - abstract problems (Braine, Reiser & Ruman, 1984), conditional inferences, three-term and five-term series spatial inferences (Byrne & Johnson-Laird, 1989), syllogistic inferences and multiple quantification problems (Johnson-Laird, Byrne & Tabossi, 1989).

Some of the simple and complex reasoning tasks are generally assumed to rely on verbal processing and representation. The complex verbal tasks include the syllogistic task from Experiments 1 and 3, as well as a multiple quantification task (Johnson-Laird, Byrne & Tabossi, 1989). Both these tasks are based on quantifiable statements and require the integration of information between quantifiers in a sentence, either single- or multiply-quantified premises. The simple verbal tasks include an abstract propositional task (Braine, Reiser & Ruman, 1984) and conditional task. Both these tasks rely on reasoning about propositions. The other simple and complex reasoning tasks are generally assumed to rely on spatial processing and representation. The simple spatial task is the three-term series spatial inferences, complimented by the complex spatial task - the five-term series spatial inferences.

Now that the general rationale has been introduced for Chapter 4, the specifics for Experiment 4 will be discussed. Experiment 4 concentrates on two types of reasoning task -

spatial and temporal inference. In the spatial domain, it has been proposed that humans reason from analogical representations, called mental models, that are “representing objects, states of affairs, sequence of events, the way the world is...” (Johnson-Laird, 1983). Byrne and Johnson-Laird (1989) designed their experiments to test a theory of spatial inference based on mental models. Indeed, their results show that it is easier to draw a valid spatial inference when a description corresponds to just a single layout as opposed to multiple layouts.

It has been argued (e.g., Baddeley, 1993) that temporal and causal relationships, in addition to spatial relationships, are also coded spatially in working memory. Schaeken, Johnson-Laird and d’Ydewalle (1996) report five experiments investigating reasoning based on temporal relations. The problems were based on producing a conclusion. They used problems of the following form (Example 4.1), substituting the letters with everyday events.

a before b,

b before c,

d while b,

e while c,

What is the relation between d and e?

Example 4 . 1

The results showed that problems requiring one mental model elicited more correct responses than problems requiring multiple models, which in turn elicited more correct answers than multiple model problems with no valid answers. This can be mapped directly onto the examples given by Byrne and Johnson-Laird (1989) in relation to spatial reasoning, summarised in Chapter 2. Schaeken, Johnson-Laird and d’Ydewalle (1996) experiments

corroborate the mental model theory for temporal reasoning, as did Byrne and Johnson-Laird (1989) for spatial reasoning.

Vandierendonck and De Vooght (1997) report two experiments testing the use of working memory components during reasoning with temporal and spatial relations in four-term series problems. The problems were based on evaluating a set of given conclusions. They used the following form (Example 4.2) for the spatial problems.

The guitar is to the right of the violin,
The guitar is to the left of the drum kit,
The drum kit is to the left of the piano,
Where is the violin in relation to the piano?

Example 4.2

The temporal problems were based on this form also. However, the temporal problems related to actions performed by a person. An example would be, "Charles went to the movies before going to the play". The results of Vandierendonck and De Vooght's (1997) dual-task studies suggested that visuo-spatial resources are used by persons solving linear reasoning problems with a temporal or a spatial content. Articulatory, visuo-spatial, and central executive interference tasks resulted in degraded solution performance in both spatial and temporal cases.

From the review of these studies it might be expected that similar patterns of correlations may be observed for both spatial and temporal reasoning tasks. As yet, no studies tackle the relationship between spatial and temporal reasoning from an individual differences perspective. Thus, Experiment 4 aims to investigate the relationships between spatial and

temporal reasoning problems, and their relationships with working memory span measures. Through a correlational and factor analytic investigation, these results may add to the conclusions made in the first part of this thesis regarding the processing and representation of problems with a spatial content.

Experiment 4 in this series is designed to look at the intercorrelation between spatial and temporal reasoning and their relationship with working memory capacity. Experiment 4 uses measures which include simple and complex verbal and spatial working memory span measures (Shah & Miyake, 1996; Daneman & Carpenter, 1980; Daneman & Tardif, 1987), together with spatial and temporal reasoning measures (Byrne & Johnson-Laird, 1989; Schaeken, Johnson-Laird & d'Ydewalle, 1996). Comparisons have been made between temporal versus spatial reasoning problems, and suggest that they may rely on similar resources and use similar representations.

4 . 2 EXPERIMENT 4

4 . 2 . 1 Aim

The primary aim of Experiment 4 was to build on the assumptions made by Stanovich and West (1998) but using spatial and temporal problems together with capacity measures of working memory. Another aim of Experiment 4 was to further assess the results found in Experiment 2, in relation to spatial reasoning, but with the addition of a temporal reasoning task. The results from Experiment 2 suggested that spatial reasoning relied upon a general resource of working memory. Experiment 4 investigates the extent to which capacity measures of working memory predict performance on both spatial and temporal problems. Four working memory span tasks were used - one complex spatial measure (TTT), one

complex verbal measure (verbal), and a simple spatial (arrow) and verbal (word) measure (Daneman & Tardif, 1987; Shah & Miyake, 1996; Daneman & Carpenter, 1980). The capacity tasks were shown to separate onto different factors in the factor analysis section of Experiments 1-3, identifying a verbal and spatial resource as well as a general one. The span tasks were given together with two reasoning tasks, the spatial inference task used in Experiment 2, and a temporal order task (based on that used by Schaeken, Johnson-Laird and d'Ydewalle, 1996). The reasoning tasks were both presented visually.

4.2.2 Method

Pilot Study

A pilot study was run to ensure that the new Visual Basic temporal inference computer task was easy to understand, use, and at a reasonable difficulty level. The task was found to be suitable for use in the main study.

Participants

Forty-seven participants took part in this study. The participants were undergraduate students and postgraduate students at the University of Plymouth, and they received course credit or cash payment for participating. None of the sample had prior training in logic. They were all native English speakers.

Procedure

Experiment 4 was carried out in one testing sessions per participant, the session being approximately one and a half hours long. Participants signed up in a time slot on an experimenter sheet for the session. The session consisted of six tasks. All participants carried out these computerised tasks in the following order:-

temporal inference visual measure.

complex spatial tic-tac-toe span,

complex verbal sentence span,

simple spatial arrow span,

simple verbal word span,

spatial inference visual measure.

A description of the tasks in Experiment 4, not already explained in previous experiments, is presented next. The procedure for all other tasks was identical to that described previously.

Simple Verbal Word Span (word)

The procedure and scoring was identical to Experiment 1.

Simple Spatial Arrow Span (arrow)

The procedure and scoring was identical to Experiment 1.

Complex Verbal Sentence Span (verbal)

The procedure and scoring was identical to Experiment 1.

Complex Spatial Tic-Tac-Toe Span (TTT)

The procedure and scoring was identical to Experiment 2.

Spatial Inference (spatial)

The procedure and scoring was identical to Experiment 2.

Temporal Inference (temporal)

Procedure

The procedure followed was based on that employed by Schaeken, Johnson-Laird and d'Ydewalle (1996). Each participant received 16 two-dimensional temporal problems in the visual modality. Similar representations have been suggested to underlie both spatial and temporal reasoning, thus the temporal problems were mapped onto the 16 spatial problems (used in Experiments 1 and 2) as consistently as the model structures would allow. For instance - 'left' was substituted for 'before', 'right' for after, and 'behind' and 'in front' became 'while'. Some of the spatial layouts would not translate into temporal ones, due to structural differences between 'behind'/'in front' versus 'while'. Therefore, some indeterminate temporal layouts were introduced. 14 temporal problems had valid conclusions, 2 were indeterminate or invalid. Half of the valid problems, according to model-based theory, had one model answers. Figure 4.1 shows an example of this.

Determinate valid problem

Models

The lever is pressed before the alarm rings.

G L A

The gate opens before the lever is pressed.

B D

The bulb goes on while the gate opens.

The door shuts while the alarm rings.

Hence, what is the relation between the bulb going on and the door shutting?

Figure 4 . 1 - A one model temporal problem

Only one model of the premises can be used here to describe the relation of the events within this set. The correct answer would be that the bulb goes on before the door shuts.

The other half of valid problems, according to model-based theory, had multiple model answers. Figure 4.2 shows an example of this.

<i>Indeterminate valid problem</i>	<i>Models</i>	
Tom drinks coffee before he watches television.	S C T	C S T
He has a shave before he watches television.	B M	B M
He eats a biscuit while he has a shave.		...
He reads a magazine while he watches television.		
Hence, what is the relation between Tom eating a biscuit and Tom reading a magazine?		

Figure 4 . 2 - A multiple model temporal problem

More than one model of the premises can be used here to describe the relation of the events within this set. The correct answer would be that Tom eats a biscuit before he reads a magazine. The order of presentation of problems was randomised before the start of the experiment but remained the same for all participants. A full list of all the problems used can be found in Appendix A2.1. The screen displayed a set of instructions explaining that in this part of the experiment the participants would be asked about the order of events in a given description. A problem would appear on the screen describing the order of events of a set of five activities or actions. All statements for each problem were presented at the same time. Shortly afterwards, a question would also appear on the screen asking them to relate two of the activities or actions within a set. They were told that there would be 16 descriptions using the following terms to describe the order in time in which each activity or action occurs: before, after, and while. The participants were told that their task was to link two of the activities or actions in the set using the above relations. For example, Tom may

eat a biscuit *before* he watches television; Tom may eat a biscuit *after* he watches television; or Tom may eat a biscuit *while* he watches television. Of note here is that some problems had less believable content and participants were reminded that these were hypothetical situations and believability was of no consequence to the conclusions they gave. They were also told that for some of the descriptions they may feel there was not enough information to determine the relation between the activities or actions, in which case they should answer “not enough information”. The instructions continued with a summary on how to respond to the descriptions. The participants were told that the screen would also display a blank box in which they were to type their response to each problem. There was a button to press to move on to the next problem. They were given one practice problem (see Appendix A2.2 for a full set of instructions).

Scoring

The number of temporal problems correctly solved (out of a maximum of 16) was taken as a participant’s score for the temporal reasoning task. This figure was then used in the main analysis. In addition, it was noted how many of each type of problem (one-model, two-model, or indeterminate) was solved. This was used in further analysis to test the assumptions made by the mental models theory of reasoning.

4 . 2 . 3 Results

In addition to correlational analyses and t-tests, confirmatory factor analysis was also performed on the data from this experiment. The correlational and significance analysis will be reported in this experimental section, but the factor analysis for the three experiments (Experiments 4-6) will be discussed in a factor analytic section (Section 4.5) after Experiment 6. This is due to a multi-group analysis consisting of Experiments 4 through to 6.

As discussed previously, there are multiple scoring methods for many of the span tasks. However, only one method per task will be used throughout the results section. The cross-correlations of the global and normal span scores (as defined in Section 3.2.2) across different measures were between .91*** and .96***. This suggests that both scoring methods reflected the same underlying construct. Thus, for arrow, word and verbal spans, the global score will be used throughout.

4.2.3.1 Descriptive Statistics

Descriptive statistics for the six measures used in Experiment 4 are presented in Table 4.1.

The table also contains data relevant to the reliability of the measures used.

	Mean	SD	Min	Max	Skewness	Reliability
<i>Span Tasks</i>						
arrow	15.17	9.40	1	37	.83	.81
word	46.53	14.83	19	96	.74	.85
verbal	36.26	18.73	6	78	.35	.91
TTT	6.94	6.23	0	25	1.22	.80
<i>Reasoning Tasks</i>						
temptot	9.51	3.74	1	15	-.43	.81
temp1M	4.66	1.77	0	7	-.48	-
tempMM	4.51	2.00	0	7	-.53	-
tempIND	.32	.52	0	2	1.29	-
spattot	8.91	3.37	2	15	-.40	.70
spat1M	4.47	2.01	0	7	-.53	-
spatMM	4.45	1.93	1	8	-.07	-

Table 4.1 - Descriptive statistics for Experiment 4 tasks (n=47)

Table 4.1 shows that all tasks produced a good overall range of scores. The spatial and temporal tasks have been broken down into problem type - one, multiple and indeterminate problems. The one and multiple model types show similar performance rates, whilst the indeterminate temporal problems show a lower accurate response rate. Both the span and reasoning measures show a varying level of performance. The distribution of scores in all tasks was satisfactory, there were no heavily skewed distributions for any of the measures used.

The reliability of the span tasks was again assessed to ensure they were internally reliable. Each span task was inspected for each participant to ensure that each set in each span task was answered equally often. Their reliability was assessed by performing a split-half calculation on each of the four span measures used. The Spearman-Brown formula was used to provide an estimate of reliability for the full task. The corrected correlations for the span tasks are shown in Table 4.1. Table 4.1 shows that the arrow, word, verbal and TTT span tasks produced a reliable odd-even corrected correlation of over .80. All span measures showed a highly satisfactory level of reliability (Rust & Golombok, 1999).

Each reasoning measure was also assessed for internal reliability using Cronbach's alpha. As seen in Table 4.1, the reliability estimate for the spatial reasoning task was adequate for group measurement ($r = .70$). This can be compared to Experiment 2 where the spatial task, presented visually, also showed an adequate level of reliability (where $r = .75$). The reliability estimate for the temporal reasoning task was highly satisfactory for group measurement ($r = .81$).

4.2.3.2 The Span Measures

The first analysis examined whether similar patterns of cross-correlations could be observed between the span measures (see Appendix B1.10 for the full matrix). This would further substantiate the claim for a general CE resource, as opposed to separate CE resources for verbal and spatial information (e.g., Turner & Engle, 1989; Shah & Miyake, 1996). The correlation matrix for Experiment 4 span measures is presented in Table 4.2.

	arrow	word	verbal
arrow	1		
word	.38**	1	
verbal	.45**	.71***	1
TTT	.38**	.18	.24 ⁺

(two-tailed sig. levels: ⁺ = $p < 0.1$, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$)

Table 4.2 - Experiment 4 correlations between the four span tasks

As Table 4.2 shows, a positive matrix can be seen between the span measures used. Firstly, the highest correlation is between the simple word and complex verbal spans ($r = .71$, $p < .001$). This is entirely consistent with the experiments so far. The passive word span correlates with the passive arrow span ($r = .38$, $p < .01$), which was observed in Experiments 2 and 3 ($r = .28$ and $r = .26$, $p < .10$), and less strongly in Experiment 1 ($r = .23$). This does not fit the picture of two separate subsystems of working memory. However, the correlation could still arise if both tasks draw on the general resource. Experiments 1-3 multiple CFA three-factor model showed that both the simple span measures, word and arrow, loaded the general factor as well as the specific factors. Thus, some of the correlation between the simple measures could be explained if they measure domain independent parts of the CE. The arrow span also correlates significantly with the two complex span measures, the verbal and TTT spans ($r = .38$, $r = .45$, $p < .01$). The arrow span would be expected to correlate

with the TTT span due to them both relying on the manipulation of spatial information. It is not so clear as to why the arrow span correlates with the complex verbal span. However, again as indicated in Experiments 1-3 CFA, it might be that all the span measures tap some general processes. This correlation between the arrow and verbal spans was also found in Experiment 3 ($r = .31, p < .05$). The correlation between the two complex spans, TTT and verbal, reached a borderline significant correlation ($r = .24, p < .10$). Overall, the findings fail to replicate the dissociation found by Shah and Miyake (1996).

4.2.3.3 The Span and Reasoning Measures

The next set of analyses relates to the influence of the span measures on the temporal and spatial reasoning tasks. Similar relationships between both types of reasoning task and working memory spans would be expected. Past literature shows that the solution of linear reasoning problems with a spatial or a temporal content require visuo-spatial resources of working memory (e.g., Vandierendonck & De Vooght, 1997). The relevant correlations from Experiment 4 are presented below in Table 4.3.

	arrow	word	verbal	TTT
temporal	.40**	.32*	.31*	.52***
spatial	.47***	.31*	.27 ⁺	.41**
temporal - spatial correlation	.68***			

(two-tailed sig. levels: ⁺ = $p < 0.1$, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$)

Table 4.3 - Experiment 4 correlations between the span and reasoning tasks

Table 4.3 presents data relevant to whether individual differences in working memory predict individual differences in both temporal and spatial reasoning in similar ways. For both tasks the matrix is generally positive in relation to the span measures used. The pattern

of correlations for both the spatial and temporal task are very similar. The data shows that the best predictors of both reasoning tasks are the spatial spans - simple arrow and complex TTT ($r = .40$, $r = .52$ for the temporal task; $r = .47$, $r = .41$ for the spatial task). This result is entirely consistent with that found in Experiment 2 for the spatial reasoning task (for arrow, $r = .25$; for TTT, $r = .41$). The spatial and temporal task would be expected to correlate highest with the spatial spans due to the spatial nature of the measures. These results are also consistent with research that indicates the similar nature of processing and representation in temporal and spatial problems. However, there is also an indication that the reasoning tasks correlate marginally with the verbal spans. This is also consistent with Experiment 2 results.

The spatial and temporal problems appear to draw on processes that are common with spatial working memory capacity, but there is also some evidence of a relationship with verbal working memory capacity. The last finding to further strengthen the view that temporal and spatial problems may rely on similar resources is the correlation between them, a very strong relationship ($r = .68$, $p < .001$). This lends support to the idea that temporal and spatial reasoning rely on similar representations and processes.

4.2.3.4 Further Analysis

The spatial and temporal tasks were also divided into easy and difficult problems. Past research suggests one model problems are easier to solve than multiple model problems. Problem difficulty percentages for the temporal and spatial tasks are presented in Table 4.4.

	% correct
temp1M	67
tempMM	64
tempIND	15
spat1M	56
spatMM	55

Table 4 . 4 - Experiment 4 % of correct responses to 1M, MM and IND problems

An average of 10 out of 16 (59%) visually presented temporal inferences were solved, 5 of which were single model problems. It should be noted that there were only 2 indeterminate temporal problems, therefore this data may be noisy. An average of 9 out of 16 (56%) visually presented spatial inferences were solved, 5 of which were single model problems. This is a lower overall accuracy rate for the spatial problems than in Experiment 2, where 69% were solved correctly.

A number of t-tests were also performed on the data (see Appendix B1.2 for full results). These would identify whether there were significant differences between the single model and multiple model problems, and whether there were differences between the reasoning tasks used. Overall, in both the temporal and spatial inference task, participants performed at similar rates (d.f. = 46, $t = 1.43$, $p = .16$). This might reflect the fact that both tasks rely on similar processing and representations. Participants performed significantly better on both the temp1M and tempMM problems in contrast to the tempIND problems (d.f. = 46, $t = 17.94$, $p = .001$; d.f. = 46, $t = 14.721$, $p = .001$). This is in line with work that showed that more difficult indeterminate problems are harder to solve (Schaeken, Johnson-Laird & d'Ydewalle, 1996). There were no significant differences between the temp1M and

tempMM problems (d.f. = 46, $t = .81$, $p = .42$). Similarly, participants performed equally well on the spat1M versus spatMM problems (d.f. = 46, $t = .07$, $p = .94$). This finding is inconsistent with Experiment 2 results and previous studies which show an effect of the number of models to be constructed. However, in Experiment 2, it was shown that visual presentation reduced the difficulty between one model and multiple model problems. Perhaps in this case, the reduction in memory load through visual presentation eliminated this difference in difficulty.

4.2.4 Discussion

The results from Experiment 4 suggest that individual differences in working memory capacity can explain variation in performance on five-term series spatial and temporal tasks, when visually presented. Experiment 4 indicates that temporal and spatial inference tasks correlate positively, and mostly significantly, with complex measures of both verbal and spatial working memory. The complex span tasks were differentiated by the type of information being stored and processed - the verbal span required the storage and processing of words and sentences, whilst the spatial TTT span required the storage and processing of winning lines in a three-dimensional game. There was no clear separation of central resources for temporal and spatial reasoning dependent on the nature of these span tasks. Both the verbal and spatial spans predicted individual differences on the reasoning tasks. A similar pattern of correlations was observed for both the temporal and spatial measures, when visually presented, although there was no significant correlation between the spatial task and the complex verbal working memory measure.

Past studies have shown that tasks that interfere with all three components of working memory resources (i.e., articulatory, visuo-spatial, and central executive suppression) affect

performance on both types of reasoning task - spatial and temporal (e.g., Vandierendonck & De Vooght, 1997). Vandierendonck et al. (1997) claim that although there has been no evidence that indicates that the representation constructed is in the visuo-spatial mode (Evans, Newstead & Byrne, 1993), their experiment is consistent with both the older spatial array view (e.g., De Soto, London & Handel, 1965) and with the more recent mental models theory (Johnson-Laird, 1983). The results of Experiment 4 are in line with the finding from Vandierendonck et al.'s (1997) work, from an individual differences perspective. It was found that both types of reasoning problem relate to all resources of working memory capacity tasks, and that a difference could be seen in performance on simple single model and difficult indeterminate temporal problems - a finding consistent with Johnson-Laird's (1983) account. These results suggest that both reasoning tasks draw on verbal and spatial resources of working memory, although the correlation with the spatial spans is highest.

A possible explanation of the relationship between span and reasoning measures is that the spatial and temporal inferences rely on both verbal and spatial resources which enables particular aspects of the deductive process to occur. It may be that performance on both types of problems first requires the interpretation of the linguistic information given in each premise. The role of verbal comprehension processes may explain the relationships between both temporal and spatial inferences with the simple and complex verbal working memory measures. The verbal information may then be processed by the manipulation of some form of mental model. This link with the manipulation of a mental model may explain the relationships between both temporal and spatial inferences with the simple and complex spatial working memory measures.

The findings for the spatial and temporal task have implications for theories of reasoning. Both tasks showed strongest correlations with spatial working memory, but were also correlated with verbal working memory. Interpretations of this finding, in relation to the three classes of reasoning theory (Chapter 2), will be left until the general discussion (Section 4.6). In this way, the factor analysis in Section 3.5 can be used to further identify which resources each reasoning task utilises.

Further investigation was needed to identify the function of the three factors identified in Experiments 1-3 CFA. Therefore, the next two experiments examine a wider range of reasoning tasks, some that are verbal or propositional in nature whilst others are spatial in nature. The wider range will allow the investigation into underlying resources required in performing both simple and complex reasoning measures. The tasks used so far - syllogistic, spatial and temporal tasks - appear to consistently relate to the complex span measures. In general, the reasoning tasks do seem to rely on centrally allocated resources of working memory, as defined by the general factor in Experiments 1-3. What has been less clear is how these tasks utilise the storage systems of working memory (or domain-specific parts of the CE), that may be best reflected by the simple span measures. Thus, in Experiment 5, only the simple word and arrow spans were used in order to try to identify the verbal (F1) and spatial factor (F2) from Experiments 1-3 CFA more fully. Thus, Experiment 5 investigated the interrelationships between the simple span measures and performance on a set of reasoning tasks. These included tasks assumed to rely on verbal processing and storage - abstract and conditional tasks. Also included were tasks assumed to rely on spatial processing and storage - three-term and five-term series spatial reasoning. Other complex tasks were also investigated, and as yet there are uncertain findings as to how these tasks are processed and stored - syllogistic and multiple quantification tasks.

4.3 EXPERIMENT 5

4.3.1 Aim

A range of simple and complex reasoning tasks is introduced in Experiment 5. The first set of experiments in this series identified a working memory model that consisted of three factors - with a verbal, spatial and general factor. The reasoning tasks that were used in these experiments were all complex in nature - they were all assumed to require the maintenance and storage of information, and generally loaded heavily on the general factor. Due to this general factor that emerged throughout the first three experiments, it was decided that Experiment 5 would concentrate solely on identifying what the specific factors were measuring. It might be that factor 1 and factor 2 represent the PL and VSSP of working memory. However, it might also be the case that these two factors represent domain specific parts of the CE. Thus, Experiment 5 attempted to adjudicate between these two alternatives by using additional reasoning tasks.

A finding from the first three experiments was that the arrow and word span showed the lowest correlations, perhaps suggesting a dissociation between them. In the CFA, the word span loaded factor 1 and factor 3, in contrast, the arrow span loaded factor 2 and factor 3. These loadings were significant and consistent across the three experiments. Thus, it could be that the simple spans load factor 1 and factor 2 due to their passive storage element, but also both load factor 3 due to the commonality in processing. By using the two span tasks that showed the lowest association, it may be possible to identify whether factor 1 and factor 2 are passive or active resources.

The five-term series spatial inferences from the previous experiments loaded heavily on the general factor, although the visually presented problems had the additional negative relationship with the verbal factor. The syllogistic inferences, whether visually or verbally presented, seemed to load all three of the factors, including the spatial one. Experiment 5 will further test the best way to interpret factors 1 and 2, the verbal and spatial resource, by using the word and arrow span to identify their passive or active nature.

Experiment 5 investigates simple and complex reasoning tasks. A reason for this is due to past dual-task research, although the tasks are not identical. For instance, Farmer, Berman and Fletcher (1986) used a simple verbal or propositional task and a simple spatial task, together with secondary tasks assumed to load the sub-systems of working memory, the PL and VSSP. They found that a detriment in performance on both the simple reasoning tasks only occurred with same modality secondary tasks, for instance, articulatory suppression or spatial tracking. This, they suggested is evidence for a distinction between verbal and spatial resources at the level of the sub-systems. Thus the reasoning tasks to be used in Experiment 5 may also be differentiated at this level. An expectation might be that simple reasoning tasks will show higher correlations with simple span measures, if the span measures reflect passive storage. Three simpler reasoning tasks, that are assumed to rely mainly on passive resources, with some processing requirement, will be used in Experiment 5. Two of these are based on propositional logic (abstract and conditional tasks), and one is based on three-term series spatial descriptions (simple spatial task). Although dual-task work has identified conditionals as being affected by CE secondary tasks (Toms, Morris & Ward, 1993), they were chosen to be labelled together with the other propositional and simple spatial task in contrast to the quantifiable and complex spatial tasks described next. Other dual-task work has investigated more complex reasoning tasks, such as syllogistic and

five-term series inferences (e.g., Gilhooly, Logie, Wetherick & Wynn, 1993; Klauer, Stegmaier & Meiser, 1997). A detriment in performance on these reasoning tasks was most seen when carried out with secondary tasks assumed to load the CE component of working memory. Thus, three of these more complex reasoning tasks, that may rely on executive processing as well as storage, were chosen to be used in Experiment 5. Two of these are based on syllogisms (syllogistic and multiple quantification tasks), and one is based on five-term spatial descriptions (complex spatial task).

The role that factor 1 and factor 2 play in performance on the six reasoning tasks might be better identified by using a mixture of simple and complex reasoning measures. The factors may be easier to interpret by the addition of propositional and simple spatial tasks (labelled the simple reasoning measures). If factor 1 and factor 2 do represent the PL and VSSP, it might be expected that the simpler reasoning tasks load more on the verbal and spatial factor, whilst the complex reasoning tasks show weaker relationships with these two factors.

This experiment also allows a test of the proposals put forward by Stanovich and colleagues. Stanovich and West (1998) used syllogistic and statistical reasoning tasks, together with a version of the selection task and an argument evaluation test. These reasoning tasks were found to be significantly correlated and the authors suggested that it was systematic response tendencies, not random performance errors, that could account for individual differences across very different reasoning tasks. If Stanovich and West (1998) were correct in the explanation of the correlations between their reasoning tasks, they would expect positive correlations between all the reasoning tasks used in Experiment 5.

Stanovich and West (1998) also implemented tests of cognitive ability and used them as measures of cognitive capacity (e.g., SATs). Stanovich and colleagues would expect the

reasoning tasks used here to correlate with measures of capacity (in comparison to their ability measures).

As previously mentioned, six reasoning tasks will be used in Experiment 5. The simple reasoning tasks consist of three-term series spatial inferences, abstract problems (Braine, Reiser & Rumin, 1984) and conditional inferences. The three-term series problems require reasoning about relations. The abstract and conditional problems require reasoning about propositions. The complex reasoning tasks consist of two tasks already used in prior experiments, five-term series spatial and syllogistic inferences. An additional complex task is multiply quantified inferences (Johnson-Laird, Byrne & Tabossi, 1989), similar to syllogistic reasoning but with double the quantifiers in each premise.

Two alternative explanations were put forward for the three-factor model, where factor 1 and factor 2 could be the sub-systems of working memory, or that they might be domain specific components of the CE. Therefore, the two span measures that will be used in Experiment 5 are those tasks that were found to be most associated with those two factors, the word span and the arrow span (Shah & Miyake, 1996). The span tasks were given together with the six reasoning tasks. The reasoning tasks were presented visually and in order to maintain consistency across all the reasoning tasks, an evaluative paradigm was used. This involved the evaluation of conclusions, rather than their production. The tasks also differed from those in the experiments so far due to the use of abstract content throughout (Experiments 1-4 used thematic/arbitrary content), which also helped to maintain consistency.

4 . 3 . 2 Method

Pilot Study

A pilot study was run to ensure that the new paper and pencil reasoning tasks (all six) were easy to understand, use, and at a reasonable difficulty level. A number of participants carried out the tasks and were happy with the procedure and type of problem. The abstract reasoning and three-term series tasks were found to be too easy when not timed (ceiling effect). Therefore a time limit was introduced of 15 and 10 seconds per problem, respectively, in order to overcome this effect. No other changes were made to the initial design.

Participants

Fifty-one participants took part in this study. The participants were undergraduate students and postgraduate students at the University of Plymouth, and they received course credit or cash payment for participating. None of the sample had prior training in logic. All participants were native English speakers.

Procedure

Experiment 5 was carried out in one testing session per participant, the session being approximately one and a half hours long. Participants signed up in a time slot on an experimenter sheet for the session, and were run in groups. The session consisted of eight tasks. All participants carried out these tasks in the following order (where an asterix (*) indicates a computerised task):-

simple spatial inference,

abstract reasoning,

simple arrow span*.

simple word span*,

conditional inference,

sylogistic reasoning.

complex spatial inference,

multiple quantification.

A description of the tasks in Experiment 5 not already explained in previous experiments. are presented next.

Simple Verbal Word Span (word)

The procedure and scoring was identical to Experiment 1.

Simple Spatial Arrow Span (arrow)

The procedure and scoring was identical to Experiment 1.

The Reasoning Tasks

All reasoning problems were presented as paper and pencil tasks. They were all, to a greater or lesser degree, evaluation tasks. In Experiments 1-4 the participants were asked to produce a conclusion to the problems given. In Experiment 5, participants were given a choice of conclusions to pick from. This was in order to provide consistency across all the reasoning tasks used. The reasoning tasks all contained abstract content (non-thematic). They were all based on hypothetical relationships between letters and numbers. In this respect, the reasoning tasks differed from those used in previous experiments, which required the production of a conclusion and employed thematic content.

Abstract Reasoning Task (abstract)

Procedure

9 multi-step problems, rated at 5.00 and above on the scale presented by Braine, Reiser & Romain (1984) were used. This task was concerned with people's ability to reason logically with sentences in various forms. They were presented with a total of 9 problems in the booklet. For each of the problems they were given a set of statements. An example of those statements is shown in Figure 4.3.

Abstract problem

If there is both an N and an I, then there's not a B

It is false that there's not an N

There is an I

Response options

There's a B

There's not a B

There may or may not be a B

Figure 4 . 3 - An abstract problem

Each problem concerned the presence or absence of letters on an imaginary black-board.

The set of statements contained some facts about the black-board. The facts or premises contained the information that was known about the black-board and they should be accepted as true. A full list of all the problems used can be found in Appendix A3.1.

The participant's task in each case was to decide on whether a presented conclusion necessarily followed from the statements about the black-board. A conclusion was necessary if it must be true, given that the statements were true. A list of the possible answers was presented below each problem. There were three response options per problem. They were to tick one of the response conclusions for each set of statements (see Figure 4.3). The correct response to this problem would be that there is not a B. If they

thought that no conclusion necessarily followed given that the statements were true, then they were to tick the relevant box. for instance, that the letter B may or may not be there. Participants were instructed to move onto the next problem after 15 seconds. A final requirement for participants was that no diagrams or notes were to be made during the task. Any questions about the task were asked before the start. There was one practice problem. A set of full instructions may be found in Appendix A3.2.

Scoring

The number of abstract problems correctly solved (out of a maximum of 9) was taken as a participants score for the abstract reasoning task. This figure was then used in the main analysis.

Conditional Inferences (condition)

Procedure

Participants received a total of 16 problems - 4 modus ponens, 4 modus tonens, 4 affirmation of the consequent, 4 denial of the antecedent. Following is Table 4.5, taken from Evans, Newstead and Byrne (1993), which presents the logical structure of the problems used.

Inference	Abbreviation	First Premise	Second Premise	Conclusion
Modus Ponens	MP	If p then q	p	q
Denial of the Antecedent	DA	If p then q	not-p	not-q
Affirmation of the Consequent	AC	If p then q	q	p
Modus Tollens	MT	If p then q	not-q	not-p

Table 4 . 5 - The logical structure of conditional inferences

The MP and MT inferences are logically valid, whilst the DA and AC inferences are logically fallacious in nature. Participants were instructed that this task concerned people's ability to reason logically with sentences in various forms. They were presented with a total of 16 problems in the booklet. All four forms of conditional were used by systematically manipulating negation within the first conditional premise. A full list of all the problems used can be found in Appendix A3.3. For each of the problems they were given two statements as shown in Figure 4.4.

<i>Conditional problem (MP)</i>	<i>Response options</i>
If the letter is M then the number is not 3	The number is a 3
The letter is M	The number is not 3
	The number may or may not be a 3

Figure 4 . 4 - A conditional problem

They were told that each problem concerned an imaginary letter-number pair and contained an initial statement or rule which determined which letters could be paired with which numbers. In each case, they were to assume that the rule held and then combine it with the information given in the second statement. This would concern either the letter or the number of an imaginary pair. Their task in each case was to decide on a conclusion that necessarily followed from the statements. A conclusion was necessary if it must be true, given that the statements were true. A list of the possible answers was presented below each problem. They were to tick one of the response conclusions for each pair of statements. There were three response options for each problem (see Figure 4.4). If they thought that no conclusion necessarily followed given that the statements were true, then they were to tick

the relevant box, for instance, that the number may or may not be a 3. In the above example, Figure 4.4, the correct response would be that the number is not a 3. Participants were also instructed that no diagrams or notes were to be made during the task. A full set of instructions may be found in Appendix A3.4.

Scoring

The number of conditional problems correctly solved (out of a maximum of 16) was taken as a participants score for the conditional reasoning task. The logically correct answers were used, i.e., MP and MT were taken as logical, whereas DA and AC were assigned a point if an indeterminate conclusion or ‘no conclusion necessarily follows’ was chosen. This figure was then used in the main analysis.

Simple Spatial Inference (simspat)

Procedure

Each participant received 16 one- and two-dimensional spatial problems. 1-D problems relate to three letters in a set, 2-D problems relate to five letters in a set, as previously explained. The order of presentation of problem in the tasks was randomised before the start of the experiment but remained in this order for each participant. For the three-term task, spatial problems from six possible orders were used, as shown in Figure 4.5.

A B C A C B B A C B C A C A B C B A

Figure 4.5 - The six possible orders for three-term series problems

Eight problems had valid conclusions, and eight had invalid conclusions. A full list of the problems used can be found in Appendix A3.5. Figure 4.6 shows an example of a valid problem.

<i>Valid three-term series problem</i>	<i>Response options</i>
A is left of B	A is left of C
B is left of C	A is right of C
	invalid

Figure 4 . 6 - A valid three-term series problem

Participants were given three response options for each problem (see Figure 4.6). The correct response to this problem would be that A is left of C. The instructions for the three-term task were similar to the five-term series task. On each page of a booklet was a problem which would describe the layout of a set of either three or five letters. In each task the participants were told that there would be 16 descriptions using the following terms to describe the layout of the letters in the three-term task: left and right. The participants were told that their task was to link two of the letters in the set using the above relations. They were told to imagine the three letters as if they were in front of them and then apply these relations to them. They were also told that for some of the descriptions they may feel there was not enough information to determine the relation between the letters, in which case they should answer “invalid” for the three-term problems. The instructions continued with a summary on how to respond to the descriptions. To remind them of the responses they might give, a list of all the possible answers was given below each problem. They were told to tick the answer that they thought was correct. Participants were instructed to move onto

the next problem after 10 seconds. A full set of instructions can be found in Appendix A3.6.

Scoring

The number of spatial problems correctly solved (out of a maximum of 16) was taken as a participant's score for the simple spatial reasoning task. This figure was then used in the main analysis. In addition it was noted how many of each type of problem were solved, be it one-model or invalid.

Complex Spatial Inference (comspat)

The procedure and scoring was identical to Experiment 2.

Syllogistic Inference (syllogs)

The procedure and scoring was identical to Experiment 3.

Multiple Quantification Task (mulquan)

Procedure

As in the syllogistic task, participants were asked to solve 16 problems, based on the logic of syllogisms. A syllogism was explained as a pair of statements providing them with information about the relationships between three classes. The three classes were labelled by letters (e.g. Bs, Ps and Cs etc.). On the basis of the information provided by the two statements, they were asked to draw a conclusion about the relationship between two of the classes. One of the letters was common to both statements; their task was to produce a conclusion linking the other two letters. The participants were presented with a total of 16 problems in each booklet, on separate pages. A full set of problems may be found in Appendix A3.7. Participants were instructed to find a logically valid conclusion that linked

the classes not explicitly related in the two premises. A conclusion was logically valid if it must follow given that the statements were true. If a conclusion *could* but did not *have* to or necessarily follow, then it was not logically valid. For each of these problems they were given two statements. An example is shown in Figure 4.7.

Multiple quantification problem

Some of the Ps are in the same place as some of the Bs

Some of the Bs are in the same place as some of the Fs

Response options

___ of the Ps are in the same place as ___ of the Fs

OR

No Valid response

Figure 4 . 7 - A multiple quantified problem

The participants had the following quantifiable options for the conclusion - All, Some, None, and Any. They also had the option of ticking 'no valid response'. They were given a response conclusion with two gaps indicating where they must fill in a response. They were told to write the answer that they believed was the correct, and strongest conclusion to each problem (i.e. If the answer could have been some... *or* all..., they were to tick the stronger conclusion, all...). Figure 4.7 shows the response options to the problem. Their task was to fill in the two gaps with the correct quantifiers (all, some, none or any), for instance for the example above they might fill in the gaps with - some of the Ps are in the same place as some of the Fs. If they thought that no one conclusion logically followed, then they were to tick no valid conclusion. Participants received the following problems - 10 one-model, 4 multi-model indeterminate, and 2 multi-model determinate problems. They were asked to

take their time and be certain they had the logically correct answer before stating it. If they had any questions, they were asked to please ask them now, as the experimenter could not answer after the task had commenced. They could not make notes or draw diagrams of any kind to aid them in these tasks. A practice problem was given to show them the format of each of the 16 problems in each task. A full set of instructions may be found in Appendix A3.8.

Scoring

The number of multiple quantification problems correctly solved (out of a maximum of 16) was taken as a participant's score for this reasoning task. This figure was then used in the main analysis. In addition, it was noted how many of each type of problem (one-model or multi-model) were solved. This was used in further analysis to test the assumptions made by the mental models theory of reasoning.

4 . 3 . 3 Results

In addition to correlational analyses and t-tests, confirmatory factor analysis was also performed on the data from this experiment. The correlational and statistical for Experiment 5 will be reported in this experimental section, but a multi-group factor analysis for the three experiments (Experiments 4-6) will be discussed after Experiment 6.

As discussed in relation to previous experiments there are multiple scoring methods for many of the span tasks. However, only one method per task will be used throughout the results section. The cross-correlations of the global and normal span scores across different measures were between .85*** and .93***. This suggests that both scoring methods reflected the same underlying construct. Thus, for simple arrow, simple word and complex verbal spans, the global score will be used throughout.

4.3.3.1 Descriptive Statistics

Descriptive statistics for the eight measures used in Experiment 5 may be found in Table 4.6. The table also contains data relevant to the reliability of the measures used.

	Mean	SD	Min	Max	Skewness	Reliability
<i>Span Tasks</i>						
arrow	23.24	12.93	2	60	.75	.64
word	48.33	15.08	19	87	.45	.88
<i>Reasoning Tasks</i>						
abstract	5.90	1.59	2	9	-.11	.34
condition	8.75	1.80	5	14	.53	.30*
simspat	7.78	2.99	3	15	.56	.67
simspat1M	5.27	1.58	2	8	-.28	-
simspatINV	2.51	2.20	0	7	.75	-
comspat	9.76	3.44	1	16	-.58	.76
comspat1M	5.71	2.11	1	8	-.85	-
comspatMM	4.06	2.04	0	8	-.05	-
syllogism	8.37	3.41	0	16	.65	.84
syll1M	6.61	1.65	0	8	-1.92	-
syllMM	1.76	2.67	0	8	1.31	-
mulquan	7.71	1.95	3	12	-.16	.28
mulquan1M	6.10	1.33	3	8	-.40	-
mulquanMM	1.61	1.27	0	5	.92	-

* = Cronbach alpha could not be calculated with all MPs problems. Estimations can be made by dropping problems 2 and 4 (see Appendix 3.3A).

Table 4.6 - Descriptive statistics for Experiment 5 tasks (n=51)

Table 4.6 shows that all tasks, both span and reasoning measures, produced a good overall range of scores. Of note here is that the range of scores is less varied than in the first three experiments, probably due to the evaluative rather than productive nature of the reasoning

tasks set. The distribution of scores in all tasks is acceptable, there are no heavily skewed distributions for any of the measures used.

The reliability of the two span tasks was assessed by performing a split-half test calculation on each of them. The Spearman-Brown formula was then used to provide an estimate of reliability for each task. The corrected correlations for the span tasks are shown in Table 4.6. Table 4.6 shows that the word span task produced a reliable odd-even corrected correlation of over .80, of highly satisfactory value in group measurement. This is consistent with the estimates found for the word span in Experiments 1-4. The arrow span did not produce as high reliability as this, at a corrected .64, although this is still of some value in group measurement (Rust & Golombok, 1999). This is the lowest estimate of the arrow span so far in this experimental series.

The reliability of the reasoning measures was assessed using Cronbach's alpha. Table 4.6 shows that the simple and complex spatial inferences, together with the syllogistic task, were all satisfactorily reliable (between .67 and .84). The abstract task exhibited a poor reliability of .34, as did the conditional task (.30) and the multiple quantification task (.28). This will be taken into account when looking at the correlations, to be discussed next.

4.3.3.2 The Span Measures

Due to the fact that the complex working memory span measures were not used here, the first set of correlations relates to the relationship between the arrow and word spans and then to their influence on the reasoning tasks (see Appendix B1.11 for the full matrix). In the prior experiments it has been shown that these two measures act relatively independently of each other, and two interpretations have been considered. The first is that these two span

measures tap into the passive storage systems of working memory for verbal and spatial information. The second is that these two span measures tap into domain specific parts of the CE component of working memory. The results can be found in Table 4.7.

	arrow
arrow	1
word	.17

(two-tailed sig. levels: + = $p < 0.1$, * = $p < 0.05$,
** = $p < 0.01$, *** = $p < 0.001$)

Table 4.7 - Experiment 5 correlations between the two span tasks

As Table 4.7 shows, a dissociation can be seen between the word span and arrow span, the lowest correlation to be found to date between these two measures (Expt.1 to Expt.4 correlations ranged between $r = .23$ and $r = .38$). However, the arrow span in Experiment 5 produced the lowest estimate of reliability seen so far (at $r = .64$). The procedure and sampling for Experiment 5 was identical to Experiments 1-4, and the standard deviation for this task was constant throughout the series. Therefore, the lower estimate could be down to chance variation. So the conclusion coming from this study is that the two tasks are largely independent of each other in what they measure. However, the overall finding that is more consistent throughout the experiments is that both measures possess some commonalities due to the positive, and usually significant, nature of the correlation. In addition, both these tasks loaded the general factor in the CFA. These two measures could be interpreted as tapping the relatively autonomous PL and VSSP sub-systems, but they may also have a CE component, and this would explain the marginal correlations between them.

4.3.3.3 The Span and Reasoning Measures

	arrow	word
abstract	.26 ⁺	.32*
condition	.19	.21
simspat	.33*	.39**
comspat	.27 ⁺	.08
syllogism	.20	.04
mulquan	.32*	.23

(two-tailed sig. levels: ⁺ = $p < 0.1$, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$)

Table 4.8 - Experiment 5 correlations between the span and reasoning tasks

Correlations between the two span measures and each of the reasoning measures used can be found in Table 4.8. Please note that the complex spatial task, named to be able to distinguish it from the simple spatial task, is the same as the five-term series visually presented task seen in Experiments 2 and 4 (but based on evaluation not production).

The five-term series spatial inferences (complex spatial) show a marginally significant correlation with the arrow span (.27, $p < .10$). They do not correlate significantly with the word span ($r = .08$). In previous experiments the correlations with the arrow and word spans were, on average, higher than those seen here, but the positive direction of the relationships are consistent through Experiments 2, 4, and 5. The syllogistic inferences show similar relationships as the five-term series spatial problems, although the correlation with the arrow span is smaller ($r = .20$). In past experiments these correlations were also larger and word span was significantly correlated with the syllogistic task in Experiment 3 ($r = .44$ and $r = .57$ for the syllogistic task in both modalities). An explanation for the low correlations with these two tasks in Experiment 5 may be because the tasks were evaluative

in nature. The tasks may not place the same load on working memory due to the conclusions being in front of the participant at all times.

The other complex task used here was the multiple quantification task. Again, low correlations can be seen between it and the word span ($r = .23$), although there is a moderate correlation with the arrow span ($r = .32, p < .05$). These correlations need to be interpreted in light of the low reliability of the multiple quantification task ($r = .28$). This might explain the weak correlations. This said, there are certainly stronger relationships between the complex reasoning tasks and the arrow span (rather than the word span) perhaps indicating that spatial resources are needed in order to solve them. This finding would be consistent with previous experiments that showed a connection between these complex deductive measures and spatial working memory.

Next, the simple reasoning tasks will be discussed. The abstract task correlates marginally with both the span measures ($r = .26, p < .10$ for arrow; $r = .32, p < .05$ for word). The conditional task also correlated positively, though not significantly, with both the span tasks. The low correlations seen between the spans and the conditional task might be due to the low reliability of this reasoning task ($r = .30$). The simple spatial task correlated both positively and significantly with both the span tasks ($r = .33, p < .05$ for arrow; $r = .39, p < .01$ for word). The findings for the simple reasoning tasks show that they correlate similarly with both the arrow and word spans. Thus, the simple reasoning tasks seem to draw on processes common to both the simple span measures.

In conclusion, it seems that the complex reasoning tasks consistently correlate with the arrow span, suggesting common processes between them. However, the correlations were

marginal, perhaps suggesting that complex reasoning tasks draw on a more active spatial resource. The simple reasoning tasks consistently correlate with both the arrow and word spans. The simple spans loaded the general factor in Experiments 1-3 CFA, and this could be the commonality between them and the simple reasoning tasks.

4.3.3.4 The Reasoning Measures

The next set of correlations concerns the relationships between the reasoning tasks themselves. Table 4.9 data relates to the work of Stanovich (e.g., 1999) who suggests that individuals giving the normative response on one task are more likely to give it on another. He suggests that these patterns of covariance would not be observed if deviations from normative responding were simply error variance.

	abstract	condition	simspat	comspat	syllogism
abstract	1				
condition	.45**	1			
simspat	.45**	.40**	1		
comspat	.23	.06	.30*	1	
syllogism	.16	.34*	.28*	.48***	1
mulquan	.19	.26 ⁺	.53***	.36**	.22

(two-tailed sig. levels: ⁺ = $p < 0.1$, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$)

Table 4.9 - Experiment 5 correlations between the reasoning tasks

The data in Table 4.9 relates to the relationships between the reasoning measures used. Two clusters of significant correlations can be seen when looking at Table 4.9. The first cluster refers to the simple reasoning tasks. There are moderately significant correlations between the abstract, conditional and simple spatial problems ($r = .40$ to $r = .45$). Both the abstract and conditional measures rely on understanding propositional connectives, whilst simple

spatial problems have been shown to be solved using either verbal and/or spatial strategies (e.g., Marquer & Pereira, 1990).

The results in Table 4.8 showed that the simple reasoning tasks correlated with both the simple arrow and word spans. The arrow and word spans loaded the general factor, as well as the independent factors, in Experiments 1-3 CFA, perhaps suggesting this general resource as a commonality between the two. The propositional based tasks might require only a low memory demand, and therefore draw only on general resources. The only one of these simple reasoning tasks to cross correlate with the complex reasoning tasks is the simple spatial measure, and to some extent the conditional task. This simple spatial task was shown to be the one that correlated most significantly with both the arrow and word spans (see Table 4.8). Thus, this could be interpreted as the simple spatial task drawing on spatial and general resources.

The second cluster of correlations refers to the complex reasoning tasks. There are moderate, mostly significant, correlations between the complex spatial, syllogistic and multiple quantification problems ($r = .22$ to $r = .53$). The multiple quantification task is based upon syllogistic arguments and thus it is not surprising that they are positively related. The complex spatial task (five-term series) correlates significantly with both the syllogistic and multiple quantification task which is consistent with research showing that quantifiers can be dealt with using spatial strategies (Ford, 1994). These tasks seem to correlate highest with each other and with the arrow span. This may reflect the fact that they rely mostly on capacity measured by a complex range of span measures, but also that they require a spatial resource for correct responding.

Stanovich and colleagues would expect a generally positive matrix for the reasoning measures - where correlations cross all tasks. A matrix where two clusters of correlations for the reasoning tasks have been identified is not consistent with this. The general picture to emerge here is a positive matrix, where the lowest cross-correlation was between the abstract and syllogistic task ($r = .16$), and the highest between the simple spatial and multiple quantification task ($r = .53$). The direction and general positivity of the matrix conforms partly to Stanovich and West's (1998) claims. What is more difficult to explain in terms of their proposals is the two clusters of correlations that emerged within the general matrix. However, the abstract, conditional, and multiple quantification task showed low reliabilities here, and this could be the reason for an incomplete cross-correlated pattern.

4.3.3.5 Further Analysis

The reasoning tasks were also investigated dependent on whether they could be split between easy and difficult problems. Past research suggests that problems built on the assumption that only one model needs to be constructed in order to reach a conclusion should be easier than problems where multiple models are needed.

	% correct
simspat1M	63
simspatINV	38
comspat1M	75
comspatMM	50
syllogism1M	88
syllogismMM	25
multi1M	60
multiMM	33

Table 4.10 - Experiment 5 % of correct responses to 1M, MM and INV problems

Problem difficulty percentages for the reasoning tasks are presented in Table 4.10. On the whole, the table shows that one model problems appear easier than multi model problems throughout the reasoning set. The most difficult model type appear to have been the multi-model syllogistic and multi-model multiple quantification problems. Table 4.10 shows an average of 6 out of 9 (67%) abstract problems were solved accurately. An average of 9 out of 16 (56%) conditional inferences were solved logically. An average of 8 out of 16 (50%) simple spatial inferences were solved correctly, whilst 10 out of 16 five-term problems (63%) were answered accurately. The complex spatial accuracy rate is similar to that found in previous experiments. An average of 9 out of 16 (56%) syllogisms were solved (consistent with previous experiments), whilst an average of 8 out of 16 (50%) multiple quantifiable problems were correctly answered.

A number of t-tests were also performed on the data in order to assess whether there were significant differences between the number of single model versus multiple model problems solved (see Appendix B1.3 for full results). Participants performed significantly better on the *simspat1M* versus *simspatINV* problems ($d.f. = 50, t = 8.24, p < .001$). Participants performed significantly better on the *comspat1M* versus *comspatMM* problems ($d.f. = 50, t = 5.05, p < .006$). This difference was not found in Experiment 4, although Experiments 1-3 support the re-emergence of this distinction. Participants performed significantly better on the *syllogism1M* versus *syllogismMM* problems ($d.f. = 50, t = 12.19, p < .001$). Participants performed significantly better on the *mulquan1M* versus *mulquanMM* problems ($d.f. = 50, t = 18.72, p < .001$).

4.3.4 Discussion

A weak relationship was shown between the two simple span measures used in Experiment 5, arrow and word. Apart from Experiment 1 and 5, this relationship was significant and always positive, perhaps suggesting some commonalities underlying the two tasks. In addition to this, the CFA in Experiments 1-3 produced a model where there were significant loadings of the word span on factor 1 and the arrow span on factor 2, and they both significantly loaded factor 3. It might be the case that these two simple measures, proposed in previous studies to be purely passive tasks for verbal and spatial storage (e.g., Shah & Miyake, 1996), are actually measuring some common processes as well. The results for the simple span tasks could be interpreted as measuring the capacity of separate parts of the CE component, as well as a general executive system, and this would explain the moderate correlations between them, and the loadings of both tasks on factor 3.

The results from Experiment 5 suggest that individual differences in simple working memory resources, as measured by the arrow and word span, do not readily predict differences in performance on the whole range of reasoning tasks investigated. The following interpretation accounts for the factor loadings of the simple word and arrow span tasks on the general factor in Experiments 1-3 CFA, as well as their loadings on the verbal and spatial factors, respectively. In other words, the results collectively suggest the need to assume some processing commonalities between the two simple span tasks, and that they perhaps reflect more than passive storage of information alone.

There appear to be stronger correlations between the simple reasoning tasks and both the arrow and word spans. In contrast, the complex reasoning tasks showed strongest correlations with the arrow span. When examining the correlations with the simple

reasoning tasks, it might be that these deductive tasks draw more on a general resource, requiring a low memory demand, as measured by both the simple spans. However, the simple spatial task also showed moderate correlations with the complex reasoning tasks. This might suggest that as well as a general resource, in line with the more complex reasoning tasks, it might also require additional spatial processing. In contrast, the complex reasoning measures only correlated significantly with the arrow span, perhaps suggesting that there is a higher memory demand, requiring spatial representations, drawing on independent spatial and general resources. However, it is important to highlight that the reasoning tasks used in Experiment 5 were evaluation tasks rather than production tasks. Having all conclusions in front of the participant during the problem solving process may place less demands on working memory, and hence reduce the chances of finding significant relationships overall.

The results in Experiment 5, in respect to the syllogistic and spatial inferences, are consistent with those found in previous experiments. For the complex deductive tasks, there were some significant relationships with the simple arrow span, indicating the need for some spatial resource in order to accurately solve these problems. This holds across all the experiments - in respect to the five-term series spatial and syllogistic inferences.

These results suggest that for the simple reasoning tasks, performance depends upon general resources of working memory, as measured by both the simple spans. These results also suggest that for the complex reasoning tasks, accurate performance depends upon spatial resources of working memory. This is due to the marginal correlations with the arrow span, but not with the word span. Again, this lends support to the notion that, for complex

deductive reasoning, spatial executive resources of working memory resources may be associated with this process.

It could still be the case that the two simple spans represent specific components within a general CE resource of working memory. As in Experiments 1-3, this may imply that a general CE resource accounts for the processing of complex and simple reasoning tasks. However, it might also be the case that the complex reasoning tasks additionally rely to some extent on domain-specific parts of this same sub-system, especially the spatial component. As in Experiments 1-3, a more sophisticated analysis is needed, over and above the correlational data, in order to best determine the functions of these resources (CFA in Section 4.5). Factor analysis will also enable the memory resources to be loaded by the reasoning tasks in order to find significant influences on their performance.

The correlational matrix from this experiment was found to contain two significant clusters of results. One cluster included the simple reasoning tasks, whilst the other included the complex reasoning tasks. This finding is only partly in agreement with the explanation put forward by Stanovich and West (1998). They found that a range of problem-solving and deductive reasoning tasks cross-correlated significantly. The findings here are consistent in that a generally positive matrix was observed. What is inconsistent is the fact that two significant clusters of correlations emerged from this general matrix. It might be that the significant predictor of cognitive ability, proposed by Stanovich and West (1998), might only represent the capacity required for performing complex deductive tasks, but not the simpler ones. This finding will be followed up in Experiment 6 to try and substantiate the two cluster result.

In relation to the representations that people must construct when reasoning, the experimental results support the notion that more errors are made on the multiple model reasoning problems. This holds for all the tasks where this split can be made between single and multiple model problems. Although the correlations were smaller in this experiment, compared to the others, the findings still have implications for reasoning theories. By combining all the results from Chapter 4, taken in conjunction with the CFA results, a discussion of the implications for the three classes of theory, introduced in Chapter 2, can be made. These implications will be summarised in the general discussion (Section 4.6).

Experiment 5 suggests that working memory resources, as measured by the arrow and word span, account for only a small proportion of the variance in reasoning ability. Maybe it is the case that, even for simpler deductive tasks, the magnitude of correlations would be highest with the complex working memory spans. So in order to further investigate this, Experiment 6 replicated Experiment 5, but with the addition of the complex span measures.

4 . 4 EXPERIMENT 6

4 . 4 . 1 Aim

From Experiment 5 it can be seen that some of the simpler reasoning tasks were associated with the simple working memory span measures. This may imply that these simpler deductive tasks required mainly general resources of working memory in order for correct responding, as implied by the arrow and word span's common loading on the general factor. However, the more complex reasoning measures showed weaker relationships with the simple spans, especially the word span. This may imply that these tasks required both processing and storage resources afforded by spatial and general executive resources. Thus

in Experiment 6, the complex working memory span tasks were re-introduced to substantiate these results.

Experiment 6 investigates simple and complex reasoning tasks, together with simple and complex working memory span tasks. Experiment 6 provides a further test of whether the simple reasoning tasks rely on passive storage resources, such as the PL and VSSP, in contrast to the complex reasoning tasks, or whether they rely on a general processing resource. In light of Experiment 5, it might be expected that the simpler reasoning tasks draw on general resources, and the correlations with the complex span measures may determine this. This experiment also makes comparisons with the work of Stanovich and colleagues (e.g., 1998). Stanovich would expect positive correlations between all the reasoning tasks used. He would also expect the tasks to correlate with measures of capacity (in comparison to his ability measures).

Experiment 6 investigated the degree to which individual differences in working memory capacity can explain variation in performance on a wide variety of reasoning types. Four working memory span tasks were used - the simple word and arrow spans, plus the complex TTT and verbal spans (Shah & Miyake, 1996; Daneman & Tardif, 1987; Daneman & Carpenter, 1980). This was in order to assess the relationship of reasoning to the three systems indicated by the CFA in Experiments 1-3. The span tasks were given together with six reasoning tasks - three-term and five-term series spatial inference tasks, abstract and conditional inference tasks, and syllogistic and multiple quantified inferences. The tasks were all presented visually and were evaluative in nature. They can be compared with Experiment 5 tasks due to the non-thematic content in both.

4 . 4 . 2 Method

Participants

Forty-eight participants took part in this study. The participants were undergraduate students and postgraduate students at the University of Plymouth, and they received course credit or cash payment for participating. None of the sample had prior training in logic. All participants were native English speakers.

Procedure

Experiment 6 was carried out in two testing sessions per participant, the sessions being of approximately an hour long each. Participants signed up in a time slot on an experimenter sheet for the first session. Session 1 consisted of five tasks. All participants carried out these tasks in the following order, (where an asterix (*) indicates a computerised task):-

complex spatial TTT span* ,

complex sentence span* ,

simple arrow span* ,

simple word span* ,

conditional inference.

Session 2 consisted of five tasks. All participants carried out these tasks in the following order:-

simple spatial inference,

abstract reasoning,

syllogistic reasoning,

complex spatial inference,

multiple quantification.

All the tasks have been explained in previous experiments. Therefore, the next section relates to the results found.

Simple Verbal Word Span (word)

The procedure and scoring is identical to Experiment 1.

Simple Spatial Arrow Span (arrow)

The procedure and scoring is identical to Experiment 1.

Complex Verbal Sentence Span (verbal)

The procedure and scoring is identical to Experiment 1.

Complex Spatial Tic-Tac-Toe Span (TTT)

The procedure and scoring is identical to Experiment 2.

Abstract Reasoning Task (abstract)

The procedure and scoring is identical to Experiment 5.

Conditional Inferences (condition)

The procedure and scoring is identical to Experiment 5.

Simple Spatial Inference (simspat)

The procedure and scoring is identical to Experiment 5.

Complex Spatial Inference (comspat)

The procedure and scoring is identical to Experiment 2.

Syllogistic Inference (syllogs)

The procedure and scoring is identical to Experiment 3.

Multiple Quantification Task (mulquan)

The procedure and scoring is identical to Experiment 5.

4 . 4 . 3 Results

In addition to correlational analyses and t-tests, confirmatory factor analysis was also performed on the data from this experiment. The correlational analysis and discussion of findings for this experiment will be reported here. However, the factor analysis for Experiment 6 will be explored in the multi-group factor analytic section following this study (Section 4.5).

As discussed in previous experiments, there are multiple scoring methods for many of the span tasks. However, only one method per task will be used throughout the results section. The cross-correlations of the global and normal span scores across different measures were between .87*** and .94***. This suggests that both scoring methods reflected the same underlying construct. Thus, for simple arrow, simple word and complex verbal spans, the global score will be used throughout.

4.4.3.1 Descriptive Statistics

Descriptive statistics for the ten measures used in Experiment 6 may be found in Table 4.11.

The table also contains data relevant to the reliability of the tasks used.

	Mean	SD	Min	Max	Skewness	Reliability
<i>Span Tasks</i>						
arrow	23.58	10.46	6	52	.58	.85
word	49.71	17.07	25	108	1.18	.88
verbal	37.13	20.52	10	100	1.02	.91
TTT	7.96	5.04	0	22	1.01	.84
<i>Reasoning Tasks</i>						
abstract	5.40	1.97	1	9	-.25	.56
condition	8.63	1.99	5	14	.29	.29
simspat	8.10	2.94	2	15	.22	.66
simspat1M	5.27	1.55	2	8	-.26	-
simspatINV	2.83	2.14	0	8	.40	-
comspat	9.21	3.14	1	15	-.44	.68
comspat1M	5.27	1.85	1	8	-.46	-
comspatMM	3.94	2.14	0	8	-.08	-
syllogism	8.02	3.00	1	16	.45	.75
syll1M	6.29	1.57	1	8	-1.33	-
syllMM	1.73	2.18	0	8	1.06	-
mulquan	6.54	1.92	2	10	-.26	.31
mulquan1M	5.25	1.59	2	8	-.36	-
mulquanMM	1.29	1.11	0	4	.46	-

Table 4.11 - Descriptive statistics for Experiment 6 tasks (n=48)

Table 4.11 shows that all tasks produced a good overall range of scores, both span and reasoning measures. The distribution of scores in all tasks was at an acceptable level, there are no heavily skewed ranges of data in any of the measures used.

The reliability of the span tasks was again assessed by performing a split-half test calculation on each of the four measures used. The Spearman-Brown formula was then used to provide an estimate of reliability for each task. The corrected correlations for the span tasks are shown in Table 4.11. Table 4.11 shows that all the span tasks produced a reliable odd-even corrected correlation of over .80. These corrected correlations are highly satisfactory for use in group measurement (Rust & Golombok, 1999). It can be seen that the arrow span in this experiment produced a much higher corrected correlation than in the previous experiment (in Experiment 5 it was .64). This low correlation in Experiment 5 may have reduced the correlations overall for this task. There were no differences in procedure for the arrow span between experiments, and thus the contrast in corrected correlations is probably due to chance variation.

The reliability of the reasoning measures was assessed using Cronbach's alpha. Table 4.11 shows that the simple and complex spatial inferences, together with the syllogistic task, were all satisfactorily reliable (between .66 and .75). These are comparable to Experiment 5 where these correlations were between .67 and .84. The abstract task produced a reliable corrected correlation of .56, which is of some use in group measurement. This can be compared to Experiment 5 where the abstract task did not produce a reasonable corrected correlation (at .34). As in Experiment 5, the conditional task (.29 in Expt. 6, .30 in Expt. 5) and the multiple quantification task (.31 in Expt. 6, .28 in Expt. 5) did not produce a reliable corrected correlation. This will be taken into account when looking at the correlations, to be discussed next.

4.4.3.2 The Span Measures

The first set of correlations relates to the relationship between the simple and complex span measures and then to their influence on the reasoning tasks (see Appendix B1.12 for the full matrix). In the previous experiments it could be seen that the correlations between the simple span tasks, arrow and word, were ambiguous. In Experiment 1 and 5 the relationship was positive, but not significant. In Experiments 2, 3 and 4 the relationship was positive and significant. The word span also loaded the verbal factor, and the arrow span the spatial factor, whilst they both loaded the general factor. The relationship between the arrow and word spans was interpreted as reflecting either the VSSP and PL components, respectively, or the activity of modality dependent components of the CE. Also seen previously, the complex span measures, verbal and spatial, were to some extent associated together, and loaded on the general and specific factors. The relationship between the complex span measures was interpreted as reflecting a general working memory resource accounting for both the storage and processing of information - the modality independent CE. The results can be found in Table 4.12.

	arrow	word	verbal
arrow	1		
word	.42**	1	
verbal	.33*	.69***	1
TTT	.33*	.11	.08

(two-tailed sig. levels: + = $p < 0.1$, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$)

Table 4.12 - Experiment 6 correlations between the four span tasks

Table 4.12 presents the relationships between the span measures used. The simple span measures will be discussed first. The relationship between the arrow and word spans was significant in Experiment 6 ($r = .42$, $p < .01$), as it was in Experiments 2-4 ($r = .28$ to $r =$

.38, $p < .10$ to $p < .01$). However this correlation did not reach significance in Experiments 1 and 5 (in Expt.1 it was .23, in Expt.5 it was .17). This finding, overall, disagrees with the notion that the two tasks are purely representative of underlying passive storage systems for verbal and spatial information, which act independently. However, the resources of the PL and VSSP could still be part of what is being measured by these two simple span measures. The arrow span showed higher reliabilities in the experiments where the results show this significant correlation. This may undermine the interpretation made at the start of this thesis that the simple word and arrow spans tap just the PL and VSSP components of working memory, and will be discussed in detail later on.

Secondly, the complex span measures will be considered from Table 4.12. The highest correlation is again between the complex verbal and word spans ($r = .69$, $p < .001$). This result is consistent throughout all the experiments. The arrow span correlates significantly with both the complex span measures, TTT and the complex verbal span ($r = .33$, $p < .05$ for both), as seen in Experiment 4. The data indicates there are no significant correlations between the complex span measures ($r = .08$). This is inconsistent with Experiments 1-4 where significant relationships were found between the complex spatial and complex verbal measures ($r = .24$ to $r = .41$). The result here is more compatible with Daneman and Tardif's (1987) data, where they found dissociations between spatial and verbal span measures.

4.4.3.3 The Span and Reasoning Measures

The span measures were looked at in relation to all the reasoning measures used, and this can be found in Table 4.13.

	arrow	word	verbal	TTT
abstract	.05	.24	.14	-.20
condition	.18	.24	.13	.20
simspat	.53***	.39**	.28	.26
comspat	.28 ⁺	.07	-.14	.14
syllogism	.13	.04	.21	.34*
mulquan	.26 ⁺	.00	-.14	.52***

(two-tailed sig. levels: ⁺ = $p < 0.1$, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$)

Table 4 . 13 - Experiment 6 correlations between the span and reasoning tasks

Table 4.13 presents the correlations between the span measures and the reasoning measures. As in Experiment 5, the first comparisons will be made between the tasks already seen in prior experiments, namely the five-term series spatial and syllogistic measures. The five-term series spatial inferences, as in Experiment 5, correlate with the arrow span ($r = .28$, $p < .10$ here and $r = .27$, $p < .10$ in Expt.5). They do not correlate significantly with the word span, or the complex measures ($r = .07$, $r = -.14$, $r = .14$). In previous experiments the correlations between the five-term series spatial task with the complex span measures were higher, and usually significant with the complex spatial spans (an average $r = .40$ represents the relationship between the spatial reasoning and complex spatial tasks seen in previous experiments). The differences seen here could be explained in reference to the type of spatial reasoning task used. In this instance it was an evaluative task with non-thematic content. The syllogistic inferences only show significant relationships with the complex TTT span ($r = .34$, $p < .05$). This is consistent with links found previously with spatial working memory spans (an average of $r = .34$ for both the complex spatial spans in Expts. 1 and 3). An explanation for the lower correlations, especially for the five-term series task, with the span measures in Experiment 6 may be because the reasoning tasks were evaluative

in nature. The evaluative tasks may not draw on working memory resources as heavily as the production tasks used previously.

The other complex task used here was the multiple quantification task. As in Experiment 5, low correlations can be seen between it and the arrow and word span ($r = .26$, $r = .00$), although there is a significant correlation with the complex TTT span ($r = .52$, $p < .001$). Multiple quantification problems are related to the syllogistic task due to both measures relying on quantifiable statements, and they therefore might be expected to show similar relationships to the working memory capacity measures. The correlations between both these tasks and the span measures are similar in Experiment 6. Both these types of problem appear to have links with spatial working memory. The major finding for the complex reasoning tasks is that they are most strongly linked with spatial working memory spans than verbal ones, although most of the correlations are low. An explanation might be that the complex tasks rely on spatial executive processing, as measured by the arrow and TTT spans. These findings are consistent with Experiment 5 and with previous experiments that showed a connection between these complex deductive measures and spatial working memory.

The simple reasoning tasks will be discussed next. The findings for the abstract and conditional tasks are consistent with Experiment 5. The abstract and conditional tasks correlate positively with both the simple span measures, but only marginally significantly with the word span ($r = .24$ for both reasoning tasks). The low correlations seen between the spans and the conditional task might be due to the low reliability of this reasoning task ($r = .29$). These tasks could be interpreted to rely more on a verbal resource, as measured by the simple word span, assuming a role for partly passive and partly active processing. The

simple spatial task correlated both positively and significantly with both the simple arrow and word spans ($r = .53, p < .001$; $r = .39, p < .01$), as in Experiment 5. The only significant correlations between the simple reasoning tasks and the complex span measures was in relation to the simple spatial reasoning task ($r = .28, r = .26, p < .10$ for both). As in Experiment 5, the simple reasoning tasks nearly all correlate with both the arrow and word spans. The simple spans loaded the general factor in Experiments 1-3 CFA, and this could be the commonality between them and the simple reasoning tasks. This may suggest a contrast between the way these simple problems are dealt with in working memory and the way the complex reasoning tasks are handled.

4.4.3.4 The Reasoning Measures

The next set of correlations concerns the relationships between the reasoning tasks themselves. Table 4.14 data relates to the work of Stanovich (e.g., 1999) who suggested that cross-correlations should be observed across a wide range of deductive tasks, and that these tasks should be predicted from a combined influence of cognitive ability and thinking style measures. This data also relates to Experiment 5 where generally two clusters of correlations were observed. One cluster related to the simple reasoning measures which cross correlated quite highly. The other cluster of correlations related to the complex reasoning measures which also cross correlated quite highly.

The data in Table 4.14 relates to the relationships between the reasoning measures used. These results are entirely consistent with Experiment 5. As in that experiment, two clusters of correlations can be identified. One cluster relates to the simple reasoning tasks, and the other to the complex.

	abstract	condition	simspat	comspat	syllogism
abstract	1				
condition	.31*	1			
simspat	.31*	.23	1		
comspat	.00	.16	.44**	1	
syllogism	.19	.10	.30*	.25	1
mulquan	.08	.10	.47***	.36*	.41**

(two-tailed sig. levels: + = $p < 0.1$, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$)

Table 4 . 14 - Experiment 6 correlations between the reasoning tasks

Again, there are moderate relationships between the simple reasoning measures (ranging from $r = .23$ to $r = .31$, $p < .05$). Also, as in Experiment 5, the only simple task to correlate with the complex ones is the simple spatial task (ranging from $.30$ to $.47$, $p < .05$).

Additionally, the simple spatial task is the only one to correlate with both the arrow and word spans (see Table 4.13).

The second cluster of correlations refers to the complex reasoning tasks. There are moderate, and mostly significant, correlations between the complex spatial, syllogistic and multiple quantification problems (ranging from $r = .25$ to $.47$). This result is nearly identical as that found in Experiment 5 (ranging from $r = .22$ to $r = .53$). These tasks seem to correlate highest with each other and with the spatial spans. This may reflect the fact that they rely mostly on capacity afforded by an executive system that is best measured by the complex range of span measures, but also that they require mostly spatial resources for correct responding.

These results are inconsistent with the work of Stanovich and colleagues. They would expect significant cross correlations between all the reasoning measures used (Stanovich &

West, 1998). Due to the two clusters of correlations identified, the notion that participants giving the normative response on one task are usually significantly more likely to give it on another can not be fully supported. This might suggest that rather than one type of cognitive capacity being relevant, there may be two capacities that relate to certain deductive tasks, one to the simple and one to the complex reasoning measures.

4.4.3.5 Further Analysis

As in Experiment 5, the reasoning tasks were further analysed by looking at the easy versus difficult divide between problems. Past research showed that performance on multiple model problems was harder than on single model problems.

	% correct
sims pat1M	66
sims patINV	38
com spat1M	66
com spatMM	50
syll 1M	79
syll MM	22
mul quan1M	53
mul quanMM	22

Table 4.15 - Experiment 6 % of correct responses to 1M, MM and INV problems

Percentage accuracy for the reasoning tasks are presented in Table 4.15. On the whole the table shows that one model problems appear easier than multi model problems throughout the reasoning set. The most difficult model type appear to have been the multi-model syllogistic and multiple quantification problems, as in Experiment 5. The following percentages relate to Experiment 6 findings, those shown in italics relate to Experiment 5

findings for comparison. Table 4.15 shows an average of 5 out of 9 (56%, 67%) abstract problems were solved accurately. An average of 9 out of 16 (56%, 56%) conditional inferences were solved logically. An average of 8 out of 16 (50%, 50%) simple spatial inferences were solved correctly, whilst 9 out of 16 five-term problems (56%, 63%) were answered accurately. An average of 8 out of 16 (50%, 56%) syllogisms were solved, whilst an average of 7 out of 16 (44%, 50%) multiple quantifiable problems were correctly answered.

A number of t-tests were performed on the data in order to examine the relative difficulty of single and multiple model problems (see Appendix B1.4 for full results). As in Experiment 5, participants performed significantly better on the *simspat1M* versus *simspat1NV* problems ($d.f. = 47, t = 7.33, p < .001$). Participants performed significantly better on the *comspat1M* versus *comspat1MM* problems ($d.f. = 47, t = 3.73, p < .005$). This was also observed in Experiment 5. Participants also performed significantly better on the *syllogism1M* versus *syllogism1MM* problems ($d.f. = 47, t = 13.55, p < .001$); as they did on the *mulquan1M* versus *mulquan1MM* problems ($d.f. = 47, t = 14.02, p < .001$). These findings are all consistent with those found in Experiment 5.

4.4.4 Discussion

The independent verbal and spatial factors from Experiments 1-3 CFA were investigated in Experiment 5 by using the two simple span tasks and additional simple reasoning tasks. These results were still ambiguous concerning the role these factors played in the reasoning process, and therefore the complex span measures were re-introduced in Experiment 6 to try and better identify the factors that different reasoning tasks draw upon. Experiment 6 allowed the final test of how to interpret the three factors from the first series of

experiments. It is worth emphasising at this stage that the reasoning tasks differ in Experiments 5 and 6, from the rest of the experimental series. The other experiments used production measures to assess performance on the problems. In Experiments 5 and 6, the reasoning tasks were evaluative in nature. This may account for some of the discrepancies seen in Experiments 5 and 6 as compared with the other experiments. What follows is a summary of findings of Experiment 6, accounting for the consistencies and inconsistencies between this and the other experiments.

The reasoning task cross correlations, as well as their relationships with the working memory capacity measures, were much lower in magnitude in Experiments 5 and 6 than in the other experiments. In response to these differences, one explanation may be that evaluation tasks require less working memory resources in order to correctly respond to a set of conclusions. Experiments 1-4 required participants to produce a conclusion, whereas in Experiments 5 and 6 the participants were only required to choose a conclusion they thought was correct from a list of possibilities. This information should be kept in mind when discussing the results from this experiment, and that of Experiment 5. Significant cross correlations were seen between the verbal and spatial span measures, as in the previous experiments, although this was the first experiment where a non-significant correlation was observed between the complex TTT and complex verbal span tasks. Again, there was no total dissociation between verbal and spatial spans, as in Shah and Miyake (1996).

The correlational results from Experiment 6 suggest that complex reasoning task performance required spatial working memory resources. The five-term series, syllogistic and multiple quantification problems all correlated highest with the arrow and TTT spans.

There were lower correlations here with verbal working memory resources. This is consistent with Experiment 5 results where the complex reasoning tasks correlated with the arrow span. For the simple abstract and conditional reasoning tasks, the only significant correlations here were with the simple word span, whereas the simple spatial task correlated significantly with both simple and complex verbal and spatial resources. This is not entirely consistent with Experiment 5 results, where all the simple reasoning tasks correlated similarly with both the word and arrow spans. However, with the addition of the complex span measures in Experiment 6, it can be seen that only the simple spatial task correlated significantly with them. This might be expected given that in Experiment 5 it was the only simple task to correlate significantly with both the arrow and word spans, but also with the complex reasoning tasks, perhaps suggesting some commonalities between this task and the more complex ones.

These results may suggest that performance on the complex reasoning tasks, which rely on processing and storage resources, do relate to central pools of working memory resources, as measured by the complex and simple spatial span tasks. This is in contrast to the simple reasoning tasks, which have stronger relationships with just the simple verbal resources, although the simple spatial reasoning task required resources more in line with the complex reasoning tasks. Thus, it might be that the complex reasoning tasks, and the simple spatial task, require both general and spatial CE resources in order to perform them. For the other simple reasoning tasks it is most likely that the memory load involved is not as demanding, as shown by the lower correlations with the complex spans, which could be interpreted as their reliance on only general resources of working memory, without recourse to additional independent executive processes for verbal and spatial information.

Again, this lends support to the notion that, especially for complex reasoning exercises, central working memory resources may be associated with this process, but that this entails more than just general processing. This finding that spatial processing is an important factor in their performance holds across all the experiments - in respect to the five-term series spatial and syllogistic inferences. For the simpler tasks, only general executive resources may be required. It could be the case that the simple tasks would only load the general factor in a CFA, reflecting the fact that they don't have a huge storage component, but do have a processing component. It might also be the case, as in Experiments 1-3, that the complex reasoning tasks, in addition to loading the general factor, also load the specific factors, perhaps reflecting the fact that they need additional spatial and/or verbal processing. As in Experiments 1-3, this may imply that a general domain explanation of working memory is required, but that additionally, verbal and spatial executive processing factors are also essential. This will be further addressed in the factor analysis section at the end of this section.

As in Experiment 5, the data in the correlational matrix in Experiment 6 support two clusters of correlations for the reasoning tasks. The simple reasoning tasks again cross correlate together, as do the complex reasoning tasks. This finding can be related to Stanovich and West (1998) who proposed that a general capacity (as measured by cognitive ability) accounted for the total cross correlations seen in their studies. Due to the two clusters of correlations being consistent over both Experiments 5 and 6, the results can not be explained by chance alone. It is true that the tasks used here are different to the ones used by Stanovich et al. (1998), except for the syllogistic task. Thus, it might be that the tasks used by Stanovich fit into the cluster of correlations for the complex reasoning tasks seen in Experiments 5 and 6. These complex deductive measures seem to have

commonalities that could be interpreted as relying on one capacity resource. What Stanovich did not find was a second cluster of correlations in relation to another identifiable group of tasks, in this instance, the simple reasoning task cluster. It could be that there are two capacity resources available, one more implicated in the complex tasks, and another more implicated in the simple tasks.

The experimental results, as with all the other experiments, also support the notion that the more models that need to be manipulated in order to reach a valid conclusion, the more errors will be made on reasoning problems. This was consistent for all the tasks where this split between single and multiple models was made. These results again have implications for theories of the reasoning process. As for Experiments 4 and 5, these theoretical accounts will be covered in the general discussion in Section 4.6. The CFA, as well as the correlational analysis, can then be combined from all three experiments to provide the most consistent interpretation of the data.

Experiments 4-6 involved a wide range of reasoning tasks and a number of working memory span measures. In order to find a best fit model for the three correlational data sets, a multiple group factor analysis was carried out on Experiments 4-6. A factor analytic investigation would help to fully understand the pattern of data in terms of models of working memory. What follows is an investigation into the best fit models of the data sets for Experiments 4-6, and the results and interpretation made of this.

4 . 5 CONFIRMATORY FACTOR ANALYSIS

EXPERIMENTS 4, 5 AND 6

4 . 5 . 1 Rationale for the Factor Analysis

The spatial and temporal inference data from Experiment 4 suggested that individual differences were predicted by both complex and simple verbal and spatial working memory spans. The five-term spatial inference data from Experiments 5 and 6 suggested that individual differences were predicted by both complex and simple spatial working memory spans, although there was some influence from the verbal spans (negative correlation in Experiment 6). The syllogistic inference and multiple quantification task data from both Experiments 5 and 6 suggested that individual differences were best predicted by both complex and simple spatial spans. A potential explanation for the findings for the complex reasoning tasks - temporal, five-term series, syllogistic and multiple quantification, is that they draw more heavily on general CE resources, with additional executive processing for spatial information.

The abstract and conditional task data from Experiments 5 and 6 showed highest correlations with the simple spans, especially the word span. The three-term spatial inference data from Experiments 5 and 6 showed similar correlations with the simple and complex spans, both verbal and spatial. One potential explanation for these findings for the simpler reasoning tasks - abstract, conditionals and three-term series, is that they draw more heavily on general resources of working memory (as measured by the simple spans), requiring only minor processing, although the simple spatial task seems to require additional spatial processing, as did the complex tasks. This general processing, partly measured by

the arrow and word spans, seems to be the most important component in performance of the simple reasoning tasks.

The interpretation that the complex spans measure verbal and spatial executive capacity resources and the simple spans measure a general executive capacity resource is evidenced by intercorrelations between the span measures across all three experiments. Cross-correlations between all the span measures were positive, as they were in Experiments 1-3. In Experiments 4 and 6, the complex verbal and spatial spans intercorrelated positively. Thus, what might underlie this finding is variations in central executive function, be it verbal or spatial in nature. There were positive correlations between the simple word and arrow spans throughout Experiments 4-6, a finding that is consistent with Experiment 1-3. A possibility is that the complex spans tap into executive processes that are verbal or spatial, and the simple spans partly tap the general, independent, CE.

There was a clearer separation between the complex verbal and spatial spans than in Experiments 1-3, although there were positive relationships between them. Shah and Miyake's (1996) complete dissociation between verbal and spatial spans was still not observed here. Overall, the general cross-correlations between span tasks, both simple and complex, through Experiments 4-6 showed positive relationships, as in the first three experiments. Therefore, these findings are not entirely consistent with Baddeley's (1986) model of a central executive with two peripheral subsystems. The inconsistency lies in the result between the simple span measures. Due to the positive, and mostly significant, relationships between the arrow and word span throughout the experimental series, it is difficult to interpret these measures as drawing on two separate sub-systems. Further evidence for the difficulty in interpreting these two measures as simple passive storage tasks

was seen in the factor analysis for Experiments 1-3. These two tasks were shown to load both the specific and general factors. This will be further investigated in the factor analysis for Experiments 4-6.

The following confirmatory factor analytic (CFA) models were run in order to further test the findings of the correlational data obtained from Experiments 4-6, as well as the patterns observed from the first three experiments. A number of different models were tested, including three and four factor models. The following summary contains the best fit models found from those that were originally investigated. Firstly, a three-factor multiple group analysis was run on Experiments 4, 5 and 6. This was in order to further test the best fit model found for Experiments 1-3 data. Section 3.5 contains the CFA results produced for Experiments 1-3. The best fit model of the data was a three factor, uncorrelated one. The simple word and complex verbal spans loaded factor 1 and factor 3 significantly. The simple arrow and complex letter spans loaded factor 2 and factor 3 significantly, whilst the complex TTT span only loaded factor 3 significantly. It must be emphasised that this is a multiple group analysis, which pools data across experiments to reach conclusions that are valid for all experiments. As in Experiments 1-3, for the tasks that were common between experiments, the parameters of the model were constrained for equality across experiments. This is an important point given the sample sizes used in relation to individual differences work.

4.5.2 Method and Results

A multiple group confirmatory factor analyses was run on all three experiments where the factors were loaded as shown in Table 4.16. The span measures were loaded as in

Experiments 1-3. The verbal and word spans were loaded on both factor 1 and factor 3.

The arrow and TTT spans were loaded on both factor 2 and factor 3.

The loadings of the reasoning measures on the three factors will be explained next. The complex reasoning tasks, five-term series spatial, syllogistic and multiple quantification tasks, were loaded on all the factors, as in Experiments 1-3. The simple reasoning measures were loaded onto specific factors - statistical and theoretical reasons for these specific loadings are discussed below. If the simple reasoning measures were allowed to load all factors, as the complex ones, it might make the definition of the factors by the span measures problematic. Thus, due to the addition of many more reasoning tasks, the simple tasks were also used to define the factors, past theory being used as a guide to define which tasks should load which factor. Past research showed that simple spatial tasks are disrupted by secondary tasks assumed to load the VSSP of working memory (Farmer, Berman and Fletcher, 1986). This research implied that simple spatial problems tend to be solved by using simple spatial resources. Therefore, the simple spatial inferences (three-term series) were loaded on both the spatial factor and the general factor (F2 and F3). Past research showed that conditional inferences are disrupted by secondary tasks assumed to load the PL and CE components of working memory (Toms, Morris & Ward, 1993; Klauer, Stegmaier and Meiser, 1997). The abstract task also requires reasoning about propositions, the conditional inferences being a sub-set of this type of problem. Therefore, the conditional and abstract reasoning tasks were loaded on both the verbal factor and general factor (F1 and F3).

	F1	F2	F3
arrow		✓	✓
word	✓		✓
verbal	✓		✓
TTT		✓	✓
three-term spatial inference		✓	✓
five-term spatial inference	✓	✓	✓
temporal inference	✓	✓	✓
abstract inference	✓		✓
conditional inference	✓		✓
syllogistic inference	✓	✓	✓
multiple quantified inference	✓	✓	✓

Table 4 . 16 - Set up of the orthogonal three-factor model for Experiments 4-6

The factors are interpreted as F1 - verbal resource, F2 - spatial resource, and F3 - general resource. The three factors were orthogonal in nature. The above loadings were constrained to equality across the experiments where the same tasks were involved. For instance, the simple span measures were constrained across Experiments 4, 5 and 6. The complex span measures were constrained across Experiments 4 and 6. The reasoning tasks used in Experiments 5 and 6 were also constrained to equality. The only reasoning task that was constrained to equality across all three experiments was the five-term series spatial problems. Even though the task was a production one in Experiment 4, as opposed to an evaluation task in Experiments 5 and 6, the structure of the problems were identical. The error variance associated with each of the tasks was also constrained to be equal across experiments in the same way. The appropriate EQS methods were set up to handle the data. The full EQS output can be found in Appendix B1.14.

chi-square	74.716
degrees of freedom	73
probability of chi-square	0.42
akaike`s information criterion	-71.28
bentler-bonett normed fit index	0.794
bentler-bonett nonnormed fit index	0.992
comparative fit index	0.994

Table 4 . 17 - Goodness of fit summary (three-factor model, uncorrelated) - multiple group analysis of Experiments 4-6

Table 4.17 gives the results of fitting the hypothetical three-factor model outlined above.

The Chi-square for testing the goodness of fit of the model is 74.716, based on 73 degrees of freedom. This result may be likely ($p < .42$) to occur with a sample of this size if the model truly holds in the population. Thus, considering the high fit indices, we may accept the model. The factor loadings for the tasks in Experiments 4-6 (three-factor model) can be found in Tables 4.17A-C.

MEASURE	LOADINGS			R-sq
	Factor 1	Factor 2	Factor 3	
	verbal resource	spatial resource	general resource	
arrow		.31***	.45***	.30
word	.40***		.62***	.54
verbal	.79***		.61***	1.00
TTT		.60***	.17	.39
five-term spatial inference	-.22	.56***	.36**	.37
temporal inference	.06	.89***	.25	.85

(significant loadings * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.005$)

Table 4 . 17A - Experiment 4 factor loadings for the three-factor uncorrelated model

MEASURE	LOADINGS			R-sq
	Factor 1	Factor 2	Factor 3	
	verbal resource	spatial resource	general resource	
arrow		.31***	.45***	.30
word	.40***		.62***	.54
three-term spatial inference		.43***	.67***	.64
five-term spatial inference	-.22	.56***	.36**	.37
abstract inference	-.20		.62***	.42
conditional inference	-.17		.51***	.29
syllogistic inference	-.01	.38***	.30*	.23
multiple quantified inference	-.16	.69***	.31*	.60

(significant loadings * = p < 0.05, ** = p < 0.01, *** = p < 0.005)

Table 4 . 17B - Experiment 5 factor loadings for the three-factor uncorrelated model

MEASURE	LOADINGS			R-sq
	Factor 1	Factor 2	Factor 3	
	verbal resource	spatial resource	general resource	
arrow		.31***	.45***	.30
word	.40***		.62***	.54
verbal	.79***		.61***	1.00
TTT		.60***	.17	.39
three-term spatial inference		.43***	.67***	.64
five-term spatial inference	-.22	.56***	.36**	.37
abstract inference	-.20		.62***	.42
conditional inference	-.17		.51***	.29
syllogistic inference	-.01	.38***	.30*	.23
multiple quantified inference	-.16	.69***	.31*	.60

(significant loadings * = p < 0.05, ** = p < 0.01, *** = p < 0.005)

Table 4 . 17C - Experiment 6 factor loadings for the three-factor uncorrelated model

In the three-factor model, especially for Experiment 5, the largest standardised residuals were mostly in relation to the correlations between reasoning tasks. It might be that there were correlations amongst the reasoning measures that were not fully explained by the three factors. In addition to this, it can be seen from the factor loadings in Tables 4.17A-C that none of the reasoning tasks loaded on factor 1 (the verbal resource). This is inconsistent with Experiments 1-3 where significant loadings for the syllogistic and five-term series reasoning problems were found on factor 1. It was therefore decided that, even though the fit was good for the three-factor model, another factor might be justified in explaining variance on the reasoning measures.

It can be seen from the cross-correlations of all the reasoning tasks used in Experiments 4-6 that there were positive relationships between them. However, two clusters of correlations did emerge - one group related to the simple reasoning measures, and one to the complex reasoning measures. This suggests that performance on one simple task can predict performance on another simple task, with similar predictions relating to the complex tasks. It may also highlight an underlying theme to all the reasoning tasks. Due to the addition of many more reasoning tasks in Experiments 4-6, a possibility could be that cognitive motivation or thinking styles influence performance. This motivational factor would be independent of ability and may connect these tasks on another level.

Stanovich and West (1998) found that cognitive capacity alone could not explain the variance amongst the reasoning tasks they studied. They found that thinking dispositions could serve as a predictor independent of differences in cognitive ability. The thinking styles that were investigated included thinking of alternative explanations to the problem at hand, how dogmatic people's beliefs were, and how open-minded they were. All of these

thinking dispositions are important features of the reasoning process. For instance, especially in relation to mental model theory, the motivation to search for alternative models, is necessary for correct responding. Thus, Stanovich and West (1998) could not explain variation in performance on a range of reasoning tasks without recourse to both a cognitive ability and thinking disposition factor.

Stanovich and West's (1997, 1998) results can be brought in here to expand on what was found in the current experimental series. Their cognitive capacity measure (in comparison to the working memory capacity measures used here) significantly correlated with performance on the deductive tasks they investigated. They suggest that errors in performance can be accounted for by limitations in a person's cognitive capacity. They also suggested that predictions can be made through looking at thinking dispositions, which can be related to the ability of the reasoner to search for alternative models or representations of the problem.

Comparing the experimental tasks used here with Stanovich's work, it would be expected that if reasoning performance does rely on variation in cognitive capacity, then they should load on what has been called the general factor, that reflects that capacity. But, another variable, that is linked to cognitive motivation or style, could underlie these tasks too. So, rather than it being solely capacity limitations, there could be something else explaining variations on these tasks, such as searching for alternative models of the problem, which is a major component of the mental model theory of reasoning (Johnson-Laird, 1983). These separate activities may not rely on working memory resources, but on the thinking style adopted by the person in order to solve the problem.

In Experiments 1-3, there were only a few reasoning tasks involved within each experiment, and therefore the extra factor accounting for thinking style did not emerge in the factor analysis. In Experiments 1-3 CFA only one correlation could be used in which to look for discrepancies between predicted and observed scores. In Experiment 4-6, many reasoning tasks were used and therefore allowed the addition of a thinking style factor in order to investigate more of Stanovich and West's (1997, 1998) claims.

The factor loadings for the three-factor model will not be discussed until the discussion of the next four-factor model. This is because the loadings are very similar between these two models. Therefore, the loadings on factors 1, 2, and 3 remain static and can be considered together with the results found for the four-factor model. The interpretation of factor loadings will follow the goodness of fit summary for a four-factor model.

The following confirmatory factor analytic model was run in order to further test the notion of a fourth underlying factor that influences reasoning performance. It may be that some other factor, as well as capacity, may explain variance on the reasoning tasks. A four-factor multiple group analysis was run on Experiments 4, 5 and 6. This involved combining the data from all three experiments into a single model in order to strengthen the final outcome. A multiple group confirmatory factor analyses was run on all three experiments where the factors loaded as shown in Table 4.18. The tasks in the four-factor model were loaded on factor 1, 2 and 3 in exactly the same way as in the three-factor model. The only difference with the four-factor model was that, additionally, all the reasoning measures were loaded onto a fourth factor that represented thinking style or disposition.

	F1	F2	F3	F4
arrow		✓	✓	
word	✓		✓	
verbal	✓		✓	
TTT		✓	✓	
three-term spatial inference		✓	✓	✓
five-term spatial inference	✓	✓	✓	✓
temporal inference	✓	✓	✓	✓
abstract inference	✓		✓	✓
conditional inference	✓		✓	✓
sylogistic inference	✓	✓	✓	✓
multiple quantified inference	✓	✓	✓	✓

Table 4 . 18 - Set up of orthogonal four-factor model for Experiments 4-6

The factors are interpreted as F1 - verbal resource, F2 - spatial resource, F3 - general resource, and F4 - thinking style or disposition. The factors were uncorrelated. The above loadings were constrained to equality across the experiments. The error variance associated with each of the tasks was also constrained across experiments. The appropriate EQS methods were set up to handle the data.

Table 4.19 gives the results of fitting the hypothetical four-factor model outlined above, where all four factors were orthogonal. The Chi-square for testing the goodness of fit of the model is 57.807, based on 66 degrees of freedom. This result may be likely ($p < .75$) to occur with a sample of this size if the model truly holds in the population. Thus, considering the high fit indices, we may accept the model.

chi-square	57.807
degrees of freedom	66
probability of chi-square	0.75
akaike's information criterion	-74.19
bentler-bonett normed fit index	0.840
bentler-bonett nonnormed fit index	1.040
comparative fit index	1.000

Table 4 . 19 - Goodness of fit summary (four-factor model, uncorrelated) - multiple group analysis of Experiments 4-6

The conclusion that the four-factor model (Table 4.19) fitted the data better than the three-factor model (Table 4.17) for Experiments 4-6 data was supported by the direct statistical comparison of alternative models. The chi-square difference test produced a significant chi-square, $\chi^2(7) = 16.91, p < .05$, suggesting that the three-factor model fits the data significantly worse than the four-factor model. However, both produced a non-significant chi-square to the data sets and this must be noted. From the factor loadings it can be seen that there were a number of significant loadings on factor 4 which also strengthens the case for the inclusion of this factor. Another important point is that the AIC value in the four-factor model was smaller than in the three-factor one (-74.19 as opposed to -71.28 in the three-factor model). The AIC provides a tool for comparing models that differ in their free parameters, and therefore is a good indicator as to whether the fourth factor is necessary. The three-factor model can be seen as a constrained version of the four-factor model. The four-factor model is less parsimonious than the three-factor model, but with a significant chi-square difference result and a lower AIC, the finding is strengthened. Considering the similarity between loadings for the three-factor versus four-factor model, the results for both will be discussed as for the four-factor model below.

MEASURE	LOADINGS				R-sq
	Factor 1	Factor 2	Factor 3	Factor 4	
	verbal resource	spatial resource	general resource	thinking disposition	
arrow		.34***	.48***		.34
word	.35***		.68***		.58
verbal	.80***		.59***		1.00
TTT		.61***	.22		.42
five-term spatial inference	-.09	.50***	.22*	.23*	.36
temporal inference	.01	.81***	.20	.55*	1.00

(significant loadings * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.005$)

Table 4 . 19A - Experiment 4 factor loadings for the four-factor uncorrelated model

MEASURE	LOADINGS				R-sq
	Factor 1	Factor 2	Factor 3	Factor 4	
	verbal resource	spatial resource	general resource	thinking disposition	
arrow		.34***	.48***		.34
word	.35***		.68***		.58
three-term spatial inference		.40***	.65***	.26*	.64
five-term spatial inference	-.09	.50***	.22*	.23*	.36
abstract inference	-.05		.44***	.55***	.50
conditional inference	.00		.34**	.41***	.29
sylogistic inference	.45***	.67***	-.10	.59***	1.00
multiple quantified inference	-.23*	.64***	.34*	.11	.59

(significant loadings * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.005$)

Table 4 . 19B - Experiment 5 factor loadings for the four-factor uncorrelated model

MEASURE	LOADINGS				R-sq
	Factor 1	Factor 2	Factor 3	Factor 4	
	verbal resource	spatial resource	general resource	thinking disposition	
arrow		.34***	.48***		.34
word	.35***		.68***		.58
verbal TTT	.80***		.59***		1.00
three-term spatial inference		.61***	.22		.42
five-term spatial inference		.40***	.65***	.26*	.64
abstract inference	-.09	.50***	.22*	.23*	.36
conditional inference	-.05		.44***	.55***	.50
sylogistic inference	.00		.34**	.41***	.29
multiple quantified inference	.45***	.67***	-.10	.59***	1.00
	-.23*	.64***	.34*	.11	.59

(significant loadings * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.005$)

Table 4 . 19C - Experiment 6 factor loadings for the four-factor uncorrelated model

Tables 4.19A-C give the results of the factor loadings of each of the eleven tasks on the four factors (see Appendix B1.14 for the full analysis). Note the similarity in loadings between tasks on factors 1, 2 and 3 in comparison to the three-factor model. Thus, the results above are supported by the previous three-factor model in Tables 4.17A-C. The following summary relates to Tables 4.19A-C, and to a four-factor orthogonal model. The loadings for the span measures will be discussed first, as these can be related to the results found in Experiments 1-3. Note that the loadings for the span measures are not much affected by whether a three-factor or four-factor solution are accepted. The loadings in Experiments 4-6 show that the word and verbal spans significantly load factor 1 (.35 and .80, respectively) and factor 3 (.68 and .59, respectively). This was also found in Experiments 1-3. However the loadings in Experiments 1-3 were slightly different. The word span loaded factor 1 similarly to Experiments 4-6 (.48). However, the verbal span did not load factor 1 as highly

in Experiments 1-3 (.47) as it did in Experiments 4-6 (.80). The word span loaded factor 3 similarly to Experiments 4-6 (.64), as did the verbal span (.64). The verbal span loaded factor 1 more than the word span. If the word span was only a passive measure it would only be expected to load factor 1. If the verbal span was an active measure it would only be expected to load factor 3, when in fact it loaded factor 1 most significantly. An interpretation of factor 1 might be more compatible with the idea of a domain-specific component of the CE, rather than it being the PL component of working memory.

The arrow span loadings in Experiments 4-6 show that it significantly loads factor 2 (.34) and factor 3 (.48). This is fairly consistent with the loadings found in Experiments 1-3, where the arrow span significantly loaded factor 2 (.57) and factor 3 (.33). The TTT span in Experiments 4-6 only significantly loads factor 2 (.61). This was not found in Experiments 1-3, where the TTT span only loaded factor 3 significantly (.60). If the arrow span was only a passive measure it would be expected to load factor 2 alone. If the TTT span was an active measure then it would be expected to load factor 3 alone. Due to the complex span measures (verbal and TTT) loading the specific factors more than factor 3, the Baddeley interpretation is not the best fit to the data. Thus, this pattern of results, again, might be more compatible with the notion that factor 2 is actually a domain-specific part of the CE, rather than being the VSSP component of working memory.

Next, the span measures will be discussed in relation to their R-squared values and the reliability estimates made for each of them. The R-squared values can be compared to those found in Experiments 1-3, where they ranged from .37 to .63. The R-squared values for the span measures in this series of experiments range from .34 to 1.00, where the only problem value is the one for the verbal span task. When error variance equals 0.00 ($r = 1.00$) the case

is called the Heywood effect (Harman, 1971) and this result could be explained as a chance finding in a relatively small sample (Dillon, Kumar & Mulani, 1987).

The simple tasks will be discussed next, as they were only loaded onto certain factors. The three-term spatial inferences, significantly load factors 2, 3 and 4 (.40, .65, .26, respectively). The abstract and conditional inferences significantly load factor 3 and factor 4 (.44, .55 for abstract and .34, .41 for conditionals, respectively). The consistent finding across all the simple reasoning tasks is that they all load the general and thinking style factors. An interpretation might be that all the simple tasks rely upon a general capacity working memory resource and in addition, upon the thinking style adopted by the reasoner.

The complex reasoning tasks were loaded onto all four factors. The temporal inferences significantly load factor 2 and factor 4 (.81, .55, respectively). The five-term spatial inferences significantly load factors 2, 3 and 4 (.50, .22, .23, respectively). In Experiments 1-3 the five-term spatial inferences loaded factor 1 and factor 3 significantly (-.60, .78, respectively). The overall finding for the complex spatial reasoning tasks is that they positively load the spatial, general and thinking style factors. An interpretation could be that these types of complex spatial task rely upon general and spatial executive resources of working memory, and additionally that the thinking style adopted by the reasoner also affects accurate performance.

The syllogistic inferences significantly load factors 1, 2 and 4 (.45, .67, .59, respectively), losing significance when loading factor 3. It should be noted that the loading of the syllogistic task on factor 1 was not found in the three-factor model. This is a major difference between the three-factor and four-factor model, the four-factor model findings

being more consistent with those found in Experiments 1-3. The multiple quantification task significantly loads factors 1 (-.23), 2 and 3 (.64, .34, respectively). This negative loading on factor 1 is of interest in relation to the loading found for the syllogistic task (.45). Both these tasks rely upon similar assumptions and reasoning about quantifiers. However, the multiple quantification task requires the processing of double quantified statements in relation to the syllogistic arguments. The negative loading of the multiple quantification task on factor 1 could be due to the fact that some reasoners rely upon verbal strategies in order to perform the task, which in the long run leads to poorer performance. However, the multiple quantification task also showed low reliability estimates in Experiments 5 and 6 and the findings should be taken in light of this consideration. The consistent finding for both the syllogistic and multiple quantification tasks is that they both significantly load the spatial factor, a finding consistent with Experiments 1-3. An interpretation might be that reasoning about quantifiers relies upon some sort of spatial executive resource of working memory in order for accurate responding.

The syllogistic and multiple quantification problems rest on similar assumptions about the relationships between quantifiers. From Experiments 5 and 6 correlational analyses it was shown that they relate positively ($r = .22$ and $r = .41$). However, the relationship between their loadings on the factors in this model are quite different. These differences could be due to the large R-squared value for the syllogistic task, or the low reliability of the multiple quantification task in both Experiments 5 and 6 (.28 and .31, respectively).

Interestingly, eight of the eleven tasks significantly load factor 3 (the general resource).

Also of interest, six of the seven reasoning tasks significantly load factor 4 (thinking disposition). This could be interpreted as showing that the majority of the span measures

and reasoning tasks rely upon a general capacity resource (Stanovich's cognitive ability) as well as the reasoning tasks being further affected by the thinking style adopted by the reasoner (Stanovich's thinking disposition).

Next, the reasoning measures will be discussed in relation to their R-squared values and the reliability estimates made for each of them. Comparisons with Experiments 1-3 are difficult here due to the fact that Experiments 5 and 6 used evaluation tasks, as opposed to production tasks. The R-squared values for the reasoning measures in this series of experiments range from .29 to 1.00, where the only problem values are for the temporal and syllogistic inferences. These problem values are again related to the Heywood case and may be explained by chance variation from sampling fluctuations. The reliability estimates for the reasoning measures used in Experiments 4-6 were variable, ranging from .28 to .84. Those with unsatisfactory reliability estimates were particularly the conditional and multiple quantification tasks. This is mirrored in the low R-squared value for the conditional inference task (.29).

4 . 5 . 3 Factor Analysis Discussion

Consistent with Experiments 1-3, the confirmatory multiple group four-factor analysis carried out for this series of experiments shows similar results. Due to the range of reasoning tasks used in these three experiments, it was possible to evaluate the influence of a fourth factor, thinking disposition, which did not emerge in the first three experiments.

The results here are consistent with the first three experiments in that some form of general factor loads all of the reasoning and capacity measures. This factor may be interpreted as a general working memory capacity resource which executes all higher level thinking, be it verbal or spatial in nature. In other words, this general factor could be interpreted as a low

demand processing resource, that deals with all information, be it span or reasoning measures. In contrast, the specific verbal and spatial factors could be interpreted as high demand processing resources, important in performing complex reasoning tasks. The data analysis and discussion below can not be explained by the Baddeley interpretation of a centrally controlled executive with two relatively passive storage systems (Baddeley, 1986). What this pattern of data does suggest is that the factors that emerged in all the CFA carried out on Experiments 1-6 support the notion of a general CE, with domain-specific components for verbal and spatial information. In addition to this finding, the present analysis adds that some form of thinking style influences the ability to reason deductively (Stanovich & West, 1997, 1998).

From Experiment 4-6 three-factor model it can be seen that the loadings were very similar to the four-factor model, with only a few discrepancies which did not change significances. With the additional χ^2 difference test and lower AIC value showing that the four-factor model was a better fit, the following discussion deals with this fourth factor as well, representing thinking dispositions. However, it should be noted that the three-factor and four-factor model can both be accepted as good fits to the data. In Experiment 1-3 CFA (Section 3.5) it was found that a three-factor model of working memory could account best for the pattern of data observed. What was not so clear was whether the factors represented a general executive with two relatively passive storage systems; or whether the factors could be interpreted as a general executive with two additional active components beyond the scope of the PL and VSSP.

An explanation of the span measure loadings in the four-factor model are considered next, together with comparisons with Experiments 1-3 CFA. The word span significantly loaded

factor 1 (verbal) and factor 3 (general), although highest on the general factor. These significant loadings are consistent with Experiments 1-3, where the same result was found. The verbal span loaded the verbal factor and the general factor significantly, although higher on factor 1. These findings, together with Experiments 1-3, suggest that the most plausible possibility is that factors 1, 2 and 3 represent specific components of the CE of working memory, as measured by the simple and complex span measures.

The arrow span significantly loaded factor 2 (spatial) and factor 3 (general). However, as did the word span, the arrow span loaded the general factor most significantly. These significant loadings were also found in Experiments 1-3, although the arrow span loading on the spatial factor was higher. The TTT span, on the other hand, loaded the spatial factor significantly, but not factor 3, the general resource. The reverse of this was found in Experiments 1-3, where the TTT span loaded the general factor and not the spatial one. An interpretation of these findings for both sets of experiments might be that instead of being representative of the VSSP, factor 2 is actually involved in both the storage and processing of spatial representations. As previously suggested, the original idea that the three factors might represent working memory resources from the Baddeley perspective are again questioned here. A more plausible explanation might be that factors 1, 2 and 3 are all active in nature, and represent specific components of the CE.

Considered next is an explanation of the reasoning task loadings in the four-factor model. From the abstract and conditional reasoning data, it can be seen that performance relied on the general working memory resource and was influenced by thinking style. The factor 3 loading is consistent with the knowledge that these types of reasoning task rely on a general CE resource (Toms, Morris & Ward, 1993; Klauer, Stegmaier and Meiser, 1997). The use

of different thinking styles, or motivation, can also account for the performance on these tasks. These simpler reasoning tasks appear to require central processing and be influenced by the thinking style adopted by the reasoner. Further factor 3 and factor 4 interpretations will be made after all the reasoning tasks have been summarised below.

From the three-term and five-term spatial reasoning data, it can be seen that performance relied on the general and the spatial resource of working memory, together with an influence of thinking style. The five-term series problems differ in this factor analytic investigation from that of the prior one (Experiments 1-3) due to the fact that it was an evaluative task here (with non-thematic content). This said, both types of spatial inference task used here require the use of a spatial resource in order to correctly respond (Farmer, Berman and Fletcher, 1986). As well, they use central resources, also seen in Experiments 1-3, in order to produce valid conclusions. Individual differences in thinking style can account for some of the variation in performance.

From the syllogistic reasoning data, it can be seen that performance relied on the verbal and spatial working memory resources, as well as having an influence of thinking style. This is consistent with Experiments 1-3 and recent work that suggests that syllogisms are solved by using spatial and/or propositional representations (Ford, 1994). As stated previously, one explanation may be that participants have to keep in mind the premise information, which logically relies on a verbal working memory resource, and then formulate a mental model that holds spatial (or propositional) information from the problem, which may rely on a spatial working memory resource. The addition of a thinking style influence may also account for variation in performance on the task.

From the multiple quantification reasoning data, it can be seen that performance relied on the spatial and general working memory resources, whilst being negatively related to the verbal resource. The negative verbal store loading was also found for the five-term series spatial inference when using a production task. As previously explained in the last factor analytic discussion (Section 3.5), the reason for the negative loading on the verbal store in both cases may be due to the use of a verbal strategy, rather than a spatial one. Phillips, Wynn, Gilhooly, Della Sala and Logie (1999) found that articulatory suppression increased accurate performance on the Tower of London task. This was interpreted as indicating that the PL resources were not available and reasoners were forced into using the VSSP resources, which produced better results. The reasoners in this set of experiments may have used verbal resources when in fact spatial resources would have achieved increased logical responses. This might explain the negative loading of the multiple quantification task on factor 1, although this finding must be taken in light of its low reliability estimates in both Experiments 5 and 6.

The loadings of the reasoning tasks on factor 4 show that all the measures load it significantly, except the multiple quantification task. This factor could be interpreted as a commonality between tasks that is independent of a general resource capacity (factor 3). Indeed, this is what was found by Stanovich and West (1998) who proposed that a combination of both cognitive ability and a reasoner's thinking disposition accounted for reasoning performance. Stanovich's explanation of this thinking style rests on the notion that the reasoner's performance might be influenced by their ability to search for counter-examples. One component of the mental models theory also relies on the assumption that the reasoner must search for alternatives where the conclusion does not hold (Johnson-Laird, 1983). The CFA from Experiments 4-6 shows a range of reasoning tasks load factor

4. which was interpreted as this thinking disposition of the reasoner. The only task not to load this factor was the multiple quantification task. A reason for this might be that it was the only task where the majority of the problems only required one model of the premises in order to correctly respond. Therefore, the search for counter-examples, as represented by factor 4, might not have improved the reasoner's performance.

The pattern of results from the measures used in Experiments 4-6 could be interpreted as the reasoning tasks drawing on CE resources. The CFA is most consistent with the notion that the CE component of working memory consists of specific modules which are involved in actively processing a variety of deductive tasks. The simpler reasoning tasks, that are assumed to need a low working memory load, are most associated with the general factor, whilst the complex reasoning tasks, assumed to need a high working memory load, are most associated with all the processing factors (factors 1-3). In addition to this mental workspace notion, Experiments 4-6 results support the idea that an individual's thinking style might affect the ability to reason deductively. This thinking style may represent the motivation a reasoner has for the searching for counter-examples. The following discussion expands on these possibilities, as well as discussing the findings in the context of reasoning theory.

4 . 6 GENERAL DISCUSSION FOR EXPERIMENTS 4-6

A central goal of the Experiments 4-6 was to demonstrate that individual differences in working memory capacity can explain variation in performance on a range of reasoning tasks. As in Experiments 1-3, the patterns of correlations do support this idea. Overall, the correlational data suggested that the simple deductive reasoning tasks were best predicted by simple capacity measures. This could be interpreted as the participant's reliance on a general executive resource that is independent of modality. Also, the correlational data

suggested that the complex deductive tasks were best predicted by the spatial capacity measures, although the magnitude of these was lower than seen in Experiments 1-3. Thus, it could be suggested that these complex deductive tasks require general and spatial resources capable of both storing (passive) and processing (active) the information. The factor analysis, consistent with Experiments 1-3, yielded results suggesting that a general factor of working memory capacity underlies all these deductive tasks. The factor analysis additionally yielded results suggesting that the reasoner's thinking style or disposition influences their performance on the deductive measures.

As in Experiments 1-3, this series of experiments also investigated the issue of the separability of central working memory resources for verbal and spatial information (Shah & Miyake, 1996). The patterns of correlations in Experiments 1-3 were inconsistent with the view that CE resources can be separated for verbal and spatial information. Experiments 4-6 correlations between the complex capacity measures ranged between .24 ($p < .10$) and .08 (In Experiments 1-3 the range was between .31 and .55). The correlations between the simple capacity measures ranged between .17 and .42 (In Experiments 1-3 the range was between .23 and .28). The Shah and Miyake (1996) interpretation and the Baddeley interpretation (CE, PL and VSSP) are not compatible with the pattern of CFA results found here. As in Experiments 1-3, the interpretation of the factor analysis suggests that the deductive tasks require a centrally executed working memory resource, but that some tasks additionally require verbal and spatial processing resources. The simple spans may be measuring part of the PL and VSSP components, in respect to storage requirements. However, more than these passive stores are required to explain the simple span loadings on factor 3 (a processing component). A possible interpretation from the results found for the

simple span measures is that the commonality between the two may be likely to represent general resources of working memory, requiring part storage and part processing.

Experiment 4 results will be discussed next, in relation to spatial and temporal thinking.

Past research has highlighted the similarity in the way reasoners deal with spatial as opposed to temporal information (Byrne & Johnson-Laird, 1989; Schaeken, Johnson-Laird and d'Ydewalle, 1996). These studies have shown, through dual-task methodology, that articulatory, visuo-spatial, and central executive interference tasks degrade logical performance on both spatial and temporal tasks. In Experiment 4, the correlation between the five-term series and temporal inferences was significant, suggesting common processes underlying both the measures. This is consistent with past research that has demonstrated a link between these two types of reasoning (Vandierendonck and De Vooght, 1997). The relationships between the spatial/temporal tasks with the working memory span measures was also investigated. Experiment 4 results showed that positive and significant correlations were obtained for both these tasks and all the working memory measures used. Experiment 4 results are also consistent with Experiment 2 where the five-term spatial inferences correlated positively with all the span measures. This is again consistent with past research, suggesting that both verbal, spatial and general resources of working memory are needed for logical performance (Vandierendonck and De Vooght, 1997).

Experiment 5 and Experiment 6 results will be discussed next in relation to the range of evaluative reasoning tasks used. It must be noted that any comparisons made between these two experiments and the previous four production type experiments must take account of these differences in methodology. These tasks will be split into simple (three-term series, abstract and conditionals) and complex (five-term series, syllogistic and multiple

quantification) groups. Past research has shown that simple deductive tasks can be separated due to their verbal or spatial nature. Farmer, Berman and Fletcher (1986), through dual-task methodology, found that performance on a simple spatial task was only affected by VSSP interference tasks; whilst performance on a simple verbal task was only affected by PL interference tasks. Gilhooly, Logie, Wetherick & Wynn (1993) found that interference tasks that are assumed to load the subsystem of working memory only had small effects on more complex deductive reasoning tasks, and that they were most affected by interference tasks assumed to load the CE component of working memory.

Experiment 5 and 6 showed that the simple reasoning tasks were positively, although not always significantly, correlated with the simple span measures. The abstract and conditional tasks, that rely upon reasoning with propositions, showed the strongest relationships with the word span in both Experiments 5 and 6 (.32 and .24 for abstract and .21 and .24 for conditionals). This would be expected considering the propositional nature of both these reasoning tasks and the word span task. The factor analysis loadings for these two reasoning tasks are less consistent with this view, where the only significant loadings were on the general factor and the thinking style factor. However, the word span loaded the general factor significantly too. A possibility might be that propositional deductive tasks rely upon general resources at the CE level. However, any interpretations made of the correlations and loadings for these tasks must be taken in light of the low reliability estimates obtained for both. The three-term series spatial inferences correlated significantly with the simple span measures, and positively with the complex span measures. The factor analysis results for this task showed significant loadings on the spatial, general and thinking style factors. This lends support to the notion of spatial processing of three-term series descriptions. However, again these resources might be at the level of the CE.

Experiment 5 and 6 showed that the complex reasoning tasks were positively, although not always significantly, correlated with the spatial span measures, both simple and complex. Performance on five-term series spatial inferences has been shown to be degraded by tasks assumed to load the PL, VSSP and CE components of working memory (Vandierendonck and De Vooght, 1997), although the major interference was seen when using CE secondary tasks. The factor analysis results for the five-term series descriptions showed that they significantly loaded the spatial, general and thinking style factors. This is consistent with this past research which has assumed that processing of these types of problem rely upon the manipulation of spatial models.

Performance on the syllogistic and multiple quantification tasks will be discussed together. This is because both types of problem rely upon reasoning about quantifiers and might be expected to show similar patterns of results. As with the complex spatial task, both these tasks appeared to have stronger correlational relationships with the spatial spans, both simple and complex. These correlations and part of the factor analysis results from Experiments 1-3 are also compatible. The consistency is the loading for both these tasks on the spatial factor, which is congruous with Experiments 1-3 findings for the syllogistic task. The syllogistic task was also shown to significantly load the verbal and thinking styles factors in this CFA. This is incompatible with Experiments 1-3 where the visually presented syllogisms loaded the general factor. The multiple quantification task was shown to significantly load the verbal (negatively) and general (positively) factors. It must be noted that the multiple quantification task produced low reliability estimates in both Experiments 5 and 6. The only consistent finding here is that both these quantifiable tasks relied upon spatial resources and this is in agreement with past research which suggests that reasoners might use different strategies based upon either spatial or verbal information

(Ford, 1994). These results are less consistent with the idea that general CE resources are required for logical performance (Gilhooly, Logie, Wetherick & Wynn, 1993), unless factor 1 and factor 2 are interpreted as domain-specific parts of the CE component of working memory.

Experiments 4-6 investigated a wide range of deductive measures, together with measures of working memory capacity. Thus, the findings here can be related to Stanovich and West's (1998) work where they suggested that measures of cognitive ability could predict individual differences in reasoning ability. The patterns of data, especially in Experiments 5 and 6, indicated that two clusters of correlations emerged. One cluster of correlations contained simpler reasoning tasks, whilst the other cluster contained the more complex tasks. Stanovich and West (1998) would expect a general positive matrix of results, where all the deductive tasks correlated together. The results of the correlational analysis here only partially support this notion. However, the factor analysis for Experiments 4-6 resulted in a four-factor model, where the majority of reasoning tasks loaded the general (capacity) and thinking disposition factors. This is consistent with Stanovich and West's (1998) claim that cognitive ability and thinking style can predict individual differences in deductive reasoning ability.

As briefly reviewed earlier, Stanovich and West (1998) found that measures of cognitive ability and thinking disposition correlated with and predicted performance on a wide range of tasks from the deductive and problem-solving literature. These findings lend support to the notion that a general capacity resource (Stanovich's cognitive ability) is needed in order to perform the deductive tasks used. In addition, one interpretation of the thinking style factor supports the notion that an individual's ability to search for counter-examples

influences their logical performance (Stanovich's thinking disposition). These interpretations will be discussed briefly below.

The general factor in both Experiments 1-3 and 4-6 CFAs could be interpreted as follows. It could be the case that this is not just an executive memory resource, due to the loadings of the simple reasoning tasks. Because the simple reasoning tasks only loaded the general factor, it might be that this factor represents decontextualisation of the reasoning process. It might reflect the participant's general control processes for dealing with different tasks. For instance, reasoning task performance relies to some extent on inhibiting responses from the pragmatic system in order to produce or evaluate conclusions. Therefore, this factor could reflect the attentional system needed to control and switch responses from one system to another. However, the span measures also loaded this factor. This is also consistent with the notion that this factor is an attentional controller. For instance, the complex verbal span requires participants to switch between processing sentences, and remembering words. The general factor might be the attentional system that allows this constant switching of particular processes. A more direct measure of the capacity of working memory can be seen in the loadings on the specific verbal and spatial factors. These factors were most identified with the span tasks and the complex reasoning tasks. This could be interpreted as the complex reasoning tasks additional reliance on capacity systems able to store and process modality dependent information.

The fourth factor, again, might not be related to capacity, but to the participant's thinking style. For instance, the conditional task (for AC and DA) requires the reasoner to resist the fallacious arguments and respond with 'no valid conclusion'. This relies on recognising the fact that alternatives might exist where the antecedent does not hold, but the consequent

does, and this might involve searching for other explanations. The other reasoning tasks could all be split into single and multiple model problems. Thus, it might be that they all relied on the individual's ability to search for counter-examples. In addition to this, the fact that the multiple quantification task, where most of the problems were based on one model of the premises, did not load this factor, further substantiates the interpretation of factor 4. In this task, participants may not have needed the motivation to search for alternative models of the conclusion, and thus were not affected by this fourth factor.

The results of Experiments 4-6 also have implications for theories of deductive reasoning. What is clear throughout this experimental series is the reliance of many of the reasoning tasks on spatial resources of working memory. The first class of theory introduced in Chapter 2 was loosely labelled as the verbal account of the reasoning process. These theories proposed that propositional rule-based or verbal comprehension processes could account for variation seen in performance of deductive reasoning tasks (e.g., Rips, 1994; Polk & Newell, 1995). The results here do not conform to this picture, showing many significant correlations between spatial working memory and the reasoning tasks, as well as links to a spatial factor (factor 2). The second class of theory introduced in Chapter 2 related to spatial accounts of the reasoning process. These theories proposed that performance on deductive reasoning problems involves the manipulation of mental models (e.g., Johnson-Laird, 1983). The results of Experiments 4-6 are in accordance with this view due to the relationships with spatial resources, although the syllogistic reasoning task was also positively loaded on the verbal resource. Due to the emergence of a general factor in the factor analysis, together with positively associated active/complex capacity measures of working memory in the correlational analysis, claims about processing theories of reasoning can not be made clearly. However, what can be claimed is that the span

measures, presumed to measure the capacity of the executive system, did predict individual differences in performance on the deductive tasks.

The heuristic accounts of syllogistic reasoning do not draw on the notion of an explicit processing system, and would not have predicted this relationship between capacity and reasoning. This was the third class of theory introduced in Chapter 2. These accounts rely on the idea of implicit processing, where rules of thumb prevail. For instance, Wetherick and Gilhooly (1990) proposed that people endorse conclusions whose mood is similar to the premises. Other noninferential theories, such as Chater and Oaksford's (1999) PHM, also proposed that syllogistic data can be accounted for in terms of heuristic decision processes related to the information conveyed by the premises and conclusions. The patterns of results found in all six experiments showed that individual differences in reasoning ability can be predicted by working memory capacity, as measured by the span tasks. These accounts must be able to explain and account for the relationship between explicit processing and reasoning ability.

The findings from Experiments 4-6 correlational analysis and the CFA can be summarised as follows. The CFA identified four factors which influenced the ability to reason deductively. Factor 1 and factor 2 were interpreted as requiring executive capacity for processing and storing verbal and spatial information, respectively. Factor 3 was interpreted as representing a general attentional component, where switching and inhibiting responses, as well as a general capacity, were important. The fourth factor that emerged was identified as the thinking disposition of the reasoner. Thus, the results showed that individual differences in reasoning ability can be predicted by a general attentional resource, plus additional verbal and spatial resources of working memory, each of which relies on more

than just passive storage. In addition to these factors, the motivation of the reasoner to search for counter-examples was also shown to affect the reasoning process (factor 4). In the chapter that follows, the findings from both Chapter 3 and Chapter 4 will be considered together. The interpretation of the CFAs will be extended to consider the best way to account for the four factors. Comparisons will also be made with models from other areas of study, such as intelligence and executive function.

CHAPTER 5 - GENERAL DISCUSSION

5 . 1 INTRODUCTION

The first aim of this chapter is to summarise all the findings from Experiments 1-6, and the interpretations made from these. Chapter 5 is then separated into a further six sections. The first of these sections considers models of working memory in light of the experimental findings. The second section goes on to examine alternative ways of looking at this data through the tradition of intelligence research. Models of intelligence are compared to the data from this series of experiments. Considered then is the relation of these experimental findings to theories of reasoning, with a particular focus on the interpretation of the fourth factor identified in the CFA. A brief section examining the use of individual differences methodology will then be given, looking at both advantages and disadvantages of this approach. Future research directions will then be examined before the final concluding comments.

5 . 2 SUMMARY OF EXPERIMENTAL FINDINGS

The first experimental chapter was inspired by the work of Shah and Miyake (1996). From their research they concluded that central executive resources of working memory could be separated for language based and spatially based processing. Experiments 1-3 examined the extent to which different reasoning tasks draw on these distinct systems. The first three experiments were additionally designed to look at the role of working memory components in syllogistic and spatial inference tasks. To test these claims, a range of simple and complex verbal and spatial working memory span measures were used, together with syllogistic and five-term series spatial inferences. In contrast to the claims of Shah and Miyake (1996), the complex verbal and spatial working memory span measures did not

show a clear dissociation. Additionally, the simple verbal and spatial span measures did not show total independence. In terms of the reasoning tasks, the clear general finding was that individual differences in working memory capacity could account for individual differences on them both.

A CFA three-factor model was identified to best account for the data from all three experiments, and the following discussion focuses on the interpretation from this. The three working memory resource factors in Experiments 1-3 CFA were labelled as follows. Factor 1 was classified as a verbal resource, factor 2 as a spatial resource, and factor 3 as a general resource. The CFA identified the simple word span and the complex verbal span with the verbal and the general factors. The simple arrow span and the complex spatial spans were identified with the spatial and the general factors. The reasoning tasks were allowed to load all three factors. The interpretation that was most favoured for these findings was in support of a unitary view of executive function, where the complex and simple spans load a general resource, and additionally load specific verbal and spatial resources. The word span and the verbal span loaded to similar degrees on both the specific verbal and general factors.

Correspondingly, the arrow span and the complex spatial spans loaded to similar degrees on both the spatial and the general factors. Due to the significant loadings of both simple and complex span tasks on these factors, it was assumed that all three relied to some degree on processing as a requirement, and that these spans were measuring this ability to process and store information. The simple word span and complex verbal span might reflect the measurement of both processing and storage resources of verbal working memory. One possible interpretation, consistent with Baddeley's working memory model, is that they reflect passive storage and active CE involvement. These loadings could be explained by recourse to partly PL and partly CE resources. The simple arrow and complex TTT and

letter spans might reflect the measurement of both processing and storage resources of spatial working memory. These loadings could be explained by recourse to partly VSSP and partly CE resources. However, the simple spans could not be specific measures of storage requirements, and the complex spans could not be specific measures of processing requirements, due to the loadings explained above. Therefore, the pattern of loadings favour an alternative interpretation, where the verbal and spatial factors reflect more than simple storage, suggesting the separation of verbal and spatial information in the CE, as well as there being a general component. The general factor, now characterised as domain independent, will be discussed later on in terms of a more detailed interpretation.

The syllogistic inferences, whether verbally or visually presented, correlated with both complex verbal and spatial working memory spans. In the CFA, they were shown to load all three factors, regardless of modality. Given the interpretation presented above, this finding demonstrated that syllogistic inference performance depends upon both verbal and spatial resources of working memory. Of particular interest, in the context of theoretical accounts of syllogistic reasoning performance, there was clear evidence for a role of spatial working memory.

The spatial inferences, whether verbally or visually presented, also correlated with both complex verbal and spatial working memory spans. In the CFA, they were shown to load the general factor, regardless of modality. This finding demonstrated that spatial reasoning performance depends upon a general working memory capacity. There was clear evidence for a role of general working memory in the five-term series spatial inference task.

Better performance on both the syllogistic and spatial inferences was shown when in the visual modality, and on single model problems. The verbally presented reasoning tasks were suggested to rely heavily on verbal working memory capacity. By presenting the problems visually, the verbal working memory load was reduced, and logical performance rates increased.

The second experimental chapter was inspired by the work of Stanovich and colleagues (e.g., Stanovich & West, 1998). From their research they demonstrated that a range of reasoning tasks, drawn from both deduction and decision making literatures, cross correlated. They also showed that these reasoning tasks were predicted by the combined influence of cognitive ability and independently with the reasoner's thinking disposition. In Stanovich's terms, cognitive ability equals capacity, and variations in that capacity explained logical performance on the reasoning tasks. Experiments 4-6 were primarily designed to extend Stanovich's work to a wider range of reasoning tasks, using a more direct measure of capacity.

The reasoning tasks, classified as simple, were the abstract, conditional and three-term series spatial inferences. The correlational analysis identified that they were best predicted by simple working memory spans. The reasoning tasks, classified as complex, were the five-term series spatial and temporal inferences, plus the syllogistic and multiple quantification problems. The correlational data suggested that they were best predicted by complex working memory spans, especially spatial ones. Two clusters of correlations were identified, where the simple reasoning tasks cross correlated, and the complex reasoning tasks cross correlated. Therefore, what might underlie performance on the simple reasoning tasks may be different to what underlies performance on the complex reasoning tasks.

Consequently, these results only partially support Stanovich's claims. However, the CFA identified a four-factor model to best account for the results. Additionally, to the three factors identified in Experiments 1-3, Experiments 4-6 CFA identified a fourth factor, which was classified as a thinking style influence, and this will be interpreted in Section 5.5.1.

The CFA in Experiments 4-6 was consistent with that seen in Experiment 1-3, in relation to the span measures. The simple and complex verbal and spatial spans showed similar loadings to those obtained in the first CFA. Again, these results suggested that the factors could represent both general and specific verbal and spatial resources at the level of a CE - requiring both processing and storage of information, and that the simple and complex spans were also measuring these resources. Experiments 4-6 CFA also identified a fourth factor, linked to the reasoning tasks, and this was classified as the reasoner's style of thinking.

The five-term series and temporal problems showed similar correlational patterns, both complex and simple verbal and spatial spans were shown to be related to their performance. The CFA identified the spatial factor as most important in both these types of reasoning process, although some influence could be seen from the general and the thinking style factors. The result for the spatial task is different to that seen in Experiments 1-3, where they only loaded the general factor. However, they do consistently correlate with the complex spatial span tasks throughout the experimental series. The complex spatial and temporal tasks were considered to rely upon spatial working memory resources and to be affected by the reasoner's thinking disposition.

The reasoning tasks, classified as simple in Experiments 5 and 6, showed stronger correlational relationships with the simple span measures than with the complex span

measures. The reasoning tasks, classified as complex ones, showed stronger correlational relationships with the complex span measures than with the simple span measures.

However, the CFA identified that five of the seven reasoning tasks loaded the general factor and six of the seven reasoning tasks loaded the thinking style factor. These results were taken as support for a person's reliance on a domain independent CE of working memory, and that their style of thinking was also a significant predictor of their ability to reason.

Experiments 1-6 findings have strong implications for a number of distinct bodies of research within the psychological literature. The results can be applied to the areas of working memory, intelligence, and reasoning theory. The following three sections discuss these areas in turn. In addition, the results from Experiments 1-6 have implications for the alternative research methodologies of dual-task versus individual differences work, to be discussed later on in Section 5.6. The next section considers the findings in terms of existing models of working memory.

5.3 MODELS OF WORKING MEMORY

Overall, the experimental series very clearly demonstrated a role for working memory in reasoning. Individual differences in a reasoner's working memory capacity were related to individual differences in their reasoning performance. Three capacity resources were identified through CFA (factors 1, 2 and 3). The factors were all suggested to rely on storage and processing requirements - one for a verbal, one for a spatial, and one for a general resource, as measured by the simple and complex span measures. Whilst the verbal and spatial factors were interpreted as specifically dealing with verbal and spatial information, respectively, the general factor was interpreted as handling information from both these domains, capable of dealing with information using domain independent

representations. The following discussion relates to the explanation that the general resource is a domain independent processing resource, whilst the other two resources are domain specific processing components.

This thesis has concentrated on differentiating between a Baddeley (1997) perspective and a Shah and Miyake (1996) perspective of working memory. The Baddeley model proposes a general CE component with two relatively passive subsystems, the PL and VSSP, which are used in higher level cognitive thinking. The Shah and Miyake model suggests a separation of CE resources for verbal and spatial information, which are important in the ability to reason. This section aims to systematically compare and contrast these models with other existing models from the literature. The question is to what extent are working memory models compatible with the findings from this thesis?

There are at least two models of working memory that can be ruled out. The first is a model consistent with a single CE resource of working memory. Kyllonen and Christal (1990) suggested that reasoning ability was little more than working memory capacity, and a unitary resource of CE working memory could be accepted. They found that working memory test scores correlated significantly with higher-order ability scores even when the broad domains of these predictor and criterion tests did not match. The results here support the idea that working memory is a major factor in determining reasoning performance.

Working memory capacity consistently predicted deductive reasoning performance through Experiments 1 to 6. However, the unitary model can not be reconciled with the data here for two reasons. The first is that two clusters of reasoning task were identified in Experiments 5 and 6. The simple reasoning tasks mainly correlated with each other, and the complex reasoning tasks mainly correlated with each other. All the reasoning tasks would

have had to show similar relationships with each other to accept that they draw on one pool of general resources. The second reason for a rejection of this unitary view is due to the separation of simple and complex verbal and spatial spans across the three factors in the CFA. If the single executive perspective was to be accepted then there should have been no need for additional specific verbal and spatial factors, over and above the general resource.

The second model that can be rejected as incompatible with the results from Experiments 1-6 is one based on two separate but independent CEs, one for language based and one for spatially based information. Shah and Miyake (1996) found that working memory span tasks and higher-order ability tasks correlated more strongly when their domains matched. They also found a dissociation between language based and spatially based domains, indicating the necessity to separate these specific executive functions. The results from Experiments 1-6 did not show a clear dissociation between verbal and spatial working memory spans, or their relationships with verbally or spatially based deductive tasks. The Shah and Miyake (1996) uncorrelated two-factor model was not supported by the data here. However, a variant of Shah and Miyake's (1996) model in which the verbal and spatial factors were allowed to correlate might be substantiated (Section 3.5). Consequently, when Shah and Miyake's (1996) data was re-run as a three-factor model, the results were consistent with the addition of a third factor to explain the underlying commonality between the initial independent factors (the correlation).

The model that has already been used as a perspective throughout this thesis - the multiple component model of working memory (Baddeley, 1993), cannot be rejected. Baddeley's model rests on the assumption of a general CE with two slave systems for verbal (PL) and spatial (VSSP) information. The three factors identified in the CFA loaded tasks requiring

both the processing and storage of information. Due to the consistent loadings of the simple and complex verbal and spatial working memory spans on the three factors, it was concluded that individual differences in reasoning are influenced by domain specific components of the CE - verbal, spatial and general. This does not mean a rejection of the PL and VSSP components of the Baddeley working memory model. However, a particular interpretation of the three factors in terms of the CE, PL and VSSP, can be discounted. Consequently, the unitary view of CE function can also be discounted due to the spatial and verbal factors being best interpreted, not as passive storage systems, but more as active processing systems specifically associated with particular modalities.

The span measures that were labelled as simple measures in Experiments 1-6 were the word and arrow spans. These tasks require storage resources for correct responding, and originally this was the only criterion for classification. Due to the findings of the CFAs it was acknowledged that they may depend upon some processing component as well, since they loaded similarly to the complex span tasks. Participants might develop strategies to help them hold the information in mind. For instance, the arrow span could be solved by using a clock-face strategy which may have added a processing load to this task. Similarly for the word span, it was observed that some participants used the first letter of each word in a set as a cue to recalling them later. This, too, could have entailed the addition of a processing load to this task. The span measures that were labelled as complex measures in Experiments 1-6 were the verbal, TTT and letter spans. These tasks all required the storage and processing of information. Therefore, it might be said that, although the word and arrow span tasks were simpler (only arrows or words to be remembered), they too required processing resources, reflected in their loadings on the factors that were similar to the complex span tasks.

The best way to account for the three factors identified throughout the CFAs was as a general domain independent CE resource, with additional CE resources for verbal and spatial information specifically. These resources depend on both processing and storage requirements, as measured by the complex and simple spans. The verbal factor was assumed to be responsible for the capacity to store, manipulate and process verbal representations. In contrast, the spatial factor was assumed to be responsible for the capacity to store, manipulate and process spatial representations. The general factor was assumed to reflect the capacity to process information in a domain independent resource. The characteristics of the general factor will be discussed in more detail in the reasoning section. Accumulating across the six studies, there is strong evidence to suggest that the unitary view, and the Shah and Miyake (1996) perspective, are not compatible with the patterns of findings across these data sets.

The original Baddeley and Hitch (1974) model has more recently been updated, especially in relation to the functions of the CE (Baddeley & Logie, 1999). The original model assumed that the CE comprised a pool of general-purpose processing capacity that could be used to support either control processes or supplementary storage. Baddeley and Logie (1999) subsequently abandoned the assumption that the CE itself stores information, proposing instead that any increase in total storage capacity beyond that given by a slave system is achieved by accessing either long-term memory or other sub-systems. This notion could be supported by the CFA carried out on Experiments 1-6. It might be that the general factor relates to the CE as a processor, whilst the specific verbal and spatial factors could be interpreted as 'other sub-systems' in the context of Baddeley and Logie (1999).

A similar account to the one presented here is the one proposed by Engle, Kane and Tuholski (1999). They were guided by two questions in their pursuit of the role of individual differences in working memory capacity and cognition. The first question was what is measured by the complex spans that is also important to high-level cognitive tasks? The second question was what do the results of studies in individual differences in working memory capacity tell us about the nature of working memory in general? They proposed that individual differences on measures of working memory capacity primarily reflect differences in capability for controlled processing and, thus, will be reflected only in situations that either encourage or demand controlled attention (e.g., Rosen & Engle, 1997). They go on to say that the working memory/attention system is probably neither entirely unitary nor entirely separable into domain-specific systems, and this is entirely consistent with the CFA results found in Experiments 1-6. Rosen and Engle (1997) also suggested that working memory/attention may be organised similarly to intelligence (e.g., Carroll, 1993; Kyllonen, 1996, see next section). This is explained as a hierarchical structure with a domain-free factor overarching several subordinate domain-specific factors. As in intelligence research, a general factor appears to account for too much variance to be ignored. However, in some studies (Shah & Miyake, 1996) significant variance is left to be explained beyond that accounted for by a general factor. In the next section these models of intelligence will be considered.

5 . 4 INTELLIGENCE

The unitary versus non-unitary view of working memory has an interesting historical parallel in the domain of intelligence. Spearman (1904, 1927) factor-analysed the results of children's performance on various tests and found that many tests were moderately positively correlated, concluding that all the tests had something in common (a general

factor) as well as something specific to each test (a specific factor). Vernon (1947) elaborated and extended Spearman's model by identifying a series of group factors (major and minor) in between g and s factors. Vernon studied the structure of human cognitive abilities through a hierarchical group factor theory. Thirteen tests were given to 1000 Army recruits. He identified g as accounting for what all the tests are measuring, and the major group factors (verbal-education and spatial-mechanical abilities) as what some tests are measuring. The minor group factors were identified as what particular tests are measuring whenever they are given, while specific factors were identified as what particular tests measure on specific occasions. This was suggested to provide evidence that good predictions of ability in education, industry or everyday life, can be achieved by g tests alone, and that somewhat more ground can be covered by tests of the main group factors. More recent models of abilities (e.g., Carroll, 1993) are very much compatible with Vernon's hierarchical treatment.

The hierarchical model of Vernon (1947) can be compared to the findings of the two CFAs in this experimental series, when Experiments 1-3 and 4-6 are looked at in combination. A general factor was identified in this experimental series onto which 11 out of 12 of the capacity and reasoning measures loaded significantly and positively. This might represent a general capacity CE resource, although it could also be interpreted as what all the tasks are measuring - identified by Vernon as g, and by Stanovich as cognitive ability. Factor 1 was identified as a verbal resource, capable of the maintenance and processing of verbal information. This could be compared to Vernon's verbal-education factor, reliant on verbal, numerical or educational abilities. Factor 2 was identified as a spatial resource, capable of the maintenance and storage of spatial information. This could be compared to Vernon's spatial-mechanical factor, reliant on practical, mechanical, spatial and physical abilities.

Experiments 1-6 demonstrated that individual differences in working memory capacity could consistently predict individual differences in reasoning performance. Kyllonen and Christal (1990) also demonstrated that reasoning is little more than working memory. Before this series of experiments, especially from the individual differences perspective, the relationship between working memory capacity and syllogistic and spatial inference tasks was not well established. Kyllonen and Christal (1990) used working memory capacity tests and suggested that the capacity measures themselves load highly on g. Stanovich and West (1998) found that measures of general intelligence (cognitive ability) could predict reasoning performance. What this thesis did was pull this research together, providing evidence that Stanovich's cognitive capacity could be related to working memory capacity. The experimental series also clarified the appropriateness of applying Kyllonen's findings to those of Stanovich.

Stanovich's results, and the findings from Experiments 1-6 can be related to the dual process theory of reasoning (e.g., Wason & Evans, 1975). This theory proposed that the ability to demonstrate logical competence reflects the product of an explicit logical system, whereas beliefs implicitly influence this process. Another dual distinction was made between two different notions of intelligence. The first is universal and embedded in the implicit cognitive system, and has been termed rationality 1. The second involves the explicit reasoning system and corresponds to rationality 2. This second type of rationality involves individual differences in g, whereas the first type of rationality does not. If competence, when motivated, is ultimately limited by working memory capacity then this links neatly with recent evidence that such capacity is itself highly loaded on g (Kyllonen & Christal, 1990).

5 . 5 REASONING THEORY

Throughout this thesis, reference has been made to the various reasoning theories and how they might account for the findings here. Three classes of theory were introduced - two classes were related to processing accounts, the verbal versus analogical, and the third to heuristic accounts of reasoning performance. The findings in Experiments 1-6 can be related to these theories. The accounts must be able to explain why individual differences in working memory capacity predict individual differences in reasoning ability. The initial discussion below looks at patterns of loadings on the three working memory resource factors - verbal, spatial and general. The section that follows (Section 5.5.1) looks at the fourth factor identified in Experiments 4-6 in relation to the reasoning tasks, with particular attention to the work of Stanovich and cognitive thinking styles.

The first class of reasoning theory loosely fitted into those relying on verbal processes, verbal comprehension, or rule-based accounts. The formal inference rule theory proposed syntactic conceptions of logic that use proof-theoretic methods, based on deriving valid conclusions by means of rules of inference (Rips, 1994). Formal rule theorists maintain that the human being has within the mind a set of rules or inferential schemes similar to those of logic, upon which answers to the deductive problems are derived. The reasoner's task can be broken down into three distinct phases - discovering the abstract form of the premises, applying the formal rules that lead to a valid inference, and re-translating the conclusion into a specific problem-content (Evans, Newstead & Byrne, 1993). The larger the set of rules required in the process of deriving a conclusion, the greater the problem's difficulty.

The second class of reasoning theory loosely fitted into those relying on analogical models to account for the reasoning data. The mental models theory holds that the basic processes

of human reasoning competence are based on the construction and manipulation of mental models (Johnson-Laird, 1983). The mental model theory approaches reasoning in a clearly distinct way. There are three stages of deduction according to the model theory: comprehension, description, and validation. Comprehension refers to reasoners applying their knowledge to constructing models that represent the state of affairs described in the premises. Description refers to the fact that reasoners try to formulate a parsimonious conclusion that accounts for the models constructed in the prior stage. Validation refers to how they try to falsify their previous conclusion by searching for alternative models or counter-examples (see Johnson-Laird & Byrne, 1991). A prime assumption of the mental model theory is that given the limited nature of human working memory resources, people try to represent explicitly as little information as possible - the more information represented explicitly, the greater the load on the working memory.

Comparisons can be made between the results in this thesis and both the verbal and analogical based accounts in relation to the use of an explicit processing system (rationality 2). All the processing accounts rely on the notion of a mental workspace which people use in order to perform mental deduction. Experiments 1-6 showed correlations between the working memory capacity tests and the reasoning measures. Thus, the results from the experimental series showed that recourse to an explicit system could indeed account for variations in performance on a variety of reasoning tasks. Both the verbal and model based theories would expect this correlation between capacity and reasoning ability, due to the individual's reliance on underlying rules, or the construction of models, respectively. These theories, and this experimental series, implicate a role for an explicit processing system in reasoning. What is less clear from the results here is whether reasoning performance relied upon verbal/propositional based or spatial/analogical based processes. This is due to the

lack of dissociation between complex working memory capacity measures, making these conclusions difficult or ambiguous.

Having said this, however, the clearest results throughout the thesis were in relation to the syllogistic inferences. It was found that the syllogisms consistently correlated with spatial working memory, as defined by the spatial factor. The syllogistic task loaded all three of the working memory resource factors and was therefore assumed to rely on both verbal and spatial processing. The multiple quantification task also involves understanding the quantifiers in the two premises. The multiple quantification task also loaded on the spatial factor. This finding suggests that deterministic verbal accounts of the reasoning process are misguided. The results here seem unlikely to be accounted for solely on verbally based processes. Perhaps, instead, different people use different strategies, some using verbal, and others using spatial heuristics.

In relation to the spatial inferences, they correlated with both verbal and spatial working memory spans, but only positively and significantly loaded the general factor (Experiments 1-3) and the spatial factor (Experiments 4-6). The spatial inferences also correlated highly with the temporal inferences, which is in support of recent research suggesting they are solved using similar spatial processing (e.g., Schaeken, Johnson-Laird & d'Ydewalle, 1996). The strongest predictors of both the spatial and temporal tasks were the spatial spans, which are dependent on the storage and manipulation of spatial information. This result again has implications for the verbal based accounts of reasoning. However, these implications must be made in light of the cross correlations between the complex capacity measures. It seems likely that people use spatial representations in order to solve these problems, although this processing might be achieved using a CE component of working

memory (spatial and general resources). Thus, these results seem more consistent with an account based on spatial mental models.

Before moving on to the evaluation experiments (5 and 6), a more detailed interpretation will be given in relation to the general factor. It was identified as an all purpose capacity resource, capable of processing deductive arguments in a domain independent way. In relation to the work of Stanovich, the general factor could reflect the ability of the participants to resist pre-potent responses cued by the pragmatic system. This ability seems to be important to all the reasoning tasks, reflected in the fact that nearly all of them loaded this third, general factor. For instance, syllogistic reasoning involves resisting conversational inferences. Spatial reasoning involves recognising that relational terms describe deterministic, constrained relationships. The general factor might reflect the extent to which people can de-contextualise their reasoning processes. In addition, due to the loadings of the span tasks on this factor, it might lend support to the idea that it is more of an attentional controller, than a pure memory resource. The general factor might allow the switching of processes from one component of a task to another - the larger this resource, the more capacity is available in order to control these processes.

The discussion now turns to the tasks used in Experiments 5 and 6, the evaluation experiments. The three-term series spatial task loaded the spatial and general resources significantly. This links neatly with the result found for the five-term series spatial task. It seems that, for all types of spatial reasoning task examined in this thesis, spatial and general working memory resources are implicated over and above verbal resources. Thus, an interpretation might be that logical performance on the spatially based reasoning tasks relied to some extent on the manipulation of spatial mental models.

The only reasoning tasks not to load on the specific factors, verbal and spatial, were the abstract and conditional tasks. These tasks only significantly loaded the general working memory resource factor. Both these tasks rely on the manipulation of propositions, as opposed to the spatial and quantified tasks outlined above. Taking the view that the general factor reflects the participant's ability to de-contextualise their reasoning processes, this finding is very interesting. It could be that the propositional based tasks do not load the verbal and spatial factors because the memory load involved in them is quite small (classified as simple tasks). The reason for their loading on the general factor might be because accurate performance entails resisting pragmatic inferences. For instance, for the conditional task, recognising that 'if p then q' does not imply 'if q then p' entails resisting conversationally invited inferences (e.g., Geis & Zwicky, 1971). Therefore, the only consistent finding for the propositional tasks is that they rely on a general executive resource, at the level of the CE, and, consequently, differentiation between verbal and analogical accounts of the reasoning process is difficult.

Now that the processing accounts have been examined, this discussion turns to the third class of reasoning theory, the heuristic accounts. Noninferential accounts include the atmosphere theory of syllogistic reasoning which proposes that reasoners select conclusions on the basis of the mood of the premises (Woodworth & Sells, 1935), or the matching hypothesis (Wetherick & Gilhooly, 1990), which suggests that responses are elicited which match the quantified form of one of the premises. A more recent version of these noninferential accounts is grounded in information theory, and proposes that reasoners select conclusions that match the least informative syllogistic premise (Chater & Oaksford, 1999). These heuristic accounts assign a very limited role for explicit inferential processing in syllogistic reasoning.

Taking the Chater and Oaksford (1999) theory a little further, they base their heuristics on probability. They associate these heuristics with an implicit processing system. Therefore, there might be no reason to expect that measures of an explicit system (working memory capacity) predict logical performance on the task. They suggest that people do not generate proofs, or construct mental models, instead they rely on rules of thumb, and 95% of their data was predicted by these simple heuristics. Experiments 1-6 showed that efficiency in the explicit system, as measured by the working memory capacity tasks, could predict performance. This finding could not be explained if simple heuristics were the dominant processes involved in performing deductive reasoning tasks.

As briefly mentioned in the intelligence section, Evans and Over (1996) have explored the notion of two types of rationality. These rationality types have been linked to two specific processing systems. Rationality 1 refers to reasoning in such a way as to achieve one's goals, and requires implicit processing. Rationality 2 refers to reasoning by a process of logic, and requires explicit processing. It seems that the heuristic accounts rely on the notion of rationality 1. Experiments 1-6 demonstrated that logical competence could be reflected as a product of an explicit system, that rests on the assumptions of rationality 2. It seems that the heuristic theories propose a single process account, and explain performance with recourse to implicit processes. What the heuristic accounts fail to explain is that measures of the capacity of the explicit system are able to predict logical reasoning performance. These accounts explain variation in logical reasoning performance by recourse to a goal directed mechanism, or implicit processing. This thesis has been able to demonstrate that capacity measures, which reflect the efficiency of the explicit system, predict reasoning performance. The heuristic accounts are therefore challenged to explain

why deductive reasoning performance can be explained in reference to the capacity of explicit processing systems.

5.5.1 A Fourth Factor?

This discussion so far has concentrated on three of the factors identified in Experiments 1-6 CFAs, that related to the capacity resources of working memory. Experiment 4-6 CFA additionally uncovered another factor which systematically predicted performance on the range of reasoning tasks used. This fourth factor that emerged was interpreted similarly to the predictor identified by Stanovich and West (1998) as a thinking style influence.

Thinking style can be related to Johnson-Laird and Byrne's (1993) notion that people are programmed to accept inferences as valid provided that they have constructed no mental model of the premises that contradict the inference. Inferences are categorised as false when a mental model is discovered that is contradictory, however, the search for these models is not governed by any systematic or comprehensive principles. Consequently, it might be that related cognitive dispositions may in fact be reflecting this kind of process. Stanovich and West (1998) suggested that individual differences in the extensiveness of the search for contradictory models could arise from a variety of cognitive factors such as cognitive confidence, reflectivity, need for cognition, ideational generativity, dispositions toward confirmation bias and premature closure.

The fourth factor identified in Experiments 4-6 CFA might reflect similar processes to the thinking dispositions identified by Stanovich and West (1998). They found a combined influence of cognitive ability and thinking disposition affected the deductive process. The CFA was most compatible with the data sets only when this extra factor loaded the deductive tasks. Therefore, it might be that something, independent of capacity, accounted

for an individual's reasoning performance. This factor did not emerge in the first CFA carried out on Experiments 1-3. The reason for this could have been the limited number of deductive tasks used in this experimental series. When additional deductive tasks were introduced in Experiments 4-6, this fourth factor was important in accounting for performance on the tasks. Thus, there may be individual differences in cognitive style, in addition to limitations in cognitive capacity, that influence the way the individual reasons. A possibility could be that the production of alternatives correlates with reasoning performance, and those individuals with higher intellectual ability will presumably have more capacity to generate alternatives. Additionally, cognitive styles determines the extent to which the reasoner is motivated to search for these alternatives. All the reasoning tasks, bar one, included multiple model problems. For instance, conditional reasoning requires recognising that alternative antecedents are possible. The only task that consisted of mainly one model problems was the multiple quantification task, which, interestingly did not load the fourth factor significantly. This lends support to the notion that this factor entails searching for alternatives.

5 . 6 METHODOLOGICAL IMPLICATIONS

The method of choice throughout this thesis has been the individual differences approach. The individual differences methodology essentially aims to discover which test items correlate with one another and which do not and then to identify the resulting correlation clusters (or factors). The use of CFA is based on correlation, and attempts to reduce a large amount of data to a much smaller amount. This method allowed the identification of a number of factors (four) in this series of experiments that could account for the resources required to perform deductive reasoning tasks. The use of this approach through Experiments 1 to 6 led to the clear demonstration of a role of working memory in reasoning.

In contrast, the experimental approach (dual task work) has not provided as clear a message about the role of CE function in reasoning.

Despite the clear results found when using this method, there are limitations to the individual differences approach. One of its major limitations is its inability to deal readily with individual variation in the way different abilities are used in the performance of tasks. The usual assumption in this regard is that the abilities required for a task are the same for all individuals in a sample. However, in general, the limitations are not as serious as they are often thought to be. Many of the difficulties with the individual differences approach have come about because of the inappropriate design and selection of the tests, improper design of factor-analytic studies, unwisely chosen methods of computational analysis, and unfortunate interpretations of results (Carroll, 1978a). Frequently, re-analysis of data from different studies can lead to much more convergence and agreement in the findings than has appeared previously. Different factorial models of mental abilities can be translated into each other, and a reasonable choice among models and methods can be made by the application of a number of rules of parsimony. Multiple group CFAs entail pooling of data across a series of studies, as seen in Experiments 1-3 and 4-6, which supports the notion of converging results together. Very similar results were found across the two experimental chapters, which further justifies the use of this methodology.

Other problems with this methodology also need to be acknowledged. For instance, a range of tasks were used in this experimental series, all of which were carried out by every participant. This was a major constraint concerning this methodology, both in time and expense. The sample sizes were small in Experiments 1-6 ($n = 45-51$) and the results need

to be examined in light of this. A replication of these experiments using larger sample numbers would strengthen the findings here.

In contrast there are three basic assumptions behind the experimentalist approach. Firstly, it allows replicability, secondly it allows us to make statements about cause and effect, and thirdly it permits a considerable degree of control which helps to make it the most objective method. Most scientific theories are intended to apply to all cases of a similar kind, and hence results try to reflect the theory rather than the particular sample of participants used (generalisability). This can be applied to the dual task work discussed previously where it was found that the CE component is implicated in the reasoning process. However, despite the general agreement that dual tasking involves executive processes, there is still no clear consensus on what abilities or specific executive functions are implicated in dual task performance (e.g., Shah & Miyake, 1999). The results from dual tasking can be generalised to the population in question, but they cannot account for individual variation on the measures used. Most of the recent literature indicates that individual differences studies have been supplemented by experimental analyses, but that the reverse of this is not true.

The experimental method is able to demonstrate causal relationships, as opposed to the individual differences approach, which relies on correlational relationships. The results from this experimental series showed that participants with a lower working memory capacity performed less well on the reasoning tasks than participants with a larger capacity. However, it must be acknowledged that a causal interpretation was made of this result when other interpretations could have been made. Therefore, it makes sense to converge the dual task work with the individual differences work, although the individual differences method itself adds something to the total picture. Understanding the origins of individual

differences in its own right is important in the understanding of why some people reason better than others.

5 . 7 FUTURE DIRECTIONS

Much discussion has centred around the definition of factor 3 (general resource) and factor 4 (thinking style) from the CFA. Factor 3 was assumed to represent a general attentional resource, rather than a pure memory resource, capable of switching and inhibiting responses, with further processing components for verbal and spatial information (factor 1 and factor 2). The general resource was proposed to be responsible for the processing of both verbal and spatial based tasks, reasoning and memory spans alike, using domain independent representations. In relation to the reasoning tasks, the results have pointed to the general factor as a reflection of the reasoner's ability to resist pre-potent responses, and to de-contextualise their reasoning processes. In relation to the memory span tasks, the results have pointed to the general factor as a reflection of the participant's ability to switch between the storage and processing components of the task. Thus, it does seem likely that this factor is a controlled processing factor. It was compared to the cognitive ability factor of Stanovich and West (1998), who proposed that measures of this factor reflected decontextualisation, and consequently it was also linked to intelligence. In contrast, the fourth factor has been identified as a reflection of the reasoner's motivation to search for counter-examples, being indicative of their particular thinking disposition. This section outlines possible future research directions that may further shed light on the characteristics of these factors.

Thus, the factor identified as a general resource in Experiments 1-6 might well represent an element of executive function (controlled processing), with two additional distinct factors

related to verbal and spatial processing. The important question concerns the function of this general factor. Many executive functions have been examined in the literature and are implicated in relation to these findings.

The pattern of data throughout this thesis supports the notion of a general CE processor, with two further domain-specific CE components for verbal and spatial information. The CFA findings are consistent with the notion that both storage and processing functions are carried out in the specific resources, whilst the general resource entails controlled attention. Baddeley and Logie (1999) argue that their current conceptualisation of working memory no longer attributes any storage capabilities to the CE. Engle, Kane and Tuholski (1999) make the same argument by proposing that working memory is STM (maintenance) plus 'controlled attention' (executive control). In contrast, according to Lovett, Reder and Lebiere (1999), the maintenance of working memory elements (such as goals) is essential for appropriate control and regulation of cognitive behaviour to arise. Therefore, one important research question is to evaluate the degree to which the maintenance function and the executive control functions of working memory are separable.

Another unresolved issue concerns a precise specification of different executive control functions and their interrelationships. For instance, Miyake, Friedman, Emerson, Witzki and Howerter (in press) investigated the role of executive function in complex frontal lobe tasks. They were interested in specifying what the general-purpose control mechanisms are that modulate the operation of various cognitive subprocesses and thereby regulate the dynamics of human cognition. They focused on three executive functions - shifting of mental sets, monitoring and updating of working memory representations, and inhibition of prepotent responses. The aim was to examine how separable these functions were and how

they contributed to executive tasks, such as the WCST and the TOH. The CFA indicated that the three target executive functions are clearly distinguishable, but are related constructs.

The findings from Experiments 1-6 can be related to this executive function model. This more recent work has separated out different semi-independent explicit processing components. It could be that the working memory capacity tasks would have loaded the monitoring and updating of working memory representations factor. It might also be the case that a combination of these executive functions play a role in reasoning (inhibition and switching), where the factors predict different aspects of the reasoning process. Therefore, an interesting future research direction might be to examine the relationship between simple measures of executive function and a range of reasoning tasks.

From Experiments 4-6 CFA a fourth factor emerged which accounted for some of the variance seen in performing the reasoning tasks. No direct measure of this fourth factor was used in this experimental series, it emerged due to the covariances between the reasoning tasks. Other research has investigated ways of measuring this thinking style or thinking disposition. As already mentioned, Stanovich and West (1998) identified a thinking style influence by measuring individual's performance on a composite test of questionnaires. A recent study by Torrens, Thompson and Cramer (1999) examined the ability to search for alternatives in the syllogistic task. Participants were asked to draw as many diagrams as possible to represent the premises of two quantified syllogisms. This was taken as their alternatives generation score. Participants who produced more alternative representations were more likely to give logically correct answers and less likely to respond on the basis of belief. Furthermore, a range of other measures, including need for cognition, open-

mindedness and intelligence all failed to serve as predictors. Alternative generation was the only measure that served to predict believability effects. In future work, it might be fruitful to examine these relationships in more detail.

Another measure, shown to be a usable and useful instrument for the assessment of an individual's thinking style, is the REI - the Rational Experiential Inventory (Epstein, Pacini, Denes-Raj & Heier, 1996). They suggest that the rational system operates at the conscious level, whilst the experiential system operates at an automatic or preconscious level. In any situation, behaviour is determined jointly by the two systems. Individual differences are assumed to exist in the extent to which people rely on each system, with some people being more inclined to rational approaches, others to experiential. Handley, Newstead and Wright (2000) found that the REI did not correlate with other measures of thinking, such as intelligence scales. This might suggest that the REI measures are independent of cognitive ability in the same way that the thinking disposition identified by Stanovich and West (1998) was needed in addition to cognitive ability.

The work carried out in this thesis could be further investigated by using direct measures of the fourth factor that emerged in the CFA. Direct measures of working memory capacity identified three independent limited resources in relation to reasoning task performance. The fourth factor emerged due to the range of reasoning tasks used and was found to be independent of the three capacity resources. From the discussion above it can be seen that many tasks have been developed in order to measure how an individual's cognitive style influences the deductive process. These tasks could be used together with measures of capacity and reasoning in order to better identify the need for this fourth factor. These tasks

would also be able to specify exactly what this fourth factor represents and how this relates to cognitive ability/capacity and individual differences in the reasoning process.

5.8 CONCLUSION

In conclusion, individual differences in working memory capacity predicted individual differences in logical reasoning performance. These predictions covered many different reasoning tasks, classified as propositional, quantifiable, and simple and complex spatial. The working memory resources were represented by three factors. These were identified as a combination of active processing and passive storage resources for verbal and spatial information, together with a general factor entailing controlled processing. The reasoning tasks consistently loaded the general factor, but also showed relationships with both the verbal and spatial factors. It was concluded that complex deductive reasoning tasks rely on centrally controlled working memory components, consisting of general processing, plus verbal and spatial capacities.

A three-factor model of executive function best accounted for the working memory capacity tasks. The general factor can be compared to Stanovich and West's (1998) cognitive ability influence. These ability and capacity measures predicted logical reasoning performance similarly, perhaps reflecting a commonality between the two. This commonality might be what is referred to as *g*, thus showing links between cognitive capacity, cognitive ability and general intelligence. The role that this general factor played in the reasoning process was seen as reflecting the capacity the reasoner has for de-contextualising their reasoning processes. The role this general factor played in the memory process was seen as reflecting the capacity the participant has for switching between the storage and processing elements of the task. Taken together, the results for the reasoning and span tasks, suggest that the

general factor is a controlled processor. This thesis also highlighted the need for additional executive functions, over and above the general resource. These domain-specific components, as represented by factor 1 and factor 2 in the CFA, were obligatory elements in relation to the reasoning tasks used. Thus, in conclusion, a non-unitary view of CE resources can be supported, but in a way as yet not seen in the literature. These results are inconsistent with either a domain-specific or domain-general explanation of working memory executive resources, requiring both these notions in order to account for the data. The pattern of data supported a model where verbal and spatial capacities (Shah & Miyake, 1996) plus general processing resources (Turner & Engle, 1989) of executive function were needed in order to account for the findings. Future research, using a range of executive function tasks, might enable a better understanding of the general factor identified in this thesis.

A fourth factor was also identified in the CFA. This was interpreted as representing a thinking style or disposition of the reasoner. This was also found in the work of Stanovich and colleagues and was most relative to the notion that the reasoner must search for alternative conclusions that might refute initial ones in order for correct responding. Many ways have now been developed to measure thinking disposition directly. One goal of future research might be to examine the relationship between capacity and thinking style by using direct measures of both these elements.

The three classes of reasoning theory must also account for the patterns of data found here. This data highlighted that working memory capacity consistently predicts logical reasoning task performance. The heuristic accounts of reasoning do not rely on the notion of an explicit system, reasoners are suggested to use simple strategies in order to solve these

tasks. If an account of reasoning based on simple heuristics is to be accepted, then the result that measures of an explicit system predict performance must be integrated with this. Due to the lack of dissociation between the complex verbal and spatial working memory spans, these results are difficult to interpret in light of the processing theories of reasoning - rules or models for instance. However, especially for the syllogistic task, links were found with the spatial working memory spans and factor 2 (the spatial factor) which suggested the need for spatial resources. The processing accounts, especially the verbal ones, must be able to explain why this relationship exists.

Having demonstrated the usefulness of using span measures, assumed to reflect the capacity of the explicit working memory system, to predict deductive reasoning task performance, the obvious way forward is to extend this research to further distinguish the role of working memory in reasoning. For instance, future research is required to investigate more fully the notion that the general factor reflects controlled processing, whilst the specific verbal and spatial factors store, process and manipulate verbal and spatial information, respectively. In addition, future research is required to directly measure the processes underlying the fourth factor, using specific thinking style tasks. Finally, more research is needed to distinguish between theories of reasoning. Due to the reasoning tasks correlations with both verbal and spatial working memory tasks, an idea might be to draw on the notion of strategies in the reasoning process, and to integrate this as part of the experimental design.

REFERENCES

- Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General* 117, 288-318.
- Atkinson, R. C., & Shiffrin, R. M. (1968). *Human memory: A proposed system and its control processes*. In K. W. Spence (Ed.), *The psychology of learning and motivation: advances in research and theory* (Vol. 2), pp. 89-195. New York: Academic Press.
- Atkinson, R. C., & Shiffrin, R. M. (1971). The control of short-term memory. *Scientific American* 225, 82-90.
- Baddeley, A. D. (1966a). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology* 18, 362-365.
- Baddeley, A. D. (1966b). The influence of acoustic and semantic similarity on long-term memory for word sequences. *Quarterly Journal of Experimental Psychology* 18, 302-309.
- Baddeley, A. D. (1966c). The capacity for generating information by randomization. *Quarterly Journal of Experimental Psychology* 18, 119-129.
- Baddeley, A. D. (1979). *Working memory and reading*. In P. A. Kolars, M. E. Wrolstad and H. Bouma (Eds.), *Processing of visible language*, pp. 305-370. New York: Plenum Publishing Corporation.

Baddeley, A. D. (1981). The cognitive psychology of everyday life. *British Journal of Psychology* 72, 257-269.

Baddeley, A. D. (1983). Working memory. *Philosophical Transactions of the Royal Society of London B302*, 311-324.

Baddeley, A. D. (1986). *Working memory*. New York: Oxford University Press.

Baddeley, A. D. (1990). *Human memory*. Hove: Lawrence Erlbaum Associates Ltd.

Baddeley, A. D. (1993b). *Working memory or working attention?* In A. D. Baddeley and L. Weiskrantz (Eds.), *Attention: Selection, awareness, and control: A tribute to Donald Broadbent*. Oxford, UK: Oxford University Press.

Baddeley, A. D. (1993b). *Working memory and conscious awareness*. In A. Collins, S. E. Gathercole, M. A. Conway and P. E. Morris (Eds.), *Theories of memory* (pp.11-28). Hove, UK: Erlbaum.

Baddeley, A. D. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology* 49A(1), 5-28.

Baddeley, A. D. (1997). *Human memory: Theory and practice*. Revised Edition. Hove: Psychology Press.

Baddeley, A. D., Eldridge, M., Lewis, V. & Thomson, N. (1984). Attention and retrieval from long-term memory. *Journal of Experimental Psychology: General* 113, 518-540.

Baddeley, A. D., Grant, W., Wight, E. & Thomson, N. (1975). *Imagery and visual working memory*. In P. M. A. Rabbitt and S. Dornic (Eds.), *Attention and performance V*, pp. 205-217. London: Academic Press.

Baddeley, A. D., & Hitch, G. (1974). *Working memory*. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8). New York: Academic Press.

Baddeley, A. D., & Hitch, G. (1977). *Recency re-examined*. In S. Dornic (Ed.), *Attention and performance VI*, pp. 647-667. Hillsdale, NJ: Erlbaum.

Baddeley, A. D. & Levy, B. A. (1971). Semantic coding and short-term memory. *Journal of Experimental Psychology* 89, 132-136.

Baddeley, A. D. & Lieberman, K. (1980). *Spatial working memory*. In R. Nickerson (Ed.), *Attention and performance VIII*, pp. 521-539. Hillsdale, NJ: Erlbaum.

Baddeley, A. D., Lewis, V. & Vallar, G. (1984). Exploring the articulatory loop. *The Quarterly Journal of Experimental Psychology* 36A, 233-252.

Baddeley, A. D. & Logie, R. H. (1999). *Working memory: The multiple-component model*. In A. Miyake and P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge: Cambridge University Press.

Baddeley, A. D., Thomson, N. & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behaviour* 14, 575-589.

Baddeley, A. D. & Warrington, E. K. (1970). Amnesia and the distinction between long- and short-term memory. *Journal of Verbal Learning and Verbal Behaviour* 14, 575-589.

Baguley, T. & Payne, S. J. (1998). *Given-new versus new-given? An analysis of reading times for spatial descriptions*. Mind III Conference, August 1998, Dublin.

Barclay, J. R. (1973). The role of comprehension in remembering sentences. *Cognitive Psychology* 4, 229, 254.

Baron, J. (1985). *Rationality and intelligence*. Cambridge: Cambridge University Press.

Baron, J. (1988). *Thinking and deciding*. Cambridge: Cambridge University Press.

Basso, A., Spinnler, H., Vallar, G. & Zanobio, E. (1982). Left hemisphere damage and selective impairment of auditory verbal short-term memory: A case study. *Neuropsychologica* 20, 263-274.

Begg, I. & Denny, J. P. (1969). Empirical reconciliation of atmosphere and conversion interpretations of syllogistic reasoning errors. *Journal of Experimental Psychology* 81, 351-354.

- Begg, I. & Harris, G. (1982). On the interpretation of syllogisms. *Journal of Verbal Learning and Verbal Behaviour* 21, 595-620.
- Bourke, P. A. (1997). Measuring attentional demand in continuous dual-task performance. *The Quarterly Journal of Experimental Psychology* 50A(4), 821-840.
- Braine, M. D. S. (1978). On the relation between natural logic of reasoning and standard logic. *Psychological Review* 85, 1-21.
- Braine, M. D. S. & O'Brien, D. P. (Eds.) (1998). *Mental logic*. NJ, USA: LEA.
- Braine, M. D. S., Reiser, B. J. & Rumin, B. (1984). Some empirical justification for a theory of natural propositional logic. *The Psychology of Learning and Motivation* 18, 313-367.
- Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology* 10, 12-21.
- Brown, J. I., Bennett, J. M. & Hanna, G. (1981). *The Nelson-Denny reading test*. Chicago, IL: The Riverside Publishing Company.
- Byrne, R. M. J. & Handley, S. J. (1997). Reasoning strategies for suppositional deductions. *Cognition* 62, 1-49.

- Byrne, R. M. J. & Johnson-Laird, P. N. (1989). Spatial reasoning. *Journal of Memory & Language* 28, 564-575.
- Carpenter, P. A. & Just, M. A. (1975). Sentence comprehension: A psycholinguistic model of verification. *Psychological Review* 82, 45-73.
- Carpenter, P. A., Just, M. A. & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices test. *Psychological Review* 97, 404-431.
- Carroll, J. B. (1978a). How shall we study individual differences in cognitive abilities? Methodological and theoretical perspectives. *Intelligence*, 2, 87-115.
- Carroll, J. B. (1989). *Factor analysis since Spearman: Where do we stand? What do we know?* In R. Kanfer, P. L. Ackerman & R. Cudeck (Eds.), *Learning and individual differences: Abilities, motivation, and methodology*. Hillsdale, NJ: Erlbaum.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. NY: Cambridge University Press.
- Case, R. D., Kurland, D. M. & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology* 33, 386-404.
- Chapman, L. J. & Chapman, J. P. (1959). Atmosphere effect re-examined. *Journal of Experimental Psychology* 58, 220-226.

Chater, N. & Oaksford, M. (1999). The probability heuristics model of syllogistic reasoning. *Cognitive Psychology* 38, 191-258.

Clark, H. H. (1969a). Linguistic processes in deductive reasoning. *Psychological Review* 76, 387-404.

Cohen, M. R. & Nagel, E. (1934). *An introduction to logic and the scientific method*. New York: Harcourt, Brace & Co.

Colle, H. A. & Welsh, A. (1976). Acoustic masking in primary memory. *Journal of Verbal Learning and Verbal Behaviour* 15, 17-32.

Conrad, R. (1964). Acoustic confusion in immediate memory. *British Journal of Psychology* 55, 75-84.

Conrad, R. & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology* 55, 429-432.

Daneman, M. & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behaviour* 19, 450-466.

Daneman, M. & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 9(4), 561-584.

Daneman, M. & Tardif, M. (1987). *Working memory and reading skill re-examined*. In M. Coltheart (Ed.), *Attention and Performance XII: The Psychology of Reading* (pp.491-508). Hillsdale, NJ: Erlbaum.

De Renzi, E. (1982). Memory disorders following focal cortical damage. *Philosophical Transactions of the Royal Society of London B* 298, 73-83.

De Renzi, E. & Spinnler, H. (1967). Impaired performance on colour tasks in patients with hemispheric damage. *Cortex* 3, 194-216.

De Soto, C. B., London, M. & Handel, S. (1965). Social reasoning and spatial paralogic. *Journal of Personality and Social Psychology* 2, 513-521.

Deary, I. J. & Stough, C. (1996). Intelligence and inspection time. *American Psychologist* 51, 599-608.

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

Dillon, W. R., Kumar, A. & Mulani, N. (1987). Offending estimates in covariance structure analysis: Comments on the causes of and solutions to Heywood cases. *Psychological Bulletin* 101(1), 126-135.

Dunn, G., Everitt, B. & Pickles, A. (1993). *Modelling covariances and latent variables using EQS*. London: Chapman & Hall.

- Egan, D. E. & Grimes-Farrow, D. D. (1982). Differences in mental representations spontaneously adopted for reasoning. *Memory and Cognition* 10(4), 297-307.
- Ehrlich, K. & Johnson-Laird, P. N. (1982). Spatial descriptions and referential continuity. *Journal of Verbal Learning and Verbal Behaviour* 21, 296-306.
- Engle, R. W., Cantor, J. & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18, 972-992.
- Engle, R. W., Kane, M. J. & Tuholski, S. W. (1999). *Working memory and controlled attention*. In A. Miyake and P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge: Cambridge University Press.
- Epstein, S., Pacini, R., Denes-Raj, V. & Heier, H. (1996). Individual differences in intuitive-experiential and analytical-rational thinking styles. *Journal of Personality and Social Psychology* 71, 390-405.
- Erikson, J. R. (1974). *A set analysis theory of behaviour in formal syllogistic reasoning tasks*. In R. L. Solso (Ed.), *Theories of cognitive psychology: The Loyola Symposium*. Hillsdale, NJ: LEA Inc.
- Erikson, J. R. (1978). *Research on syllogistic reasoning*. In R. Revlin & R. E. Mayer (Eds.), *Human Reasoning*. Washington, DC: Winston.

- Evans, F. J. (1978). Monitoring attention deployment by random number generation: An index to measure subjective randomness. *Bulletin of the Psychonomic Society* 12, 35-38.
- Evans, J. St. B. T., Barston, J. L. & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory and Cognition* 11, 295-306.
- Evans, J. St. B. T. & Brooks, P. G. (1981). Competing with reasoning: A test of the working memory hypothesis. *Current Psychological Research* 1, 139-147.
- Evans, J. St. B. T., Handley, S., Johnson-Laird, P. N. & Harper, C. (1999). Reasoning about necessity and possibility: A test of the mental model theory of deduction. *Journal of Experimental Psychology: Learning, Memory and Cognition* 25(6), 1495-1513.
- Evans, J. St. B. T., Newstead, S. E. & Byrne, R. M. J. (1993). *Human Reasoning*. East Sussex: LEA Ltd.
- Evans, J. St. B. T. & Over, D. E. (1996). *Rationality and reasoning*. Hove, UK: Psychology Press.
- Eysenck, M. W. (1986). *Working memory*. In G. Cohen, M. W. Eysenck & M. A. Le Voi (Eds.), *Memory: A cognitive approach*. Milton Keynes: Open University Press.
- Farmer, E. W., Berman, J. V. F. & Fletcher, Y. L. (1986). Evidence for a visio-spatial scratch-pad in working memory. *The Quarterly Journal of Experimental Psychology* 38A, 675-688.

- Ferreira F. & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language* 25, 348-368.
- Fisher, D. L. (1981). A three-factor model of syllogistic reasoning: The study of isolable stages. *Memory and Cognition* 9, 496-514.
- Ford, M. (1994). Two modes of mental representation and problem solution in syllogistic reasoning. *Cognition* 54, 1-71.
- Frase, L. T. (1968). Effects of semantic incompatibility upon deductive reasoning. *Psychonomic Science* 12, 64.
- Galotti, K. M., Baron, J. & Sabini, J. P. (1986). Individual differences in syllogistic reasoning: Deduction rules or mental models? *Journal of Experimental Psychology: General* 115(1), 16-25.
- Geis, M. C. & Zwicky, A. M. (1971). On invited inferences. *Linguistic Inquiry* 2, 561-566.
- Gigerenzer, G. & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review* 103, 650-669.
- Gilhooly, K. J. (1998). *Working memory, strategies, and reasoning tasks*. In R. H. Logie and K. J. Gilhooly (Eds.), *Working memory and thinking*, pp. 7-22. East Sussex: Psychology Press.

Gilhooly, K. J. & Logie, R. H. (Eds.) (1998). *Thinking in working memory*. In Working memory and thinking, pp. 1-7. East Sussex: Psychology Press.

Gilhooly, K. J., Logie, R. H. & Wynn, V. (1999). Syllogistic reasoning tasks, working memory and skill. *European Journal of Cognitive Psychology* 11(4), 473-498.

Gilhooly, K. J., Logie, R. H., Wetherick, N. E. & Wynn, V. (1993). Working memory and strategies in syllogistic-reasoning tasks. *Journal of Memory and Cognition* 21(1), 115-124.

Glanzer, M. & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behaviour* 5, 351-360.

Guyer, B. L. & Friedman, M. P. (1975). Hemispheric processing and cognitive styles in learning-disabled and normal children. *Child Development* 46, 658-668.

Guyote, M. J. & Sternberg, R. J. (1981). A transitive chain theory of syllogistic reasoning. *Cognitive Psychology* 13, 461-525.

Hagert, G. (1983). Unpublished manuscript, Uppsala Programming Methodology and Artificial Intelligence Laboratory, Uppsala University.

Hagert, G. (1984). *Modeling mental models: Experiments in cognitive modeling of spatial reasoning*. In T. O'Shea (Ed.), *Advances in artificial intelligence*, pp. 389-398.

Amsterdam: North-Holland.

Halford, G. S., Bain, J. D. & Mayberry, M. T. (1984). Does a concurrent memory load interfere with reasoning? *Current Psychological Research and Reviews*, Summer, 1984.

Handley, S., Newstead, S & Wright, H. (2000). *Rational and experiential thinking: A study of the REI*. In R. J. Riding & S. G. Rayner (Eds.), *International perspectives on individual differences, volume one, cognitive styles*. Connecticut, USA: Ablex Publishing Corporation.

Harman, H. H. (1971). *Modern factor analysis*. Chicago: University of Chicago Press.

Hitch, G. J. & Baddeley, A. D. (1976). Verbal reasoning and working memory. *Quarterly Journal of Experimental Psychology* 28, 603-621.

Humphreys, G. W. & Riddoch, J. (1987). *To see but not to see: A case study of visual agnosia*. Hove, UK: LEA Ltd.

Humphreys, G. W. & Riddoch, M. J. (Eds.) (1987). *Visual object processing: A cognitive neuropsychological approach*. Hove, UK: LEA Ltd.

Hunter, I. M. L. (1957). The solving of three term series problems. *British Journal of Psychology* 48, 286-298.

Johnson-Laird, P. N. (1972). The three term series problem. *Cognition* 1, 58-82.

Johnson-Laird, P. N. (1983). *Mental models*. Cambridge: Cambridge University Press.

- Johnson-Laird, P. N. & Bara, B. G. (1984). Syllogistic inference. *Cognition* 17, 1-61.
- Johnson-Laird, P. N. & Byrne, R. M. J. (1990). Meta-logical reasoning: Knights, knaves and Rips. *Cognition* 36, 69-84.
- Johnson-Laird, P. N. & Byrne, R. M. J. (1991). *Deduction*. Hove: Erlbaum.
- Johnson-Laird, P. N., Byrne, R. M. J. & Tabossi, P. (1989). Reasoning by model: The case of multiple quantification. *Psychological Review* 96(4), 658-673.
- Johnson-Laird, P. N. & Steedman, M. (1978). The psychology of syllogisms. *Cognitive Psychology* 10, 64-99.
- Jorm, A. F. (1983). Specific reading retardation and working memory: A review. *British Journal of Psychology* 74, 311-342.
- Jurden, F. H. (1995). Individual differences in working memory and complex cognition. *Journal of Educational Psychology* 87, 93-102.
- Just, M. A. & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review* 99, 122-149.
- Kahneman, D., Slovic, P. & Tversky, A. (1982). *Judgement under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

- Kane, M. J. & Engle, R. W. (1998). Full frontal fluidity: Working memory capacity, controlled attention, intelligence, and the prefrontal cortex. (Unpublished manuscript).
- Kaufman, H. & Goldstein, S. (1967). The effects of emotional value of conclusions upon distortions in syllogistic reasoning. *Psychonomic Science* 7, 367-368.
- Klauer, K. C., Stegmaier, R. & Meiser, T. (1997). Working memory involvement in propositional and spatial reasoning. *Thinking and Reasoning* 3(1), 9-47.
- Kucera, H. & Francis, W. N. (1970). *Computational analysis of present-day American English*. Providence, Rhode Island: Brown University Press.
- Kyllonen, H. (1996). *Is working memory capacity Spearman's g?* In I. Dennis and P. Tapsfield (Eds.), *Human abilities: Their nature and measurement*. NJ: Lawrence Erlbaum Associates.
- Kyllonen, P. C. & Christal, R. E. (1990). Reasoning ability is (little more than) working memory capacity?! *Intelligence* 14, 389-433.
- La Pointe, L. B. & Engle, R. W. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology, Learning, Memory, and Cognition* 16(6), 1118-1133.
- Lehto, J. (1996). Are executive function tests dependent on working memory capacity? *The Quarterly Journal of Experimental Psychology* 49A, 29-50.

Levy, B. A. (1971). The role of articulation in auditory and visual short-term memory. *Journal of Verbal Learning and Verbal Behaviour* 10, 123-132.

Levy, B. A. (1975). Vocalisation and suppression effects in sentence memory. *Journal of Verbal Learning and Verbal Behaviour* 14, 304-316.

Logie, R. H. (1995). *Visuo-spatial working memory*. Hove, UK: Erlbaum.

Logie, R. H., Baddeley, A., Mane, A., Donchin, E. & Sheptak, R. (1989). Working memory in the acquisition of complex cognitive skills. *Acta Psychologica* 71, 53-87.

Logie, R. H. & Marchetti, C. (1991). *Visuo-spatial working memory: Visual, spatial or central executive?* In R. H. Logie & M. Denis (Eds.), *Mental images in human cognition*, pp. 105-115. Amsterdam: Elsevier.

Lohman, D. F. & Kyllonen, P. C. (1983). *Individual differences in solution strategy on spatial tasks*. In R. F. Dillon and R. R. Schmeck (Eds.), *Individual differences in cognition* Vol. 1, pp. 105-135. NY: Academic Press.

Lovett, M. C., Reder, L. M. & Lebiere, C. (1999). *Modeling working memory in a unified architecture: An ACT-R perspective*. In A. Miyake and P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge: Cambridge University Press.

- Macleod, C. M., Hunt, E. B. & Mathews, N. N. (1978). Individual differences in the verification of sentence picture relationships. *Journal of Verbal Learning and Verbal Behaviour* 19, 531-548.
- Mani, K. & Johnson-Laird, P. N. (1982). The mental representation of spatial descriptions. *Memory and Cognition* 10(2), 181-187.
- Marquer, J. M. & Pereira, M. (1990). Reaction times in the study of strategies in sentence-picture verification: A reconsideration. *Quarterly Journal of Experimental Psychology* 42A, 147-168.
- Mayberry, M. T., Bain, J. D. & Halford, G. S. (1986). Information-processing demands of transitive inference. *Journal of Experimental Psychology: Learning, Memory and Cognition* 12, 600-613.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63, 81-97.
- Milner, B. (1966). *Amnesia following operation on the temporal lobes*. In C. W. M. Whitty and O. L. Zangwill (Eds.), *Amnesia*, pp. 109-133. Butterworths.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H. & Howerter, A. (in press). The unity and diversity of executive functions and their contributions to complex 'frontal lobe' tasks: A latent variable analysis. *Cognitive Psychology*.

Milner, B. (1971). Interhemispheric differences in the localisation of psychological processes in man. *British Medical Bulletin 3 (Cognitive Psychology)*, 272-277.

Moeser, S. D. & Tarrant, B. L. (1977). Learning a network of comparisons. *Journal of Experimental Psychology: Human Learning and Memory 3*, 643-659.

Morton, J. (1967). A singular lack of incidental learning. *Nature 215*, 203-204.

Murdock, B. B. Jr. (1965). Effects of a subsidiary task on short-term memory. *British Journal of Psychology 56*, 413-419.

Murray, D. J. (1968). Articulation and acoustic confusability in short-term memory. *Journal of experimental Psychology 78*, 679-684.

Mynatt, B. T. & Smith, K. H. (1977). Constructive processes in linear ordering problems revealed by sentence study times. *Journal of Experimental Psychology: Human Learning and Memory 3*, 357-374.

Navon, D. (1984). Resources - a theoretical soup stone? *Psychological Review 86*, 214-255.

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: LEA Inc.

Newell, A. (1990). *Unified theory of cognition*. Cambridge, MA: Harvard University Press.

Newell, A. & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

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

Newstead, S. E. (1995). Gricean implicatures and syllogistic reasoning. *Journal of Memory and Language* 34, 644-664.

Newstead, S. E. & Griggs, R. A. (1999). Premise misinterpretation and syllogistic reasoning. *Quarterly Journal of Experimental Psychology* 52A(4), 1057-1075.

Newstead, S. E., Handley, S. J. & Buck, E. (1999). Falsifying mental models: Testing the predictions of theories of syllogistic reasoning. *Memory and Cognition*.

Norman, D. A. & Shallice, T. (1980). Attention to action. Willed and automatic control of behaviour. University of California San Diego CHIP Report 99.

Oakhill, J. V. & Johnson-Laird, P. N. (1984). Representation of spatial descriptions in working memory. *Current Psychological Research and Reviews*, Spring, 52-62.

Oakhill, J. V. & Johnson-Laird, P. N. (1985). The effect of belief on the spontaneous production of syllogistic conclusions. *Quarterly Journal of Experimental Psychology* 37A, 553-570.

Oakhill, J. V., Yuill, N. & Parkin, A. J. (1986). On the nature of the difference between skilled and less-skilled comprehenders. *Journal of Research in Reading* 9, 80-91.

Oaksford, M. & Chater, N. (1992). Bounded rationality in taking risks and drawing inferences. *Theory and Psychology* 2, 225-230.

Oaksford, M. & Chater, N. (1993). *Reasoning theories and bounded rationality*. In K. I. Manktelow & D. E. Over (Eds.), *Rationality* (pp. 31-60). London, UK: Routledge.

Ohlsson, S. (1981). Model-based vs. propositional inference in a spatial reasoning task. Unpublished paper, Department of Psychology, University of Stockholm.

Ohlsson, S. (1984). Induced strategy shifts in spatial reasoning. *Acta Psychologica* 57, 47-67.

Ohlsson, S. (1988). Trace analysis and spatial reasoning (Tech. Rep. No. KUL-87-02). Learning Research and Development Centre, University of Pittsburgh.

Osherson, D. (1975). *Logic and models of logical thinking*. In R. J. Falmagne (Ed.), *Reasoning: Representation and process in children and adults*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

- Paivio, A. (1971). *Imagery and verbal processes*. New York Holt: Rinehart & Winston.
- Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin* 116(2), 220-244.
- Perfetti, C. A. & Lesgold, A. N. (1977). *Discourse comprehension and sources of individual differences*. In M. Just & P Carpenter (Eds.), *Cognitive processes in comprehension*. Erlbaum, Hillsdale, N. J.
- Peterson, L. R. & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology* 58, 193-198.
- Phillips, W. A. & Christie, D. F. M. (1977a). Components of visual memory. *Quarterly Journal of Experimental Psychology* 29, 117-133.
- Phillips, W. A. & Christie, D. F. M. (1977b). Interference with visualisation. *Quarterly Journal of Experimental Psychology* 29, 637-650.
- Phillips, L. H., Wynn, V., Gilhooly, K. J., Della Sala, S. & Logie, R. H. (1999). The role of memory in the Tower of London task.
- Politzer, G. (1990). Non-logical solving of categorical syllogisms. In J. P. Caverni, J. M. Fabre and M. Gonzalez (Eds.), *Cognitive biases. Advances in Psychology* Vol. 68. Amsterdam: North Holland.

- Polk, T. A. & Newell, A. (1995). Deduction as verbal reasoning. *Psychological Review* 102(3), 533-566.
- Potts, G. R. (1972). Information processing strategies used in the encoding of linear orderings. *Journal of Verbal Learning and Verbal Behaviour* 11, 727-740.
- Potts, G. R. & Scholz, K. W. (1975). The internal representation of three-term series problems. *Journal of Verbal Learning and Verbal Behaviour* 14, 439-452.
- Quinlan, P. T. (1992). *The Oxford psycholinguistic database*. Oxford University Press.
- Revlis, R. (1975a). *Syllogistic reasoning: Logical decisions from a complex data base*. In R. J. Falmagne (Ed.), *Reasoning: Representation and process*. New York: John Wiley.
- Revlis, R. (1975b). Two models of syllogistic inference: Feature selection and conversion. *Journal of Verbal Learning and Verbal Behaviour* 14, 180-195.
- Rips, L. J. (1983). Cognitive processes in propositional reasoning. *Psychological Review* 90, 38-71.
- Rips, L. J. (1986). *Mental muddles*. In M. Brand & R. M. Harnish (Eds.), *The representation of knowledge and belief*, pp. 258-286. Tucson: University of Arizona Press.
- Rips, L. J. (1994). *The psychology of proof*. New York: MIT Press.

- Roberts, M. J. (1993). Human reasoning: Deduction rules or mental models, or both? *The Quarterly Journal of Experimental Psychology* 46A(4), 569-589.
- Roberts, M. J., Gilmore, D. J. & Wood, D. J. (1997). Individual differences and strategy selection in reasoning. *British Journal of Psychology* 88, 473-492.
- Rosen, V. M. & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General* 126, 211-227.
- Rust, J. & Golombok, S. (Eds.) (1999). *Modern psychometrics: the science of psychological assessment* (2nd Edition). London & New York: Routledge.
- Rylander, G. (1939). Personality changes after operations on the frontal lobes. *Acta Psychiatrica Neurologica*, Supplement No. 30.
- Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics* 2, 437-442.
- Salame, P. & Baddeley, A. D. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behaviour* 21, 150-164.
- Salthouse, T. A. & Babcock, R. L. (1990). The reading and computational span tasks. Unpublished manuscript, Georgia Institute of Technology.

Sanford, A. J. (1985). *Cognition and cognitive psychology*. London: Weidenfeld & Nicholson.

Schaeken, W., Johnson-Laird, P. N. & d'Ydewalle, G. (1996). Mental models and temporal reasoning. *Cognition* 60, 205-234.

Sells, S. B. (1936). The atmosphere effect: An experimental study of reasoning. *Archives of Psychology No. 200*.

Shah, P. & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General* 125(1), 4-27.

Shah, P. & Miyake, A. (1999). *Models of working memory: An introduction*. In A. Miyake and P. Shah (Eds.), *Models of working memory, mechanisms of active maintenance and executive control*, pp. 1-27. Cambridge: Cambridge University Press.

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London B* 298, 199-209.

Shallice, T. & Warrington, E. K. (1970). Independent functioning of verbal memory stores: A neuropsychological study. *Quarterly Journal of Experimental Psychology* 22, 261-273.

- Shiffrin, R. M. & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review* 84, 127-190.
- Smith, K. H. & Foos, P. W. (1975). Effect of presentation order on the construction of linear orders. *Memory and Cognition* 3, 614-618.
- Smith, P. & Whetton, C. (1988). *General ability tests: Spatial*. Windsor, Berkshire: NFER-NELSON Publishing Company.
- Spearman, C. (1904). 'General Intelligence', objectively determined and measured. *American Journal of Psychology* 15, 201-293.
- Spearman, C. (1927). *The abilities of man*. London: Macmillan.
- Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah, NJ: Erlbaum.
- Stanovich, K. E. & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology* 89, 342-357.
- Stanovich, K. E. & West, R. F. (1998). Cognitive ability and variation in selection task performance. *Thinking and reasoning* 4, 193-230.

- Sternberg, R. J. (1981). Reasoning with determinate and indeterminate linear syllogisms. *British Journal of Psychology* 72, 407-420.
- Sternberg, R. J. & Turner, M. E. (1981). Components of syllogistic reasoning. *Acta Psychologica* 47, 245-265.
- Sternberg, R. J. & Weil, E. M. (1980). An aptitude x strategy interaction in linear syllogistic reasoning. *Journal of Educational Psychology* 72, 226-239.
- Toms, M., Morris, N. & Ward, D. (1993). Working memory and conditional reasoning. *Quarterly Journal of Experimental Psychology*, 46A, 679-699.
- Torrens, D., Thompson, V. A. & Cramer, K. M. (1998). Individual differences and the belief bias effect: Mental models, logical necessity, and abstract reasoning. *Thinking and Reasoning* 5(0), 00-00.
- Trabasso, T., Riley, C. A. & Wilson, E. G. (1975). *The representation of linear order and spatial strategies in reasoning: A developmental study*. In R. J. Falmagne (Ed.), Reasoning: Representation and process. NY: John Wiley.
- Turner, M. L. & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language* 28, 127-154.
- Vallar, G. & Baddeley, A. D. (1982). Short-term forgetting and the articulatory loop. *Quarterly Journal of Experimental Psychology* 34, 53-60.

- Vallar, G. & Baddeley, A. D. (1984a). Fractionation of working memory: Neuropsychological evidence for a phonological short-term store. *Journal of Verbal Learning and Verbal Behaviour* 23, 151-161.
- Vallar, G. & Baddeley, A. D. (1984b). Phonological short-term store, phonological processing and sentence comprehension: A neuropsychological case study. *Cognitive Neuropsychology* 1, 121-141.
- Vandierendonck, A. & De Vooght, G. (1997). Working memory constraints on linear reasoning with spatial and temporal contents. *Quarterly Journal of Experimental Psychology* 50A(4), 803-820.
- Vernon, P. A. (1991). The use of biological measures to estimate behavioural intelligence. *Educational Psychologist* 25, 293-304.
- Vernon, P. A. (1993). *Biological approaches to the study of human intelligence*. Norwood, NJ: Ablex.
- Vernon, P. E. (1947a). Research on personnel selection in the Royal Navy and the British Army. *American Psychologist* 2, 35-51.
- Vernon, P. E. (1947b). Psychological tests in the Royal Navy, Army and A.T.S. *Occupational Psychology* 21, 53-74.

Vernon, P. E. (1947c). The variations of intelligence with occupation, age and locality. *British Journal of Psychological Statist. Sec. 1*, 52-63.

Wagenaar, W. A. (1970). Subjective randomness and the capacity to generate information. *Acta Psychologica* 33.

Wason, P. C. & Evans, J. St. B. T. (1975). Dual processes in reasoning? *Cognition* 3, 141-154.

Wason, P. C. & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content*. London: Batsford.

Waters, G. S. & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *The Quarterly Journal of Experimental Psychology* 49A(1), 51-79.

Wetherick, N. E. (1989). Psychology and syllogistic reasoning. *Philosophical Psychology* 2, 111-124.

Wetherick, N. E. & Gilhooly, K. (1990). *Syllogistic reasoning: Effects of premise order*.

In K. Gilhooly, M. T. G. Keane, R. Logie and G. Erdos (Eds.), *Lines of thought:*

Reflections on the psychology of thinking, Vol. 1. London: John Wiley.

Wood, D. J., Shotton, J. & Godden, D. (1974). An investigation of the relationships between problem solving strategies, representation and memory. *Quarterly Journal of Experimental Psychology* 26, 252-257.

Woodworth, R. S. & Sells, S. B. (1935). An atmosphere effect in syllogistic reasoning. *Journal of Experimental Psychology* 18, 451-460.

LIST OF APPENDICES

APPENDIX A - Materials and Instructions Employed in Experiments 1-6

Appendix A1 - Experiment 1

A1. 1 Simple word span words	383
A1. 2 Simple word span instructions	384
A1. 3 Complex verbal span sentences	385
A1. 4 Complex verbal span instructions	388
A1. 5 Five-term series spatial problems	389
A1. 6 Five-term series spatial inference instructions	394
A1. 7 Syllogistic problems	395
A1. 8 Syllogistic inference instructions	399

Appendix A2 - Experiment 4

A2. 1 Temporal problems	401
A2. 2 Temporal inference instructions	404

Appendix A3 - Experiment 5

A3. 1 Abstract problems	405
A3. 2 Abstract inference instructions	406
A3. 3 Conditional problems	407
A3. 4 Conditional inference instructions	409
A3. 5 Three-term series spatial problems	410
A3. 6 Three-term series spatial inference instructions	411
A3. 7 Multiple quantification problems	412

A3. 8 Multiple quantification inference instructions	414
--	-----

APPENDIX B - Summary Tables of Statistical Analyses: t-tests, ANOVAs, correlations and factor analyses

Appendix B1 - All experiments

B1. 1 t-tests for Experiment 1	416
B1. 2 t-tests for Experiment 4	417
B1. 3 t-tests for Experiment 5	419
B1. 4 t-tests for Experiment 6	420
B1. 5 ANOVA results for Experiment 2	421
B1. 6 ANOVA results for Experiment 3	422
B1. 7 Correlation matrix for Experiment 1	423
B1. 8 Correlation matrix for Experiment 2	424
B1. 9 Correlation matrix for Experiment 3	425
B1. 10 Correlation matrix for Experiment 4	426
B1. 11 Correlation matrix for Experiment 5	427
B1. 12 Correlation matrix for Experiment 6	428
B1. 13 Multiple group CFA for Experiments 1-3 three-factor model	429-451
B1. 14 Multiple group CFA for Experiments 4-6 four-factor model	452-477

APPENDIX A: Materials and Instructions Employed in Experiments 1-6

Appendix A1 . 1: Experiment 1 Materials: Simple Word Span

All words were taken from the Oxford Psycholinguistic Database (Quinlan, 1992). They were all 2 syllables long and were concrete nouns. None of the words in a set began with the same letter.

Practice Sets



BERRY
CATTLE

LEMON
SINGER

JERSEY
APPLE

2 Sets



WALLET
COFFEE

MOVIE
KITTEN

BANNER
FOREST

SILVER
CANNON

PLANET
TABLE

3 Sets



ADULT
CARPET
DOLLAR

MOTHER
TOWER
HOTEL

BANKER
INSECT
OTTER

DINNER
ANKLE
TUNNEL

FLOWER
JOCKEY
HAMMER

4 Sets



BUCKET
PEPPER
COTTON
STABLE

MEDAL
ORGAN
COLLAR
DOCTOR

BATTLE
SUGAR
ANGEL
RABBIT

RIBBON
CABIN
POCKET
BISHOP

MANOR
SUPPER
DANCER
RACKET

5 Sets



ANCHOR
LIVER
NEEDLE
WINDOW
MIRROR

METAL
BUTTER
FATHER
RIVER
CANDY

CHEERY
ENGINE
PILLOW
STATUE
MARKET

VALLEY
BUBBLE
DRIVER
CIRCLE
LOCKER

ISLAND
PARTY
CANOE
BASKET
LETTER

6 Sets



PUZZLE
MARBLE
TIGER
HUNTER
SADDLE
BASIN
CEMENT
FERRY
INFANT
PYTHON
WOMAN
LAWYER

BOTTLE
TURTLE
CANDLE
GRAVEL
JACKET
MAPLE

ARTIST
WILLOW
TICKET
SUMMER
CIRCUS
LOBBY

GIANT
BARREL
EAGLE
ROCKET
CELLAR
DEMON

7 Sets



MUTTON
SISTER
BUCKLE
OFFICE
CLOVER
LEVER
PAPER

CIGAR
RUBBER
ELBOW
POWDER
BUTTON
GARDEN
LINEN

HONEY
CABLE
PRISON
MONEY
NUMBER
BRANDY
WATER

CORAL
MONKEY
SHOWER
FABRIC
BULLET
HEAVEN
ARROW

MUSIC
PENNY
OLIVE
RESORT
SALAD
ALLEY
WALNUT

Appendix A1 . 2: Experiment 1 Instructions: Simple Word Span

In this part of the experiment, you must remember sets of words that will appear on the screen. The number of words to remember in each set will increase throughout the test.

An example of a set of words can be seen below:-

The screen displays the word: **GUITAR**

Then the screen displays the word: **LORRY**

Then the screen displays the word: **RECALL**

Your task is to remember the set of words, in the order they were presented. For the above example, you would write down the words: **GUITAR, LORRY**.

You are provided with a response sheet to mark down your responses to each set of words. Please have a look at the sheet now to ensure you understand how to mark your responses.

If you **DO NOT** understand the task, please tell the experimenter now. If you **DO** understand the task, please press the space bar and you will be given three practice sets of words.

Appendix A1 . 3: Experiment 1 Materials: Complex Verbal Span

All end words were taken from the Oxford Psycholinguistic Database (Quinlan, 1992). They were all 1 or 2 syllables long and were concrete nouns. None of the end words in a set began with the same letter.

Practice Set

A WALLET HOLDS MONEY *T
COTTON WOOL IS HARD *F

2 Sets

A BIRD HAS FEET *F
SINGERS HAVE A GOOD VOICE *T

BREAKFAST IS EATEN IN THE MORNING *T
A RULER DRAWS A STRAIGHT LINE *T

MOUNTAIN TOPS ARE HIGH *T
WINNERS COME IN SECOND *F

A LETTER IS A NUMBER *F
DOGS ARE MADE OF FELT *F

THE CONCLUSION COMES FIRST *F
ENGLAND IS A COUNTRY *T

3 Sets

THERE ARE FLOORS IN A HOUSE *T
A SQUARE IS ROUND *F
AFTER FOUR COMES FIVE *T

OFFICERS WORK IN THE POLICE FORCE *T
KITTENS LIKE TO PLAY *T
CARPETS HANG ON A WALL *F

THE ACE OF HEARTS IS BLACK *F
ENGLISH DRIVE ON THE LEFT *T
TWENTY MONTHS MAKE A YEAR *F

WINNERS COME IN LAST *F
A PRINCE IS A WOMAN *F
TOMORROW IS IN THE FUTURE *T

A PENNY IS A COIN *T
THE SUN COMES OUT AT NIGHT *F
THERE ARE TOES ON A HAND *F

Appendix A1 . 3 (cont.): Experiment 1 Materials: Complex Verbal Span

4 Sets

THE NUMBER THREE IS EVEN *F
FISH LIVE ON THE LAND *F
GOLD HAS A HIGH VALUE *T
A LEMON IS A FLOWER *F

A DINGHY IS A BOAT *T
CARTOON CHARACTERS ARE REAL *F
THE LEADER IS AT THE FRONT *T
THERE ARE BRANCHES ON A TREE *T

BILLIONS ARE SMALLER THAN A MILLION *F
POLAR BEARS ARE WHITE *T
ELEPHANTS ARE VERY SMALL *F
FAMILIES LIVE IN A HOME *T

LIFTS GO UP AND DOWN *T
ENGLISH DRIVE ON THE RIGHT *F
WINDOWS CAN BE OPEN *T
MEMORIES ARE FROM THE PAST *T

A SLOPE IS LEVEL *F
MINERS WORK IN SCHOOL *F
CLOCKS TELL THE TIME *T
CARS DRIVE ON A ROAD *T

5 Sets

A BROTHER IS A GIRL *F
KEYBOARDS ARE USED TO TYPE *T
SEVEN COMES BEFORE FOUR *F
PEOPLE LIVE IN A CITY *T
FOOTBALL IS A SPORT *T

AT NIGHT TIME IT IS LIGHT *F
ELECTRIC IS A FORM OF POWER *T
WATER IS FOUND DOWN A WELL *T
ROCK MUSIC IS SUNG IN CHURCH *F
LEGS ARE PART OF THE FACE *F

JUDGES WORK IN COURT *T
CROPS GROW IN A WOOD *F
MUSEUMS ARE OPEN TO THE PUBLIC *T
ONE PERSON MAKES A GROUP *F
CARPETS COVER THE FLOOR *T

Appendix A1 . 3 (cont.): Experiment 1 Materials: Complex Verbal Span

WATER LIGHTS A FIRE *F
PLANES FLY AROUND THE WORLD *T
SEVENTEEN COMES BEFORE THREE *F
CHESS IS A GAME *T
A FROG HAS HAIR *F

THE HOSPITAL SENDS MAIL *F
TOWNS ARE FULL OF SAND *F
THE OCEAN CONTAINS WATER *T
AN APPLE IS A FRUIT *T
CARS ARE MADE OF BONE *F

6 Sets

A HAND HAS A PALM *T
YOU FIND FUR IN A SALAD *F
PLUG RHYMES WITH LEAF *F
BALD PEOPLE HAVE A FRINGE *F
HAUNTED HOUSES HAVE A GHOST *T
BROWNIES GO TO CAMP *T

TEA IS KEPT IN A FLASK *T
LIONS SLEEP IN A CRADLE *F
CIGARETTES COME IN A PACK *T
THE SEA IS MADE OF GRAVY *F
CEMENT IS A SPICE *F
YOU WAKE UP WITH AN ALARM *T

CANOEISTS USE A PADDLE *T
YOU SIT ON A STOOL *T
VODKA IS FOUND IN A LAKE *F
NINETY SECONDS EQUALS ONE HOUR *F
PIGS TAKE OFF FROM AN AIRPORT *F
RED IS THE COLOUR OF LIME *F

YOU KEEP ANIMALS IN A FILE *F
YOU EAT WITH A SWORD *F
THE SEA IS NEAR THE COAST *T
A CINEMA HAS A SPIRE *F
YOU CAN MOW A LAWN *T
PEOPLE WALK ALONG A PIER *T

YOU CAN SUCK YOUR THUMB *T
PEASANTS LIVE IN A PALACE *F
YOU CAN DRINK FROM A GOBLET *T
A DAFFODIL GROWS FROM A BULB *T
YOU STAY DRY IN THE RAIN *F
PUBS SELL CRISPS AND BEER *T

Appendix A1 . 4: Experiment 1 Instructions: Complex Verbal Span

In this part of the experiment, your task is to decide whether sets of sentences are true or false whilst remembering the last word of each sentence. The number of sentences to verify in each set will increase throughout the test. The test can be completed by using mouse clicks.

An example of a set of sentences can be seen below:-

The screen displays a sentence:

THE SKY IS BLUE

Your task is to decide whether the sentence is true (T) or false (F). In this case you would click the mouse in the “T” box. You must remember the last word of the sentence.

Then the screen displays another sentence: **THE SUN IS PINK**

Your task is to decide whether the sentence is T or F. In this case you would click the mouse in the “F” box. You must remember the last word of the sentence.

Then the screen displays the symbol of a **RECALL LIGHTBULB**

When the recall lightbulb appears, your task is to remember the last word in each sentence, in the order in which they appeared. For the above example, you would write down the words: **BLUE** and **PINK**.

You are provided with a response sheet to mark down your responses to each set of sentences. Please have a look at the sheet now to ensure you understand how to mark your responses.

If you **DO NOT** understand the task, please tell the experimenter now. If you **DO** understand the task, please press the space bar and you will be given a practice set of sentences.

Appendix A1 . 5: Experiment 1 Materials: Five-Term Series Spatial Inference

One Model Problems

The apple is on the left of the orange.

The banana is on the right of the orange.

The melon is in front of the apple.

The peach is in front of the banana.

Hence, what is the relationship between the melon and the peach?

The melon is left of the peach.

The vase is on the right of the clock.

The ashtray is on the left of the clock.

The plant is in front of the ashtray.

The lamp is in front of the clock.

Hence, what is the relationship between the plant and the lamp?

The plant is left of the lamp.

The dish is on the right of the spoon.

The knife is on the left of the spoon.

The fork is in front of the spoon.

The plate is in front of the knife.

Hence, what is the relationship between the fork and the plate?

The fork is right of the plate.

The clock is on the right of the vase.

The ashtray is on the right of the clock.

The plant is behind the ashtray.

The lamp is behind the vase.

Hence, what is the relationship between the plant and the lamp?

The plant is to the right of the lamp.

The knife is on the left of the dish.

The spoon is on the right of the dish.

The fork is behind the knife.

The plate is behind the dish.

Hence, what is the relationship between the fork and the plate?

The fork is to the left of the plate.

The torch is on the right of the drill.

The hammer is on the left of the drill.

The spanner is behind the hammer.

The pliers are behind the drill.

Hence, what is the relationship between the spanner and the pliers?

The spanner is to the left of the pliers.

Appendix A1 . 5 (cont.): Experiment 1 Materials: Five-Term Series Spatial Inference

The spoon is on the right of the plate.

The dish is behind the spoon.

The knife is behind the dish.

The knife is on the right of the fork.

Hence, what is the relationship between the fork and the plate?

The fork is behind the plate.

The spanner is on the left of the drill.

The torch is in front of the drill.

The hammer is in front of the torch.

The pliers are on the left of the hammer.

Hence, what is the relationship between the spanner and the pliers?

The spanner is behind the pliers.

The hammer is on the left of the pliers.

The torch is in front of the hammer.

The drill is in front of the torch.

The spanner is on the right of the drill.

Hence, what is the relationship between the spanner and the pliers?

the spanner is in front of the pliers.

The apple is behind the orange.

The orange is behind the banana.

The banana is on the left of the melon.

The peach is on the right of the orange.

Hence, what is the relationship between the melon and the peach?

The melon is in front of the peach.

The torch is on the right of the hammer.

The drill is on the left of the torch.

The spanner is in front of the drill.

The pliers are in front of the torch.

Hence, what is the relationship between the spanner and the pliers?

The spanner is to the left of the pliers.

The pliers are on the right of the torch.

The hammer is in front of the torch.

The drill is in front of the hammer.

The spanner is to the right of the hammer.

Hence, what is the relationship between the spanner and the pliers?

The spanner is in front of the pliers.

Multi Model Problems

The banana is on the left of the apple.
The orange is on the right of the banana.
The melon is in front of the orange.
The peach is in front of the banana.
Hence, what is the relationship between the melon and the peach?
The melon is to the right of the peach.

The ashtray is on the right of the clock.
The vase is on the left of the ashtray.
The plant is in front of the vase.
The lamp is in front of the ashtray.
Hence, what is the relationship between the plant and the lamp?
The plant is to the left of the lamp.

The torch is on the right of the hammer.
The drill is on the left of the torch.
The spanner is behind the drill.
The pliers are behind the torch.
Hence, what is the relationship between the spanner and the pliers?
The spanner is to the left of the pliers.

The orange is on the right of the apple.
The banana is on the left of the orange.
The melon is behind the banana.
The peach is behind the orange.
Hence, what is the relationship between the melon and the peach?
The melon is to the left of the peach.

The knife is on the left of the dish.
The spoon is on the right of the knife.
The fork is behind the knife.
The plate is behind the dish.
Hence, what is the relationship between the fork and the plate?
The fork is to the left of the plate.

The clock is on the left of the ashtray.
The vase is behind the clock.
The plant is behind the ashtray.
The lamp is behind the vase.
Hence, what is the relationship between the plant and the lamp?
The plant is to the right of the lamp.

Appendix A1 . 5 (cont.): Experiment 1 Materials: Five-Term Series Spatial Inference

The apple is on the left of the orange.
The orange is in front of the banana.
The melon is behind the banana.
The peach is behind the apple.
Hence, what is the relationship between the melon and the peach?
The melon is to the right of the peach.

The clock is on the right of the ashtray.
The vase is in front of the clock.
The plant is in front of the ashtray.
The lamp is in front of the vase.
Hence, what is the relationship between the plant and the lamp?
The plant is to the left of the lamp.

The apple is on the left of the orange.
The banana is behind the apple.
The peach is behind the banana.
The melon is behind the orange.
Hence, what is the relationship between the melon and the peach?
The melon is to the right of the peach.

The knife is on the left of the dish.
The fork is in front of the knife.
The spoon is in front of the dish.
The plate is in front of the spoon.
Hence, what is the relationship between the fork and the plate?
The fork is to the left of the plate.

The vase is on the left of the clock.
The ashtray is in front of the clock.
The plant is in front of the ashtray.
The lamp is in front of the vase.
Hence, what is the relationship between the plant and the lamp?
The plant is to the right of the lamp.

The spoon is on the left of the dish.
The dish is in front of the knife.
The dish is in front of the fork.
The knife is on the right of the plate.
Hence, what is the relationship between the fork and the plate?
The fork is to the right of the plate.

A sub-set of these problems (16) was used in Experiment 2 (verbal presentation).

Appendix A1 . 5 (cont.): Experiment 1 Materials: Five-Term Series Spatial Inference

Content Substitution in Experiment 2 (visual presentation)

beetroot	notebook
carrot	biro
turnip	calculator
potato	rubber
onion	pencil
trowel	glass
fork	bowl
pot	saucer
spade	mug
broom	napkin

These four content sets were randomly assigned to four problems each (a sub-set of the problems seen above).

Appendix A1 . 6: Experiment 1 Instructions: Five-Term Series Spatial Inference

In this part of the experiment you will be asked about the relations between objects in a set. The experimenter will read to you a description of the layout of a set of objects. You should listen attentively to the descriptions. Each description will be read to you twice, and at a reasonable pace. Each time the description is read, you will be asked about the relation between two of the objects in the set.

The descriptions will use the following terms to describe the layouts of the objects within a set:-

in front of

behind

to the left of

to the right of

not enough information to tell

You should imagine the objects as if they were in front of you and then apply the above relations to them. For some of the descriptions you may feel there is not enough information to determine the relation between the objects; in which case you should say so.

There is an answer sheet provided for your responses to each description. There will be 24 descriptions plus two practice descriptions. If you understand these instructions, please tell the experimenter who will read to you two simple practice descriptions to introduce the task. You may keep this information in front of you during the test to ensure you remember the ways in which you can link the objects together.

practice descriptions

The knife is on the right of the vase.

The glass is on the left of the vase.

What is the relationship between the knife and the vase?

The answer should be that the knife is on the right of the vase.

The knife is on the right of the vase.

The glass is on the left of the vase.

The dish is in front of the glass.

The plate is in front of the vase.

What is the relationship between the dish and the plate?

The answer should be that the dish is on the left of the plate.

Appendix A1 . 7: Experiment 1 Materials: Syllogistic Inference

Mood - AB-BC

Models - 1M

Your conclusion should link Welders and Carpenters.
Some of the Welders are Builders.
All of the Builders are Carpenters.
Thus,
Some of the Welders are Carpenters.

Your conclusion should link Secretaries and Climbers.
All of the Secretaries are Cabdrivers.
None of the Cabdrivers are Climbers.
Thus,
None of the Secretaries are Climbers.

Your conclusion should link Gardeners and Nurses.
None of the Nurses are Politicians.
All of the Politicians are Gardeners.
Thus,
None of the Gardeners are Nurses.

Your conclusion should link Conductors and Models.
All of the Conductors are Farmers.
All of the Farmers are Models.
Thus,
All of the Conductors are Models.

Models - MM

Your conclusion should link Plumbers and Weightlifters.
None of the Weightlifters are Athletes.
Some of the Athletes are Plumbers.
Thus,
Some Plumbers are not Weightlifters.

Your conclusion should link Electricians and Geologists.
Some of the Electricians are Bakers.
None of the Bakers are Geologists.
Thus,
Some of the Electricians are not Geologists.

Appendix A1 . 7 (cont.): Experiment 1 Materials: Syllogistic Inference

Mood - BA-CB

Models - 1M

Your conclusion should link Managers and Psychologists.

None of the Musicians are Psychologists.

All of the Managers are Musicians.

Thus,

None of the Managers are Psychologists.

Your conclusion should link Hairdressers and Bankers.

All of the Cashiers are Bankers.

Some of the Hairdressers are Cashiers.

Thus,

Some of the Hairdressers are Bankers.

Your conclusion should link Designers and Gardeners.

All of the Sailors are Designers.

None of the Gardeners are Sailors.

Thus,

None of the Designers are Gardeners.

Your conclusion should link Clowns and Divers.

All of the Policemen are Divers.

All of the Clowns are Policemen.

Thus,

All of the Clowns are Divers.

Models - MM

Your conclusion should link Bricklayers and Singers.

None of the Teachers are Singers.

Some of the Bricklayers are Teachers.

Thus,

Some of the Bricklayers are not Singers.

Your conclusion should link Shopkeepers and Butchers.

Some of the Fishmongers are Shopkeepers.

None of the Butchers are Fishmongers.

Thus,

Some of the Shopkeepers are not Butchers.

Mood - AB-CB

Models - 1M

Your conclusion should link Managers and Plumbers.

None of the Managers are Geologists.

All of the Plumbers are Geologists.

Thus,

None of the Managers are Plumbers.

Appendix A1 . 7 (cont.): Experiment 1 Materials: Syllogistic Inference

Your conclusion should link Carpenters and Bakers.
All of the Carpenters are Electricians.
None of the Bakers are Electricians.
Thus,
None of the Carpenters are Bakers.

Models - MM

Your conclusion should link Athletes and Cabdrivers.
None of the Cabdrivers are Weightlifters.
Some of the Athletes are Weightlifters.
Thus,
Some of the Athletes are not Cabdrivers.

Your conclusion should link Secretaries and Shopkeepers.
Some of the Secretaries are Cashiers.
None of the Shopkeepers are Cashiers.
Thus,
Some of the Secretaries are not Shopkeepers.

Mood - BA-BC

Models - 1M

Your conclusion should link Builders and Teachers.
Some of the Fishmongers are Builders.
All of the Fishmongers are Teachers.
Thus,
Some of the Builders are Teachers.

Your conclusion should link Psychologists and Musicians.
All of the Hairdressers are Psychologists.
Some of the Hairdressers are Musicians.
Thus,
Some of the Psychologists are Musicians.

Models - MM

Your conclusion should link Bricklayers and Butchers.
None of the Bankers are Butchers.
Some of the Bankers are Bricklayers.
Thus,
Some of the Bricklayers are not Butchers.

Appendix A1 . 7 (cont.): Experiment 1 Materials: Syllogistic Inference

Your conclusion should link Climbers and Athletes.

Some of the Welders are Climbers.

None of the Welders are Athletes.

Thus,

Some of the Climbers are not Athletes.

A sub-set of these problems (16) was used in Experiment 3 (verbal presentation).

Content Substitution in Experiment 3 (visual presentation)

athletes

farmers

pilots

managers

singers

clowns

journalists

bankers

dancers

butchers

divers

chefs

These four content sets were randomly assigned to four problems each (a sub-set of the problems seen above).

Appendix A1 . 8: Experiment 1 Instructions: Syllogistic Inference

In this part of the experiment you will be asked to solve 20 syllogisms. A syllogism is a pair of statements providing you with information about the relationships between three classes. The three classes will be labelled by professions (e.g. Bakers, Painters etc.). On the basis of the information provided by the two statements, you will be asked to draw a conclusion between two of the classes. **One of the classes is common to both statements; your task is to produce a conclusion linking the other two classes.**

The 20 syllogisms will be read out loud by the experimenter. Each one will be read twice, and at a reasonable pace. You will be reminded which two terms need to be linked in the conclusion each time the problem is read.

For example, the experimenter might read out the following two statements:-

All of the Painters are Bakers.

All of the Bakers are Farmers.

In each of the 20 problems you will be told which terms need to be linked. In the above example, you would be asked to find a conclusion that links the terms 'Painters' and 'Farmers'. So, for the above example you may draw the conclusion:-

All of the Painters are Farmers.

You would write this conclusion down in the space provided on the response sheet.

YOU SHOULD REMEMBER THAT YOUR CONCLUSION MUST ALWAYS CONTAIN ONE OF THE FOLLOWING QUANTIFIERS:

'all.....are.....' (as in the above example).

'some.....are.....',

'none.....are.....',

and 'some.....are not.....'.

You also have the option of writing 'no valid conclusion'.

Appendix A1 . 8 (cont.): Experiment 1 Instructions: Syllogistic Inference

So for the previous example, these are all of the conclusions you might have produced:-

All of the Painters are Farmers.

All of the Farmers are Painters.

Some of the Painters are Farmers.

Some of the Farmers are Painters.

None of the Painters are Farmers.

None of the Farmers are Painters.

Some of the Painters are not Farmers.

Some of the Farmers are not Painters.

No valid conclusion.

Please take your time and be certain you have the logically correct answer before stating it.

If you have any questions, please ask them now, as the experimenter cannot answer after the task has commenced. Please keep these instructions in front of you in case you need to refer to them later on. You must not make notes or draw diagrams of any kind to aid you in this task.

Practice problem

Please write a conclusion in the space provided. Your conclusion should link Doctors and Bus Drivers. If you think no conclusion is possible, please write 'no valid conclusion'.

Remember that your conclusion must include one of the following quantifiers:

All, No, Some, or Some.....not.

Some of the Doctors are Bank Managers.

All of the Bank Managers are Bus Drivers.

Appendix A2 . 1: Experiment 4 Materials: Temporal Inference

One Model Problems

Passengers board the London train before passengers board the Glasgow train. L G M
Passengers board the Manchester train after passengers board the Glasgow train. B P
Passengers board the Bristol train while passengers board the London train.
Passengers board the Plymouth train while Passengers board the Manchester train.
Hence, what is the relation between passengers boarding the Bristol train and the Plymouth train?
Passengers board the Bristol train before they board the Plymouth train.

Tom drinks coffee after he has a shave. B S C
He eats a biscuit before he has a shave. M T
He reads a magazine while he eats a biscuit.
He watches television while he has a shave.
Hence, what is the relation between Tom reading a magazine and Tom watching television?
Tom reads a magazine before he watches television.

Tom has a shave after he drinks coffee. C S B
He eats a biscuit after he has a shave. T M
He reads a magazine while he has a shave.
He watches television while he drinks coffee.
Hence, what is the relation between Tom reading a magazine and Tom watching television?
Tom reads a magazine after he watches television.

The lever is pressed before the alarm rings. L A G
The gate opens after the alarm rings. D B
The bulb goes on while the gate opens.
The door shuts while the lever is pressed.
Hence, what is the relation between the bulb going on and the door shutting?
The bulb goes on after the door shuts.

The gate opens after the alarm rings. L A G
The lever is pressed before the alarm rings. B D
The bulb goes on while the alarm rings.
The door shuts while the gate opens.
Hence, what is the relation between the bulb going on and the door shutting?
The bulb goes on before the door shuts.

Passengers board Flight 106 after passengers board Flight 204. 2 1 3
Passengers board Flight 307 after passengers board Flight 106. 5 4
Passengers board Flight 403 while passengers board Flight 307.
Passengers board Flight 508 while passengers board Flight 106.
Hence, what is the relation between passengers boarding Flight 403 and 508?
Passengers board Flight 403 after they board Flight 508.

The lever is pressed before the alarm rings. G L A
The gate opens before the lever is pressed. B D
The bulb goes on while the gate opens.
The door shuts while the alarm rings.
Hence, what is the relation between the bulb going on and the door shutting?
The bulb goes on before the door shuts.

Appendix A2 . 1 (cont.): Experiment 4 Materials: Temporal Inference

Multi Model Problems

Passengers board Flight 307 after passengers board Flight 204. 1 2 3 2 1 3
Passengers board Flight 106 before passengers board Flight 307. 4 5 4 5
Passengers board Flight 403 while passengers board Flight 106.
Passengers board Flight 508 while passengers board Flight 307.
Hence, what is the relation between passengers boarding Flight 403 and 508?
Passengers board Flight 403 before they board Flight 508.

Passengers board the Manchester train before passengers board the London train. M L G M G L
Passengers board the Glasgow train after passengers board the Manchester train. P B P B
Passengers board the Bristol train while passengers board the Glasgow train.
Passengers board the Plymouth train while Passengers board the Manchester train.
Hence, what is the relation between passengers boarding the Bristol train and the Plymouth train?
Passengers board the Bristol train after they board the Plymouth train.

Tom drinks a coffee after he watches television. T C S T S C
He has a shave after he watches television. M B M B
He eats a biscuit while he has a shave.
He reads a magazine while he watches television.
Hence, what is the relation between Tom eating a biscuit and Tom reading a magazine?
Tom eats a biscuit after he reads a magazine.

Passengers board Flight 307 before passengers board Flight 204. 3 2 1 3 1 2
Passengers board Flight 106 after passengers board Flight 307. 5 4 5 4
Passengers board Flight 403 while passengers board Flight 106.
Passengers board Flight 508 while passengers board Flight 307.
Hence, what is the relation between passengers boarding Flight 403 and 508?
Passengers board Flight 403 after they board Flight 508.

Tom drinks coffee before he watches television. S C T C S T
He has a shave before he watches television. B M B M
He eats a biscuit while he has a shave.
He reads a magazine while he watches television.
Hence, what is the relation between Tom eating a biscuit and Tom reading a magazine?
Tom eats a biscuit before he reads a magazine.

Passengers board the Glasgow train after passengers board the London train. M L G L M G
Passengers board the Manchester train before passengers board the Glasgow train. B P B P
Passengers board the Bristol train while passengers board the Manchester train.
Passengers board the Plymouth train while Passengers board the Glasgow train.
Hence, what is the relation between passengers boarding the Bristol train and the Plymouth train?
Passengers board the Bristol train before they board the Plymouth train.

Passengers board the London train before passengers board the Glasgow train. L M G M L G
Passengers board the Manchester train before passengers board the Glasgow train. B P B P
Passengers board the Bristol train while passengers board the Manchester train.
Passengers board the Plymouth train while Passengers board the Glasgow train.
Hence, what is the relation between passengers boarding the Bristol train and the Plymouth train?
Passengers board the Bristol train before they board the Plymouth train.

Appendix A2 . 1 (cont.): Experiment 4 Materials: Temporal Inference

Indeterminate Problems

Passengers board Flight 307 after passengers board Flight 204. 1 2 3 2 1 3
Passengers board Flight 106 before passengers board Flight 307. 4 5 5 4
Passengers board Flight 403 while passengers board Flight 106.
Passengers board Flight 508 while passengers board Flight 204.
Hence, what is the relation between passengers boarding Flight 403 and 508?
Not enough information.

The alarm rings before the gate opens. A G L A L G
The lever is pressed after the alarm rings. B D D B
The bulb goes on while the gate opens.
The door shuts while the lever is pressed.
Hence, what is the relation between the bulb going on and the door shutting?
Not enough information.

Appendix A2 . 2: Experiment 4 Instructions: Temporal Inference

In this part of the experiment your task is to answer a series of questions based on the information about the order of events given in a description. A problem will appear on the screen describing the order of events of a set of five activities or actions. Shortly afterwards, a question will also appear on the screen asking you to relate two of the activities or actions within the set. There will be 16 descriptions using the following terms to describe the order in time in which each activity or action occurs:-

before

after

while

Your task is to link two of the activities or actions in the set using the above relations. For example, passengers may board the Bristol train **before** the Plymouth train; passengers may board the Bristol train **after** the Plymouth train; or passengers may board the Bristol train **while** they board the Plymouth train (hypothetically, people can't be in two places at once, but for this exercise it is not necessary to have believable content). For some of the descriptions you may feel there is not enough information to determine the relation between the activities or actions; in which case you should answer "Not enough information".

The screen will also display a blank box in which you must type your response to each problem. For example, if you thought the answer was "Passengers board the Bristol train before the Plymouth train" then this is what you should type. There will then be a button to press to move onto the next problem. If you understand these instructions, please press the bar below and a practice problem will appear for you to have a go at.

Practice problem

Passengers board Flight 106 after passengers board Flight 204

Passengers board Flight 307 before passengers board Flight 204

Passengers board Flight 403 while passengers board Flight 307

Passengers board Flight 508 while passengers board Flight 204

Hence, what is the relation between passengers boarding Flight 403 and 508?

The answer should be that Passengers board Flight 403 before they board Flight 508.

Appendix A3 . 1: Experiment 5 Materials: Abstract Reasoning

Multi-Step Problems

If there is a P, then there is a C
There is a P
It is not true that there is both a C and an M
Thus, there's not an M

There is a Y or an L
There's not a Y
If there is either an L or an R, then there's not a W
Thus, there's not a W

If there is an R, then there is an F
If there is an W, then there is an L
There is an R or a W
If there is either an F or an L, then there's not a Z
Thus, there's not a Z

It is not true that there is both a C and an H
There is a C
There's not a P
Thus, there's not an H or a P

There is a P, and there is a Q or an R
If there is both a P and a Q, then there is an S
If there is both a P and an R, then there is a T
Thus, there may or may not be an S and a T

There is a B or a Z
There's not a Z
It is not true that there is both a B and an R
Thus, there's not an R

It is not true that there is both a K and an L
It is false that there is not a K
Thus, there's not an L

There is an L or a W
If there is an L, then there is not an E
If there is a W, then there is not an E
There is an E or an O
Thus, there is an O

There is an E or an X
If there is an E, then there is not an H
If there is an X, then there is not an H
There is an H or a T
Thus, there is a T

Appendix A3 . 2: Experiment 5 Instructions: Abstract Reasoning

This task is concerned with people's ability to reason logically with sentences in various forms. You will be presented with a total of 9 problems in this booklet. For each of these problems you are given a set of statements. for instance:-

If there is both an N and an I, then there's not a B

It is false that there's not an N

There is an I

Each problem concerns the presence or absence of letters on an imaginary black-board. The set of statements contains some facts about the black-board. The facts or premises contain the information that is known about the black-board and you should accept them at face value.

Your task in each case is to decide on a conclusion that necessarily follows from the statements about the black-board. A conclusion is necessary if it must be true, given that the statements are true. A list of the possible answers is presented below each problem. You must tick one of the response conclusions for each set of statements. For instance, in the example shown you might tick:-

There's not a B

If you think that no conclusion necessarily follows given that the statements are true, then you must tick the relevant box, for instance, that the letter B may or may not be there. You will have 20 seconds to answer each problem. The experimenter will let you know when to continue onto the next problem. Please do not make diagrams or notes during the task, and do not go back through the booklet once you have answered a problem. If you have any questions about this task, please ask the experimenter now. If you understand these instructions, please wait before beginning the practice problem.

Practice problem

If there is either a K or an O, then there is an N

It is false that there's not a K

Thus, there is an N

Appendix A3 . 3: Experiment 5 Materials: Conditional Inference

Modus Ponens

If p then q
p
Therefore q

If the letter is B then the number is 7
The letter is B
Therefore the number is 7

If p then not q
p
Therefore not q

If the letter is A then the number is not 8
The letter is A
Therefore the number is not 8

If not p then q
not p
Therefore q

If the letter is not C then the number is 4
The letter is not C
Therefore the number is 4

If not p then not q
not p
Therefore not q

If the letter is not D then the number is not 3
The letter is not D
Therefore the number is not 3

Modus Tollens

If p then q
not q
Therefore not p

If the letter is D then the number is 3
The number is not 3
Therefore the letter is not D

If p then not q
q
Therefore not p

If the letter is C then the number is not 4
The number is 4
Therefore the letter is not C

If not p then q
not q
Therefore p

If the letter is not A then the number is 8
The number is not 8
Therefore the letter is A

If not p then not q
q
Therefore p

If the letter is not B then the number is not 7
The number is 7
Therefore the letter is B

Appendix A3 . 3 (cont.): Experiment 5 Materials: Conditional Inference

Denial of the Antecedent

If p then q
not p
Therefore not q

If the letter is E then the number is 9
The letter is not E
Therefore the number is not 9

If p then not q
not p
Therefore q

If the letter is F then the number is not 2
The letter is not F
Therefore the number is 2

If not p then q
p
Therefore not q

If the letter is not G then the number is 5
The letter is G
Therefore the number is not 5

If not p then not q
p
Therefore q

If the letter is not H then the number is not 1
The letter is H
Therefore the number is 1

Affirmation of the Consequent

If p then q
q
Therefore p

If the letter is H then the number is 1
The number is 1
Therefore the letter is H

If p then not q
not q
Therefore p

If the letter is G then the number is not 5
The number is not 5
Therefore the letter is G

If not p then q
q
Therefore not p

If the letter is not F then the number is 2
The number is 2
Therefore the letter is not F

If not p then not q
not q
Therefore not p

If the letter is not E then the number is not 9
The number is not 9
Therefore the letter is not E

Appendix A3 . 4: Experiment 5 Instructions: Conditional Inference

This task is concerned with people's ability to reason logically with sentences in various forms. You will be presented with a total of 16 problems in this booklet. For each of these problems you are given two statements, for instance:

- 1** **If the letter is M then the number is not 3**
2 **The letter is M**

Each problem concerns an imaginary letter-pair and contains an initial statement or rule (1) which determines which letters may be paired with which numbers. In each case, you must assume that the rule holds and then combine it with the information given in the second statement (2). This will concern either the letter or the number of an imaginary pair.

Your task in each case is to decide on a conclusion that necessarily follows from the statements. A conclusion is necessary if it must be true, given that the statements are true. A list of the possible answers is presented below each problem. You must tick one of the response conclusions for each pair of statements. For instance, in the example shown you might tick:

- 3** **The number is not 3**

If you think that no conclusion necessarily follows given that the statements are true, then you must tick the relevant box, for instance, that the number may or may not be a 3. Please do not make diagrams or notes during the task, and do not go back through the booklet once you have answered a problem. If you have any questions about this task, please ask the experimenter now. If you understand these instructions, please turn over to the first problem out of 16.

Appendix A3 . 5: Experiment 5 Materials: Three-Term Series Spatial Inference

Valid Problems

A is left of B
B is left of C
Thus, A is left of C

E is right of F
F is right of G
Thus, E is right of G

Y is right of X
Z is right of Y
Thus, X is left of Z

K is left of J
L is left of K
Thus, J is right of L

I is left of J
K is right of J
Thus, I is left of K

L is right of M
N is left of M
Thus, N is left of L

T is left of S
T is right of U
Thus, S is right of U

W is right of V
W is left of X
Thus, V is left of X

Invalid Problems

M is left of N
N is right of O
Thus, invalid

R is right of S
S is left of T
Thus, invalid

W is right of V
X is left of W
Thus, invalid

G is left of F
H is right of G
Thus, invalid

P is left of Q
R is left of Q
Thus, invalid

Q is right of R
S is right of R
Thus, invalid

C is left of B
C is left of D
Thus, invalid

G is right of F
G is right of H
Thus, invalid

Appendix A3 . 6: Experiment 5 Instructions: Three-Term Series Spatial Inference

In this task you will be asked about the relations between letters in a set. You are presented with a booklet of 16 problems each describing the layout of a set of three letters. The descriptions will use the following terms to describe the layouts of the letters:-

left

right

Your task is to link two of the letters in the set using the above relations. You should imagine the three letters as if they were in front of you and then apply these relations to them. For some of the descriptions you may feel there is not enough information to determine the relation between the letters; in which case you should answer “Not valid”.

To remind you of the responses you might give, a list of all the possible answers is given below each problem. You must tick the answer that you think is correct. For example, if you thought the answer was “F is left of P” then this is what you should tick. You will have 10 seconds to answer each problem. The experimenter will let you know when to continue onto the next problem. You must not make diagrams or notes of any kind and do not look back to questions that you have already answered. If you understand these instructions, please wait before beginning the first problem.

Appendix A3 . 7: Experiment 5 Materials: Multiple Quantification

Mood - AB-BC

Models - 1M

All of the Js are in the same place as all of the Ks

All of the Ks are in the same place as all of the Ls

Thus

All of the Js are in the same place as all of the Ls

All of the Ps are in the same place as some of the Qs

All of the Qs are in the same place as all of the Rs

Thus

All of the Ps are in the same place as all of the Rs

All of the As are in the same place as some of the Bs

All of the Bs are in the same place as some of the Cs

Thus

All of the As are in the same place as some of the Cs

None of the Ts are in the same place as any of the Us

Some of the Us are in the same place as all of the Vs

Thus

None of the Ts are in the same place as any of the Vs

None of the Ds are in the same place as any of the Es

All of the Es are in the same place as all of the Fs

Thus

None of the Ds are in the same place as any of the Fs

Models - MM Determinate

None of the Gs are in the same place as any of the Hs

All of the Hs are in the same place as some of the Js

Thus

None of the Gs are in the same place as some of the Js

Models - MM Indeterminate

All of the Xs are in the same place as some of the Ys

Some of the Ys are in the same place as all of the Zs

Thus

Not valid

None of the Bs are in the same place as some of the Cs

Some of the Cs are in the same place as all of the Ds

Thus

Not valid

Appendix A3 . 7 (cont.): Experiment 5 Materials: Multiple Quantification

Mood - BA-CB

Models - 1M

All of the Ws are in the same place as all of the Vs

All of the Xs are in the same place as all of the Ws

Thus

All of the Xs are in the same place as all of the Vs

Some of the Ss are in the same place as all of the Rs

All of the Ts are in the same place as all of the Ss

Thus

All of the Ts are in the same place as all of the Rs

Some of the Fs are in the same place as all of the Es

Some of the Gs are in the same place as all of the Fs

Thus

Some of the Gs are in the same place as all of the Es

Any of the Ls are in the same place as none of the Ks

All of the Ms are in the same place as some of the Ls

Thus

Any of the Ms are in the same place as none of the Ks

Any of the Os are in the same place as none of the Ns

All of the Ps are in the same place as all of the Os

Thus

Any of the Ps are in the same place as none of the Ns

Models - MM Determinate

Any of the Rs are in the same place as none of the Qs

Some of the Ss are in the same place as all of the Rs

Thus

Some of the Ss are in the same place as none of the Qs

Models - MM Indeterminate

Some of the Ds are in the same place as all of the Cs

All of the Es are in the same place as some of the Ds

Thus

Not valid

Some of the Gs are in the same place as none of the Fs

All of the Hs are in the same place as some of the Gs

Thus

Not valid

Appendix A3 . 8: Experiment 5 Instructions: Multiple Quantification

In this task you will be asked to solve 16 multiple quantification problems, based on the same logic of syllogisms. A syllogism is a pair of statements providing you with information about the relationships between three classes. The three classes will be labelled by letters (e.g. Bs, Ps and Cs etc.). On the basis of the information provided by the two statements, you will be asked to draw a conclusion between two of the letters. **One of the letters is common to both statements; your task is to produce a conclusion linking the other two letters.**

You will be presented with a total of 16 problems in this booklet, on separate pages. For each of these problems you are given two statements, for instance:-

Some of the Ps are in the same place as some of the Bs

Some of the Bs are in the same place as some of the Fs

Thus...

In the above example, you are asked to find a logically valid conclusion that links the letters 'P' and 'F'. A conclusion is logically valid if it must follow given that the statements are true. If a conclusion *could* but does not *have* to or necessarily follow, then it is not logically valid. You will be given a response conclusion with two gaps indicating where you must fill in a response. You must write the answer that you believe is the correct, and strongest conclusion to each problem (i.e. If the answer could be some... *or* all..., tick the stronger conclusion, all...). So, in the example given, you would be shown the following conclusion:-

___ of the Ps are in the same place as ___ of the Fs

OR

No Valid response

Appendix A3 . 8 (cont.): Experiment 5 Instructions: Multiple Quantification

Your task is to fill in the two gaps with the correct quantifiers (all, some, none or any). for instance for the example above you might fill in the gaps with - **some** of the Ps are in the same place as **some** of the Fs. If you think that no one conclusion logically follows, then tick no valid conclusion.

YOU SHOULD REMEMBER THAT THE CONCLUSION WILL ALWAYS CONTAIN TWO OF THE FOLLOWING QUANTIFIERS:

All of the are in the same place as

Some of the are in the same place as

None of the are in the same place as

Any of the are in the same place as

You also have the option of ticking 'no valid conclusion'

Please take your time and be certain you have the logically correct answer before stating it. If you have any questions, please ask them now, as the experimenter cannot answer after the task has commenced. You must not make notes or draw diagrams of any kind to aid you in this task. Please do not look back at problems you have already completed. A practice problem will now be given to show you the format of each of the 16 problems.

Practice problem

All of the Cs are in the same place as some of the Ds

None of the Ps are in the same place as any of the Cs

Thus

None of the Ps are in the same place as some of the Ds

APPENDIX B: Summary Tables Of Statistical Analyses: t-tests, ANOVAS,

Correlations and Factor Analyses

Appendix B1 . 1: Experiment 1 Statistics: Summary Tables of t-tests

Summary table of t-test to investigate syllogistic inference, model type: 1M vs. MM

	Mean	S.D.	N	Diff.	S.D. Diff	t	df	p-level
Syll 1M	8.778	1.363						
Syll MM	1.200	2.073	45	7.578	2.607	19.501	44	.000001

Summary table of t-test to investigate spatial inference, model type: 1M vs. MM

	Mean	S.D.	N	Diff.	S.D. Diff	t	df	p-level
Spat 1M	8.200	2.564						
Spat MM	5.978	2.572	45	2.222	2.883	5.170	44	.000005

Appendix B1 . 2: Experiment 4 Statistics: Summary Tables of t-tests

Summary table of t-test to investigate temporal inference vs. spatial inference

	Mean	S.D.	N	Diff.	S.D. Diff	t	df	p-level
Temporal	9.511	3.741						
Spatial	8.915	3.374	47	.596	2.856	1.430	46	.15952

Summary table of t-test to investigate temporal inference, model type: 1M vs. MM

	Mean	S.D.	N	Diff.	S.D. Diff	t	df	p-level
Temp 1M	4.660	1.773						
Temp MM	4.512	1.999	47	.149	1.268	.805	46	.42481

Summary table of t-test to investigate temporal inference, model type: 1M vs. IND

	Mean	S.D.	N	Diff.	S.D. Diff	t	df	p-level
Temp 1M	4.660	1.773						
Temp IND	.319	.515	47	4.340	1.659	17.940	46	.000001

Summary table of t-test to investigate temporal inference, model type: MM vs. IND

	Mean	S.D.	N	Diff.	S.D. Diff	t	df	p-level
Temp MM	4.511	1.999						
Temp IND	.319	.515	47	4.191	1.952	14.721	46	.000001

Appendix B1 . 2 (cont.): Experiment 4 Statistics: Summary Tables of t-tests

Summary table of t-test to investigate spatial inference, model type: 1M vs. MM

	Mean	S.D.	N	Diff.	S.D. Diff	t	df	p-level
Spat 1M	4.468	2.009						
Spat MM	4.447	1.931	47	.021	2.038	.072	46	.94324

Appendix B1 . 3: Experiment 5 Statistics: Summary Tables of t-tests

Summary table of t-test to investigate simple spatial inference, model type: 1M vs. Invalid

	Mean	S.D.	N	Diff.	S.D. Diff	t	df	p-level
Simspat 1M	5.275	1.576						
Simspat INV	2.510	2.203	51	2.765	2.397	8.238	50	.000001

Summary table of t-test to investigate complex spatial inference, model type: 1M vs. MM

	Mean	S.D.	N	Diff.	S.D. Diff	t	df	p-level
Spat 1M	5.706	2.110						
Spat MM	4.059	2.044	51	1.647	2.331	5.046	50	.000006

Summary table of t-test to investigate syllogistic inference, model type: 1M vs. MM

	Mean	S.D.	N	Diff.	S.D. Diff	t	df	p-level
Syll 1M	6.608	1.650						
Syll MM	1.765	2.665	51	4.843	2.838	12.187	50	.000001

Summary table of t-test to investigate multiple quantification, model type: 1M vs. MM

	Mean	S.D.	N	Diff.	S.D. Diff	t	df	p-level
Multi 1M	6.098	1.330						
Multi MM	1.608	1.266	51	4.490	1.713	18.718	50	.000001

Appendix B1 . 4: Experiment 6 Statistics: Summary Tables of t-tests

Summary table of t-test to investigate simple spatial inference, model type: 1M vs. Invalid

	Mean	S.D.	N	Diff.	S.D. Diff	t	df	p-level
Simspat 1M	5.271	1.554						
Simspat INV	2.833	2.137	48	2.438	2.305	7.325	47	.000001

Summary table of t-test to investigate complex spatial inference, model type: 1M vs. MM

	Mean	S.D.	N	Diff.	S.D. Diff	t	df	p-level
Spat 1M	5.271	1.854						
Spat MM	3.938	2.138	48	1.333	2.478	3.727	47	.000519

Summary table of t-test to investigate syllogistic inference, model type: 1M vs. MM

	Mean	S.D.	N	Diff.	S.D. Diff	t	df	p-level
Syll 1M	6.292	1.570						
Syll MM	1.729	2.181	48	4.563	2.333	13.549	47	.000001

Summary table of t-test to investigate multiple quantification, model type: 1M vs. MM

	Mean	S.D.	N	Diff.	S.D. Diff	t	df	p-level
Multi 1M	5.250	1.591						
Multi MM	1.292	1.110	48	3.958	1.957	14.017	47	.000001

Appendix B1 . 5: Experiment 2 Statistics: Summary Table of ANOVA

Summary table of ANOVA to investigate modality and model type

Summary of all effects where 1 = modality. 2 = model.

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
1	1	41.19	46	1.9306	21.3359	.000031
2	1	129.45	46	2.5989	49.8067	.000000
12	1	3.60	46	.8566	4.1976	.046211

Appendix B1 . 6: Experiment 3 Statistics: Summary table of ANOVA

Summary table of ANOVA to investigate modality and model type

Summary of all effects where 1 = modality, 2 = model.

Effect	df Effect	MS effect	df Error	MS Error	F	p-level
1	1	8.27	45	1.9996	4.1339	.047956
2	1	780.66	45	4.6798	166.8132	.000000
12	1	.66	45	1.1243	.5849	.448383

Appendix B1 . 7: Experiment 1 Statistics: Correlation Matrix for all Measures, n = 45 (two-tailed significance level (*) p < .05)

1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	
	1																					
	.91*	1																				
	.22	.21	1																			
	.22	.23	.95*	1																		
	.46*	.56*	.28	.28	1																	
	.47*	.59*	.14	.11	.91*	1																
	-.17	-.17	.03	.05	-.16	-.29	1															
	.37*	.40*	.16	.21	.23	.30*	-.32*	1														
	.43*	.42*	.24	.31*	.39*	.37*	-.28	.75*	1													
	.39*	.39*	.21	.25	.18	.23	-.32*	.75*	.77*	1												
	-.12	-.24	-.11	-.12	-.39*	-.41*	.49*	-.24	-.27	-.11	1											
	.39*	.39*	.11	.08	.30*	.31*	-.25	.52*	.55*	.50*	-.28	1										
	.43*	.50*	.13	.10	.32*	.36*	-.28	.14	.24	.24	-.17	.48*	1									
	.49*	.51*	.23	.18	.42*	.38*	-.17	.25	.42*	.28	-.16	.50*	.83*	1								
	.22	.31*	-.01	-.01	.10	.21	-.29	-.01	-.03	.11	-.12	.30*	.83*	.37*	1							
	.29	.40*	.40*	.43*	.35*	.30*	-.05	.39*	.37*	.19	-.15	.10	.27	.42*	.03	1						
	.20	.24	.04	.04	.25	.22	.04	.24	.45*	.23	.01	.22	.12	.35*	-.15	.48*	1					
	.19	.29*	.42*	.46*	.23	.19	-.09	.28	.12	.07	-.18	-.03	.22	.24	.13	.82*	-.11	1				
	-.05	.06	.31*	.26	.23	.19	-.08	.16	-.05	-.06	-.29	-.00	-.12	-.06	-.14	.28	-.14	.41*	1			
	.16	.12	.10	.12	.11	.10	-.25	.14	.19	.03	-.18	.22	-.09	-.10	-.05	-.04	-.16	.06	.11	1		
	.12	.20	-.12	-.10	.28	.32*	-.14	.07	.05	-.03	-.20	.03	-.12	-.17	-.03	-.13	-.01	-.14	.07	.52*	1	

1 = word span (normal) 7 = verbal span error 13 = spatial inference (total) 19 = spatial ability
2 = word span (global) 8 = letter span (normal) 14 = spatial inference (1M) 20 = verbal ability
3 = arrow span (normal) 9 = letter span (limits) 15 = spatial inference (MM) 21 = reading rate
4 = arrow span (global) 10 = letter span (global) 16 = syllogistic inference (total)
5 = verbal span (normal) 11 = letter span error 17 = syllogistic inference (1M)
6 = verbal span (global) 12 = TTT span 18 = syllogistic inference (MM) (level of significance for two-tailed test p < .10, r = .25; p < .05, r = .30; p < .01, r = .39; p < .001, r = .49)

Appendix B1. 8: Experiment 2 Statistics: Correlation Matrix for all Measures, n = 47 (two-tailed significance level (*) p < .05)

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	1												
2	.95*	1											
3	.29*	.32*	1										
4	.24	.28	.89*	1									
5	.53*	.56*	.18	.14	1								
6	.59*	.64*	.14	.10	.93*	1							
7	.29*	.40*	.30*	.19	.29*	.41*	1						
8	.24	.30*	.51*	.52*	.32*	.36*	.46*	1					
9	.33*	.36*	.54*	.55*	.34*	.32*	.34*	.86*	1				
10	.11	.18	.38*	.39*	.25	.32*	.47*	.91*	.56*	1			
11	.09	.14	.31*	.25	.19	.24	.41*	.69*	.64*	.59*	1		
12	.09	.12	.40*	.36*	.19	.19	.30*	.58*	.67*	.39*	.87*	1	
13	.07	.12	.15	.09	.15	.24	.42*	.63*	.46*	.63*	.89*	.55*	1

- 1 = word span (normal) 8 = spatial inference, verbal (total)
- 2 = word span (global) 9 = spatial inference, verbal (IM)
- 3 = arrow span (normal) 10 = spatial inference, verbal (MM)
- 4 = arrow span (global) 11 = spatial inference, visual (total)
- 5 = verbal span (normal) 12 = spatial inference, visual (IM)
- 6 = verbal span (global) 13 = spatial inference, visual (MM)
- 7 = TTT span

(level of significance for two-tailed test p < .10, r = .24; p < .05, r = .29; p < .01, r = .37; p < .001, r = .46)

Appendix B1 .9: Experiment 3 Statistics: Correlation Matrix for all Measures, n = 46 (two-tailed significance level (*) p < .05)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1																
2	.92*	1															
3	-.00	.06	1														
4	.14	.26	.85*	1													
5	.47*	.65*	.15	.27	1												
6	.50*	.66*	.12	.31*	.94*	1											
7	.36*	.33*	.33*	.43*	.39*	.44*	1										
8	.37*	.41*	.36*	.47*	.47*	.48*	.71*	1									
9	.36*	.40*	.41*	.55*	.43*	.49*	.83*	.81*	1								
10	-.23	-.26	-.42*	-.24	-.29	-.32*	-.26	-.33*	-.28	1							
11	.22	.30*	.32*	.40*	.37*	.35*	.35*	.49*	.49*	-.17	1						
12	.48*	.57*	.31*	.50*	.53*	.45*	.36*	.46*	.40*	-.20	.45*	1					
13	.33*	.38*	.37*	.43*	.35*	.27	.32*	.48*	.38*	-.36*	.42*	.77*	1				
14	.42*	.50*	.11	.35*	.46*	.42*	.24	.24	.24	.04	.29	.79*	.21	1			
15	.34*	.44*	.18	.45*	.48*	.50*	.43*	.47*	.55*	-.06	.21	.66*	.41*	.61*	1		
16	.17	.19	.05	.22	.18	.21	.18	.33*	.32*	-.09	.12	.42*	.50*	.17	.75*	1	
17	.36*	.49*	.22	.48*	.55*	.56*	.49*	.43*	.56*	-.02	.21	.63*	.23	.75*	.89*	.37*	1

1 = word span (normal) 7 = letter span (normal)

2 = word span (global) 8 = letter span (limits)

3 = arrow span (normal) 9 = letter span (global)

4 = arrow span (global) 10 = letter span error

5 = verbal span (normal) 11 = IIT span

6 = verbal span (global) 12 = syllogistic inference, verbal (total)

13 = syllogistic inference, verbal (IM)

14 = syllogistic inference, verbal (MM)

15 = syllogistic inference, visual (total)

16 = syllogistic inference, visual (IM)

17 = syllogistic inference, visual (MM)

(level of significance for two-tailed test p < .10, r = .24; p < .05, r = .29; p < .01, r = .37; p < .001, r = .46)

Appendix B1 . 10: Experiment 4 Statistics: Correlation Matrix for all Measures, n = 47 (two-tailed significance level (*) p < .05)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1													
2	.96*	1												
3	.33*	.36*	1											
4	.33*	.38*	.91*	1										
5	.65*	.68*	.45*	.43*	1									
6	.70*	.71*	.45*	.45*	.95*	1								
7	.22	.18	.44*	.38*	.21	.24	1							
8	.29*	.31*	.55*	.47*	.29*	.27	.41*	1						
9	.26	.28	.51*	.46*	.32*	.31*	.41*	.86*	1					
10	.23	.26	.43*	.35*	.18	.14	.30*	.85*	.47*	1				
11	.33*	.32*	.42*	.40*	.27	.31*	.52*	.68*	.63*	.54*	1			
12	.38*	.37*	.39*	.36*	.26	.30*	.49*	.66*	.56*	.57*	.93*	1		
13	.21	.19	.35*	.34*	.17	.21	.48*	.60*	.59*	.43*	.94*	.78*	1	
14	.31*	.33*	.39*	.36*	.47*	.47*	.28	.37*	.38*	.25	.42*	.36*	.22	1

1 = word span (normal) 8 = spatial inference, visual (total)

2 = word span (global) 9 = spatial inference, visual (IM)

3 = arrow span (normal) 10 = spatial inference, visual (MM)

4 = arrow span (global) 11 = temporal inference, visual (total)

5 = verbal span (normal) 12 = temporal inference, visual (IM)

6 = verbal span (global) 13 = temporal inference, visual (MM)

7 = III span 14 = temporal inference, visual (IND)

(level of significance for two-tailed test p < .10, r = .24; p < .05, r = .29; p < .01, r = .37; p < .001, r = .46)

Appendix B1 . 11: Experiment 5 Statistics: Correlation Matrix for all Measures, n = 51 (two-tailed significance level (*) p < .05)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
1	1																		
2	.85*	1																	
3	.24	.13	1																
4	.19	.17	.93*	1															
5	.24	.39*	.32*	.33*	1														
6	.24	.38*	.16	.18	.70*	1													
7	.16	.25	.32*	.31*	.86*	.23	1												
8	.08	.08	.25	.27	.30*	.26	.22	1											
9	.13	.11	.23	.28*	.26	.25	.17	.83*	1										
10	-.00	.03	.18	.16	.23	.17	.19	.82*	.37*	1									
11	.28	.32*	.26	.26	.45*	.31*	.39*	.23	.22	.17	1								
12	.29*	.21	.23	.19	.40*	.15	.43*	.06	.16	-.06	.45*	1							
13	.13	.04	.23	.20	.28*	.15	.27	.48*	.43*	.35*	.16	.34*	1						
14	.03	-.11	.31*	.30*	.08	-.00	.12	.30*	.28*	.21	.27	.29*	.64*	1					
15	.14	.12	.10	.06	.30*	.20	.27	.42*	.38*	.32*	.04	.25	.88*	.20	1				
16	.15	.23	.29*	.32*	.53*	.33*	.49*	.36*	.39*	.21	.19	.26	.22	.10	.21	1			
17	.09	.13	.18	.17	.27	.13	.27	.36*	.34*	.26	-.00	.05	.12	.04	.14	.77*	1		
18	.14	.21	.26	.32*	.54*	.37*	.47*	.18	.24	.06	.30*	.34*	.20	.12	.19	.74*	.13	1	

1 = word span (normal) 8 = complex spatial inference (total) 15 = syllogistic inference (MM)
2 = word span (global) 9 = complex spatial inference (1M) 16 = multiple quantification (total)
3 = arrow span (normal) 10 = complex spatial inference (MM) 17 = multiple quantification (1M)
4 = arrow span (global) 11 = abstract inference 18 = multiple quantification (MM)
5 = simple spatial inference (total) 12 = conditional inference
6 = simple spatial inference (1M) 13 = syllogistic inference (total) (level of significance for two-tailed test p < .10, r = .24;
7 = simple spatial inference (INV) 14 = syllogistic inference (1M) p < .05, r = .29; p < .01, r = .37; p < .001, r = .46)

Appendix B1 . 12: Experiment 6 Statistics: Correlation Matrix for all Measures, n = 48 (two-tailed significance level (*) p < .05)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	
1	1																					
2	.94*	1																				
3	.20	.26	1																			
4	.37*	.42*	.87*	1																		
5	.53*	.55*	.22	.30*	1																	
6	.64*	.69*	.23	.33*	.94*	1																
7	.13	.11	.32*	.33*	.01	.08	1															
8	.32*	.39*	.44*	.53*	.25	.28	.26	1														
9	.16	.19	.26	.41*	.18	.19	.21	.71*	1													
10	.32*	.40*	.42*	.44*	.21	.24	.20	.86*	.25	1												
11	.04	.07	.11	.28	-.17	-.14	.14	.44*	.45*	.28	1											
12	.13	.20	.18	.29*	-.11	-.05	.33*	.59*	.53*	.43*	.75*	1										
13	-.06	-.07	.01	.16	-.15	-.16	-.08	.14	.20	.04	.82*	.24	1									
14	.18	.24	.02	.05	.10	.14	-.20	.31*	.08	.38*	.00	.01	-.00	1								
15	.21	.24	.12	.18	.03	.13	.20	.23	.09	.26	.16	.38*	-.10	.31*	1							
16	-.02	.04	.05	.13	.17	.21	.34*	.30*	.40*	.12	.25	.37*	.04	.19	.10	1						
17	-.13	-.08	.23	.24	.06	.09	.36*	.32*	.45*	.11	.26	.24	.17	.11	-.09	.71*	1					
18	.07	.12	-.09	.00	.18	.22	.21	.18	.22	.09	.15	.34*	-.07	.19	.21	.86*	.26	1				
19	.02	.00	.21	.26	-.15	-.14	.52*	.47*	.38*	.37*	.36*	.49*	.10	.08	.10	.41*	.43*	.25	1			
20	.01	.05	.27	.28	-.05	-.01	.38*	.36*	.34*	.25	.30*	.42*	.08	.06	.08	.42*	.51*	.21	.82*	1		
21	.01	-.06	-.02	.04	-.19	-.23	.35*	.29*	.16	.28	.19	.25	.06	.04	.07	.11	.01	.14	.56*	-.02	1	

1 = word span (normal) 7 = TTT span 13 = complex spatial (MM) 19 = multiple quantifiers (tot)
 2 = word span (global) 8 = simple spatial inference (total) 14 = abstract inference 20 = multiple quantifiers (IM)
 3 = arrow span (normal) 9 = simple spatial (IM) 15 = conditional inference 21 = multiple quantifiers (MM)
 4 = arrow span (global) 10 = simple spatial (INV) 16 = syllogistic inference (tot)
 5 = verbal span (normal) 11 = complex spatial inference (tot) 17 = syllogistic inference (IM) (level of significance for two-tailed test p < .10, r = .24;
 6 = verbal span (global) 12 = complex spatial (IM) 18 = syllogistic inference (MM) p < .05, r = .29; p < .01, r = .37; p < .001, r = .46)

**Appendix B1 . 13: Experiments 1, 2 and 3 Statistics: Multi-Group Factor Analysis,
Three-Factor Model**

EQS, A STRUCTURAL EQUATION PROGRAM MULTIVARIATE SOFTWARE.

INC. COPYRIGHT BY P.M. BENTLER

VERSION 5.7b (C) 1985 - 1998.

PROGRAM CONTROL INFORMATION

```
1 /TITLE
2 Experiment 1 - Three factors
3 /SPECIFICATIONS
4 VARIABLES = 11; CASES = 45; METHOD = ML; MATRIX = CORR;
5 ANALYSIS = CORR; GROUP = 3;
6 /LABELS
7 V1 = ARROW; V2 = WORD; V3 = LETTER; V4 = TTT; V5 = VERBAL;
8 V6 = SPATver; V7 = SYLLver; V8 = SPATvis; V9 = SYLLvis; V10 = LETTerr;
9 V11 = VERerr;
10 /EQUATIONS
11 V1 = + 4*F2    + 5*F3    + E1;
12 V2 = + 10*F1   + 10*F3   + E2;
13 V3 = + 0.15*F2 + 1*F3    + E3;
14 V4 = + 4*F2    + 4*F3    + E4;
15 V5 = + 8*F1    + 11*F3   + E5;
16 V6 = + 1.5*F1   + 1*F2    + 2*F3    + E6;
17 V7 = + 0.5*F1   + 0.5*F2   + 1.5*F3 + E7;
18 V10 = + *F3    + E10;
19 V11 = + *F3    + E11;
20 /VARIANCES
21 F1 = 1.00;
22 F2 = 1.00;
23 F3 = 1.00;
24 E1 = *;
25 E2 = *;
26 E3 = *;
27 E4 = *;
28 E5 = *;
29 E6 = *;
30 E7 = *;
31 E10 = *;
32 E11 = *;
33 /MATRIX
34 1
35 0.23 1
36 0.31 0.42 1
37 0.08 0.39 0.55 1
38 0.11 0.59 0.37 0.31 1
39 0.10 0.50 0.24 0.48 0.36 1
```


40 0.43 0.40 0.37 0.10 0.30 0.27 1
 41 0.00 0.00 0.00 0.00 0.00 0.00 0.00 1
 42 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 1
 43 -0.12 -0.22 -0.27 -0.28 -0.40 -0.17 -0.16 0.00 0.00 1
 44 0.08 -0.20 -0.29 -0.25 -0.30 -0.28 -0.04 0.00 0.00 0.48 1
 45 /STANDARD DEVIATIONS
 46 10.26 15.43 1.10 8.26 17.00 4.25 2.35 0.00 0.00 5.91 5.63
 47 /TECHNICAL
 48 ITR = 50;
 49 /END

49 CUMULATED RECORDS OF INPUT MODEL FILE WERE READ (GROUP 1)

PROGRAM CONTROL INFORMATION

50 /TITLE
 51 Experiment 2 - Three factors
 52 /SPECIFICATIONS
 53 VARIABLES = 11; CASES = 47; METHOD = ML; MATRIX = CORR;
 54 ANALYSIS = CORR;
 55 /LABELS
 56 V1 = ARROW; V2 = WORD; V3 = LETTER; V4 = TTT; V5 = VERBAL;
 57 V6 = SPATver; V7 = SYLLver; V8 = SPATvis; V9 = SYLLvis; V10 = LETTerr;
 58 V11 = VERerr;
 59 /EQUATIONS
 60 $V1 = + 4 * F2 + 5 * F3 + E1;$
 61 $V2 = + 10 * F1 + 10 * F3 + E2;$
 62 $V4 = + 4 * F2 + 4 * F3 + E4;$
 63 $V5 = + 8 * F1 + 11 * F3 + E5;$
 64 $V6 = + 1.5 * F1 + 1 * F2 + 2 * F3 + E6;$
 65 $V8 = + 2 * F1 + 0.5 * F2 + 2 * F3 + E8;$
 66 /VARIANCES
 67 F1 = 1.00;
 68 F2 = 1.00;
 69 F3 = 1.00;
 70 E1 = *;
 71 E2 = *;
 72 E4 = *;
 73 E5 = *;
 74 E6 = *;
 75 E8 = *;
 76 /MATRIX
 77 1
 78 0.28 1
 79 0.00 0.00 1
 80 0.19 0.40 0.00 1
 81 0.10 0.64 0.00 0.41 1
 82 0.52 0.30 0.00 0.46 0.36 1
 83 0.00 0.00 0.00 0.00 0.00 0.00 1
 84 0.25 0.14 0.00 0.41 0.24 0.69 0.00 1
 85 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 1
 86 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 1

87 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 1
88 /STANDARD DEVIATIONS
89 9.53 16.56 0.00 5.19 17.96 3.75 0.00 3.13 0.00 0.00 0.00
90 /TECHNICAL
91 ITR = 50;
92 /END

92 CUMULATED RECORDS OF INPUT MODEL FILE WERE READ (GROUP 2)

PROGRAM CONTROL INFORMATION

93 /TITLE
94 Experiment 3 - Three factors
95 /SPECIFICATIONS
96 VARIABLES = 11; CASES = 46; METHOD = ML; MATRIX = CORR;
97 ANALYSIS = CORR;
98 /LABELS
99 V1 = ARROW; V2 = WORD; V3 = LETTER; V4 = TTT; V5 = VERBAL;
100 V6 = SPATver; V7 = SYLLver; V8 = SPATvis; V9 = SYLLvis; V10 = LETTERR;
101 V11 = VERERR;
102 /EQUATIONS
103 $V1 = +4*F2 + 5*F3 + E1;$
104 $V2 = +10*F1 + 10*F3 + E2;$
105 $V3 = +0.15*F2 + 1*F3 + E3;$
106 $V4 = +4*F2 + 4*F3 + E4;$
107 $V5 = +8*F1 + 11*F3 + E5;$
108 $V7 = +0.5*F1 + 0.5*F2 + 1.5*F3 + E7;$
109 $V9 = +0.05*F1 + 2*F2 + 3*F3 + E9;$
110 $V10 = +*F3 + E10;$
111 /VARIANCES
112 $F1 = 1.00;$
113 $F2 = 1.00;$
114 $F3 = 1.00;$
115 $E1 = *;$
116 $E2 = *;$
117 $E3 = *;$
118 $E4 = *;$
119 $E5 = *;$
120 $E6 = *;$
121 $E7 = *;$
122 $E9 = *;$
123 $E10 = *;$
124 /MATRIX
125 1
126 0.26 1
127 0.47 0.41 1
128 0.40 0.30 0.49 1
129 0.31 0.66 0.48 0.35 1
130 0.00 0.00 0.00 0.00 0.00 1
131 0.50 0.57 0.46 0.45 0.45 0.00 1
132 0.00 0.00 0.00 0.00 0.00 0.00 0.00 1

```

133 0.45 0.44 0.47 0.21 0.50 0.00 0.66 0.00 1
134 -0.24 -0.26 -0.33 -0.08 -0.32 0.00 -0.20 0.00 -0.06 1
135 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 1
136 /STANDARD DEVIATIONS
137 11.08 18.68 1.25 5.07 19.11 0.00 2.73 0.00 3.74 7.21 0.00
138 /TECHNICAL
139 ITR = 50;
140 /LMTEST
141 /CONSTRAINT
142 (1,V1,F3) = (2,V1,F3) = (3,V1,F3);
143 (1,V2,F3) = (2,V2,F3) = (3,V2,F3);
144 (1,V3,F3) = (3,V3,F3);
145 (1,V4,F3) = (2,V4,F3) = (3,V4,F3);
146 (1,V5,F3) = (2,V5,F3) = (3,V5,F3);
147 (1,V6,F3) = (2,V6,F3);
148 (1,V7,F3) = (3,V7,F3);
149 (1,V10,F3) = (3,V10,F3);
150 (1,V1,F2) = (2,V1,F2) = (3,V1,F2);
151 (1,V2,F1) = (2,V2,F1) = (3,V2,F1);
152 (1,V3,F2) = (3,V3,F2);
153 (1,V4,F2) = (2,V4,F2) = (3,V4,F2);
154 (1,V5,F1) = (2,V5,F1) = (3,V5,F1);
155 (1,V6,F1) = (2,V6,F1);
156 (1,V7,F1) = (3,V7,F1);
157 (1,V6,F2) = (2,V6,F2);
158 (1,V7,F2) = (3,V7,F2);
159 (1,E1,E1) = (2,E1,E1) = (3,E1,E1);
160 (1,E2,E2) = (2,E2,E2) = (3,E2,E2);
161 (1,E3,E3) = (3,E3,E3);
162 (1,E4,E4) = (2,E4,E4) = (3,E4,E4);
163 (1,E5,E5) = (2,E5,E5) = (3,E5,E5);
164 (1,E6,E6) = (2,E6,E6);
165 (1,E7,E7) = (3,E7,E7);
166 (1,E10,E10) = (3,E10,E10);
167 /END

```

167 CUMULATED RECORDS OF INPUT MODEL FILE WERE READ (GROUP 3)

CORRELATION MATRIX TO BE ANALYZED: 9 VARIABLES (SELECTED FROM 11 VARIABLES) BASED ON 45 CASES.

*** WARNING *** STATISTICS MAY NOT BE MEANINGFUL
DUE TO ANALYZING CORRELATION MATRIX

		ARROW	WORD	LETTER	TTT	VERBAL
		V 1	V 2	V 3	V 4	V 5
ARROW	V 1	1.000				
WORD	V 2	0.230	1.000			
LETTER	V 3	0.310	0.420	1.000		
TTT	V 4	0.080	0.390	0.550	1.000	
VERBAL	V 5	0.110	0.590	0.370	0.310	1.000

SPATVER V 6	0.100	0.500	0.240	0.480	0.360
SYLLVER V 7	0.430	0.400	0.370	0.100	0.300
LETTERR V 10	-0.120	-0.220	-0.270	-0.280	-0.400
VERERR V 11	0.080	-0.200	-0.290	-0.250	-0.300

	SPATVER V 6	SYLLVER V 7	LETTERR V 10	VERERR V 11
SPATVER V 6	1.000			
SYLLVER V 7	0.270	1.000		
LETTERR V 10	-0.170	-0.160	1.000	
VERERR V 11	-0.280	-0.040	0.480	1.000

BENTLER-WEEKS STRUCTURAL REPRESENTATION:

NUMBER OF DEPENDENT VARIABLES = 9

DEPENDENT V'S : 1 2 3 4 5 6 7 10 11

NUMBER OF INDEPENDENT VARIABLES = 12

INDEPENDENT F'S : 1 2 3

INDEPENDENT E'S : 1 2 3 4 5 6 7 10 11

NUMBER OF FREE PARAMETERS = 27

NUMBER OF FIXED NONZERO PARAMETERS = 12

CORRELATION MATRIX TO BE ANALYZED: 6 VARIABLES (SELECTED FROM 11 VARIABLES) BASED ON 47 CASES.

*** WARNING *** STATISTICS MAY NOT BE MEANINGFUL
DUE TO ANALYZING CORRELATION MATRIX

	ARROW V 1	WORD V 2	TTT V 4	VERBAL V 5	SPATVER V 6
ARROW V 1	1.000				
WORD V 2	0.280	1.000			
TTT V 4	0.190	0.400	1.000		
VERBAL V 5	0.100	0.640	0.410	1.000	
SPATVER V 6	0.520	0.300	0.460	0.360	1.000
SPATVIS V 8	0.250	0.140	0.410	0.240	0.690

	SPATVIS V 8
SPATVIS V 8	1.000

BENTLER-WEEKS STRUCTURAL REPRESENTATION:

NUMBER OF DEPENDENT VARIABLES = 6

DEPENDENT V'S : 1 2 4 5 6 8

NUMBER OF INDEPENDENT VARIABLES = 9

INDEPENDENT F'S : 1 2 3

INDEPENDENT E'S : 1 2 4 5 6 8

NUMBER OF FREE PARAMETERS = 20

NUMBER OF FIXED NONZERO PARAMETERS = 9

CORRELATION MATRIX TO BE ANALYZED: 8 VARIABLES (SELECTED FROM 11 VARIABLES) BASED ON 46 CASES.

*** WARNING *** STATISTICS MAY NOT BE MEANINGFUL
DUE TO ANALYZING CORRELATION MATRIX

		ARROW	WORD	LETTER	TTT	VERBAL
		V 1	V 2	V 3	V 4	V 5
ARROW	V 1	1.000				
WORD	V 2	0.260	1.000			
LETTER	V 3	0.470	0.410	1.000		
TTT	V 4	0.400	0.300	0.490	1.000	
VERBAL	V 5	0.310	0.660	0.480	0.350	1.000
SYLLVER	V 7	0.500	0.570	0.460	0.450	0.450
SYLLVIS	V 9	0.450	0.440	0.470	0.210	0.500
LETTERR	V 10	-0.240	-0.260	-0.330	-0.080	-0.320

		SYLLVER	SYLLVIS	LETTERR
		V 7	V 9	V 10
SYLLVER	V 7	1.000		
SYLLVIS	V 9	0.660	1.000	
LETTERR	V 10	-0.200	-0.060	1.000

BENTLER-WEEKS STRUCTURAL REPRESENTATION:

NUMBER OF DEPENDENT VARIABLES = 8

DEPENDENT V'S : 1 2 3 4 5 7 9 10

NUMBER OF INDEPENDENT VARIABLES = 12

INDEPENDENT F'S : 1 2 3

INDEPENDENT E'S : 1 2 3 4 5 6 7 9 10

NUMBER OF FREE PARAMETERS = 26

NUMBER OF FIXED NONZERO PARAMETERS = 11

3RD STAGE OF COMPUTATION REQUIRED 23755 WORDS OF MEMORY.
PROGRAM ALLOCATED 100000 WORDS

DETERMINANT OF INPUT MATRIX IN GROUP 1 IS 0.65983E-01

DETERMINANT OF INPUT MATRIX IN GROUP 2 IS 0.11548E+00

DETERMINANT OF INPUT MATRIX IN GROUP 3 IS 0.39878E-01

MULTIPLE POPULATION ANALYSIS. INFORMATION IN GROUP 1

MAXIMUM LIKELIHOOD SOLUTION (NORMAL DISTRIBUTION THEORY)

PARAMETER ESTIMATES APPEAR IN ORDER,
NO SPECIAL PROBLEMS WERE ENCOUNTERED DURING OPTIMIZATION.

RESIDUAL COVARIANCE MATRIX (S-SIGMA):

		ARROW V 1	WORD V 2	LETTER V 3	TTT V 4	VERBAL V 5
ARROW	V 1	0.000				
WORD	V 2	0.018	0.000			
LETTER	V 3	-0.108	0.045	-0.008		
TTT	V 4	-0.176	0.008	0.158	0.000	
VERBAL	V 5	-0.103	-0.042	-0.007	-0.074	0.000
SPATVER	V 6	-0.224	0.125	-0.243	0.030	-0.019
SYLLVER	V 7	-0.006	-0.013	-0.076	-0.200	-0.113
LETTERR	V 10	0.027	0.058	-0.011	-0.016	-0.120
VERERR	V 11	0.234	0.093	-0.018	0.027	-0.006

		SPATVER V 6	SYLLVER V 7	LETTERR V 10	VERERR V 11
SPATVER	V 6	-0.012			
SYLLVER	V 7	-0.051	0.007		
LETTERR	V 10	0.148	0.023	0.000	
VERERR	V 11	0.054	0.152	0.278	-0.003

AVERAGE ABSOLUTE COVARIANCE RESIDUALS = 0.0697

AVERAGE OFF-DIAGONAL ABSOLUTE COVARIANCE RESIDUALS = 0.0862

STANDARDIZED RESIDUAL MATRIX:

		ARROW V 1	WORD V 2	LETTER V 3	TTT V 4	VERBAL V 5
ARROW	V 1	0.000				
WORD	V 2	0.018	0.000			
LETTER	V 3	-0.108	0.045	-0.008		
TTT	V 4	-0.176	0.008	0.158	0.000	
VERBAL	V 5	-0.103	-0.042	-0.007	-0.074	0.000
SPATVER	V 6	-0.224	0.125	-0.243	0.030	-0.019
SYLLVER	V 7	-0.006	-0.013	-0.076	-0.200	-0.113
LETTERR	V 10	0.027	0.058	-0.011	-0.016	-0.120
VERERR	V 11	0.234	0.093	-0.018	0.027	-0.006

		SPATVER V 6	SYLLVER V 7	LETTERR V 10	VERERR V 11
SPATVER	V 6	-0.012			

SYLLVER	V 7	-0.051	0.007		
LETERR	V 10	0.148	0.023	0.000	
VERERR	V 11	0.054	0.152	0.278	-0.003

AVERAGE ABSOLUTE STANDARDIZED RESIDUALS = 0.0697
 AVERAGE OFF-DIAGONAL ABSOLUTE STANDARDIZED RESIDUALS = 0.0862

MULTIPLE POPULATION ANALYSIS. INFORMATION IN GROUP 1

MAXIMUM LIKELIHOOD SOLUTION (NORMAL DISTRIBUTION THEORY)

LARGEST STANDARDIZED RESIDUALS:

V 11,V 10	V 6,V 3	V 11,V 1	V 6,V 1	V 7,V 4
0.278	-0.243	0.234	-0.224	-0.200

V 4,V 1	V 4,V 3	V 11,V 7	V 10,V 6	V 6,V 2
-0.176	0.158	0.152	0.148	0.125

V 10,V 5	V 7,V 5	V 3,V 1	V 5,V 1	V 11,V 2
-0.120	-0.113	-0.108	-0.103	0.093

V 7,V 3	V 5,V 4	V 10,V 2	V 11,V 6	V 7,V 6
-0.076	-0.074	0.058	0.054	-0.051

DISTRIBUTION OF STANDARDIZED RESIDUALS

```

-----
!
20-
! * !
! * !
! * !
! * !
! * !
15- * -
! * ! 1 -0.5 - -- 0 0.00%
! * ! 2 -0.4 - -0.5 0 0.00%
! * * ! 3 -0.3 - -0.4 0 0.00%
! * * ! 4 -0.2 - -0.3 3 6.67%
10- * * - 5 -0.1 - -0.2 5 11.11%
! * * ! 6 0.0 - -0.1 19 42.22%
! * * ! 7 0.1 - 0.0 12 26.67%
! * * ! 8 0.2 - 0.1 4 8.89%
! * * ! 9 0.3 - 0.2 2 4.44%

```

5-	* * *	- A	0.4	-	0.3	0	0.00%
!	* * * *	! B	0.5	-	0.4	0	0.00%
!	* * * * *	! C	++	-	0.5	0	0.00%
!	* * * * * *	!	-----				
!	* * * * * *	!	TOTAL		45		100.00%

1 2 3 4 5 6 7 8 9 A B C EACH "*" REPRESENTS 1 RESIDUALS

MEASUREMENT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS

$$\text{ARROW} = V1 = -.573 * F2 + .334 * F3 + 1.000 E1$$

.122	.113
-4.686	2.957

$$\text{WORD} = V2 = .480 * F1 + .635 * F3 + 1.000 E2$$

.112	.100
4.297	6.328

$$\text{LETTER} = V3 = -.386 * F2 + .590 * F3 + 1.000 E3$$

.127	.114
-3.027	5.158

$$\text{TTT} = V4 = -.096 * F2 + .602 * F3 + 1.000 E4$$

.115	.086
-.839	7.000

$$\text{VERBAL} = V5 = .473 * F1 + .638 * F3 + 1.000 E5$$

.111	.100
4.252	6.382

$$\text{SPATVER} = V6 = -.177 * F1 - .143 * F2 + .725 * F3 + 1.000 E6$$

.122	.159	.101
-1.454	-9.01	7.194

$$\text{SYLLVER} = V7 = .310 * F1 - .518 * F2 + .416 * F3 + 1.000 E7$$

.118	.105	.120
2.625	-4.919	3.483

$$\text{LETERR} = V10 = -.438 * F3 + 1.000 E10$$

.113
-3.887

$$\text{VERERR} = V11 = -.461 * F3 + 1.000 E11$$

.158
-2.927

VARIANCES OF INDEPENDENT VARIABLES

V	F	
---	---	
I F1 - F1	1.000	I
I		I
I		I
I		I
I F2 - F2	1.000	I
I		I
I		I
I		I
I F3 - F3	1.000	I
I		I
I		I
I		I

VARIANCES OF INDEPENDENT VARIABLES

	E	D	
	---	---	
E1 -ARROW	.560*		I
	.117		I
	4.779		I
			I
E2 - WORD	.366*		I
	.090		I
	4.058		I
			I
E3 -LETTER	.511*		I
	.104		I
	4.922		I
			I
E4 - TTT	.629*		I
	.087		I
	7.231		I
			I
E5 -VERBAL	.369*		I
	.089		I
	4.126		I
			I
E6 -SPATVER	.436*		I
	.088		I
	4.937		I
			I
E7 -SYLLVER	.454*		I
	.088		I
	5.187		I
			I

E10 -LETTERR	.808*	I
	.129	I
	6.238	I
		I
E11 -VERERR	.790*	I
	.181	I
	4.362	I
		I

STANDARDIZED SOLUTION:

R-SQUARED

ARROW	=V1 =	-.573*F2	+ .334*F3	+ .749 E1		.440
WORD	=V2 =	.480*F1	+ .635*F3	+ .605 E2		.634
LETTER	=V3 =	-.384*F2	+ .588*F3	+ .712 E3		.494
TTT	=V4 =	-.096*F2	+ .602*F3	+ .793 E4		.371
VERBAL	=V5 =	.473*F1	+ .638*F3	+ .608 E5		.631
SPATVER	=V6 =	-.176*F1	-.142*F2	+ .720*F3	+ .656 E6	.570
SYLLVER	=V7 =	.312*F1	-.520*F2	+ .418*F3	+ .676 E7	.543
LETTERR	=V10 =	-.438*F3	+ .899 E10			.192
VERERR	=V11 =	-.460*F3	+ .888 E11			.212

 END OF METHOD

MULTIPLE POPULATION ANALYSIS, INFORMATION IN GROUP 2

MAXIMUM LIKELIHOOD SOLUTION (NORMAL DISTRIBUTION THEORY)

PARAMETER CONDITION CODE
 E8,E8 CONSTRAINED AT LOWER BOUND

RESIDUAL COVARIANCE MATRIX (S-SIGMA) :

		ARROW	WORD	TTT	VERBAL	SPATVER
		V 1	V 2	V 4	V 5	V 6
ARROW	V 1	0.000				
WORD	V 2	0.068	0.000			
TTT	V 4	-0.066	0.018	0.000		
VERBAL	V 5	-0.113	0.008	0.026	0.000	
SPATVER	V 6	0.196	-0.075	0.010	-0.019	-0.012
SPATVIS	V 8	0.085	-0.069	-0.042	0.024	0.045
		SPATVIS				
		V 8				
SPATVIS	V 8	0.013				

AVERAGE ABSOLUTE COVARIANCE RESIDUALS = 0.0423
 AVERAGE OFF-DIAGONAL ABSOLUTE COVARIANCE RESIDUALS = 0.0576

STANDARDIZED RESIDUAL MATRIX:

		ARROW V 1	WORD V 2	TTT V 4	VERBAL V 5	SPATVER V 6
ARROW	V 1	0.000				
WORD	V 2	0.068	0.000			
TTT	V 4	-0.066	0.018	0.000		
VERBAL	V 5	-0.113	0.008	0.026	0.000	
SPATVER	V 6	0.196	-0.075	0.010	-0.019	-0.012
SPATVIS	V 8	0.085	-0.069	-0.042	0.024	0.045

		SPATVIS V 8
SPATVIS	V 8	0.013

AVERAGE ABSOLUTE STANDARDIZED RESIDUALS = 0.0423
 AVERAGE OFF-DIAGONAL ABSOLUTE STANDARDIZED RESIDUALS = 0.0576

LARGEST STANDARDIZED RESIDUALS:

V 6,V 1	V 5,V 1	V 8,V 1	V 6,V 2	V 8,V 2
0.196	-0.113	0.085	-0.075	-0.069
V 2,V 1	V 4,V 1	V 8,V 6	V 8,V 4	V 5,V 4
0.068	-0.066	0.045	-0.042	0.026
V 8,V 5	V 6,V 5	V 4,V 2	V 8,V 8	V 6,V 6
0.024	-0.019	0.018	0.013	-0.012
V 6,V 4	V 5,V 2	V 1,V 1	V 5,V 5	V 2,V 2
0.010	0.008	0.000	0.000	0.000

DISTRIBUTION OF STANDARDIZED RESIDUALS

		RANGE	FREQ	PERCENT
20-	!	1 -0.5 - --	0	0.00%
!	!	2 -0.4 - -0.5	0	0.00%
!	!	3 -0.3 - -0.4	0	0.00%
!	!	4 -0.2 - -0.3	0	0.00%
10-	*	5 -0.1 - -0.2	1	4.76%
!	* *	6 0.0 - -0.1	10	47.62%
!	* *	7 0.1 - 0.0	9	42.86%
!	* *	8 0.2 - 0.1	1	4.76%
!	* *	9 0.3 - 0.2	0	0.00%
5-	* *	A 0.4 - 0.3	0	0.00%
!	* *	B 0.5 - 0.4	0	0.00%
!	* *	C ++ - 0.5	0	0.00%
!	* *			
!	* * * *	TOTAL	21	100.00%

1 2 3 4 5 6 7 8 9 A B C EACH "*" REPRESENTS 1 RESIDUALS

MEASUREMENT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS

$$\text{ARROW} = V1 = -.573 * F2 + .334 * F3 + 1.000 E1$$

.122 .113
-4.686 2.957

$$\text{WORD} = V2 = .480 * F1 + .635 * F3 + 1.000 E2$$

.112 .100
4.297 6.328

$$\text{TTT} = V4 = -.096 * F2 + .602 * F3 + 1.000 E4$$

.115 .086
-.839 7.000

$$\text{VERBAL} = V5 = .473 * F1 + .638 * F3 + 1.000 E5$$

.111 .100
4.252 6.382

$$\text{SPATVER} = V6 = -.177 * F1 - .143 * F2 + .725 * F3 + 1.000 E6$$

	.122	.159	.101	
	-1.454	-.901	7.194	
SPATVIS =V8 =	$-.594 * F1$	$+ .165 * F2$	$+ .778 * F3$	+1.000 E8
	.154	.240	.135	
	-3.871	.690	5.782	

VARIANCES OF INDEPENDENT VARIABLES

V	F	
---	---	
I F1 - F1	1.000	I
I		I
I		I
I		I
I F2 - F2	1.000	I
I		I
I		I
I		I
I F3 - F3	1.000	I
I		I
I		I
I		I

VARIANCES OF INDEPENDENT VARIABLES

	E	D	
	---	---	
E1 -ARROW	.560*		I
	.117		I
	4.779		I
			I
E2 - WORD	.366*		I
	.090		I
	4.058		I
			I
E4 - TTT	.629*		I
	.087		I
	7.231		I
			I
E5 -VERBAL	.369*		I
	.089		I
	4.126		I
			I
E6 -SPATVER	.436*		I
	.088		I
	4.937		I
			I

E8 -SPATVIS	.000*	I
	.019	I
	.000	I
		I

STANDARDIZED SOLUTION:

R-SQUARED

ARROW	=V1 =	-.573*F2	+ .334*F3	+ .749 E1		.440
WORD	=V2 =	.480*F1	+ .635*F3	+ .605 E2		.634
TTT	=V4 =	-.096*F2	+ .602*F3	+ .793 E4		.371
VERBAL	=V5 =	.473*F1	+ .638*F3	+ .608 E5		.631
SPATVER	=V6 =	-.176*F1	- .142*F2	+ .720*F3	+ .656 E6	.570
SPATVIS	=V8 =	-.599*F1	+ .166*F2	+ .784*F3	+ .000 E8	1.000

 END OF METHOD

MULTIPLE POPULATION ANALYSIS, INFORMATION IN GROUP 3

MAXIMUM LIKELIHOOD SOLUTION (NORMAL DISTRIBUTION THEORY)

PARAMETER	CONDITION CODE
E6,E6	CONSTRAINED AT LOWER BOUND
E9,E9	CONSTRAINED AT LOWER BOUND

ALL EQUALITY CONSTRAINTS WERE CORRECTLY IMPOSED
 E6,E6 VARIANCE OF PARAMETER ESTIMATE IS SET TO ZERO.

RESIDUAL COVARIANCE MATRIX (S-SIGMA) :

		ARROW	WORD	LETTER	TTT	VERBAL
		V 1	V 2	V 3	V 4	V 5
ARROW	V 1	0.000				
WORD	V 2	0.048	0.000			
LETTER	V 3	0.052	0.035	-0.008		
TTT	V 4	0.144	-0.082	0.098	0.000	
VERBAL	V 5	0.097	0.028	0.103	-0.034	0.000
SYLLVER	V 7	0.064	0.157	0.014	0.150	0.037
SYLLVIS	V 9	-0.022	0.049	0.103	0.058	0.114
LETERR	V 10	-0.093	0.018	-0.071	0.184	-0.040

SYLLVER	SYLLVIS	LETERR
V 7	V 9	V 10

SYLLVER V 7	0.007		
SYLLVIS V 9	0.020	0.020	
LETERR V 10	-0.017	-0.001	0.000

AVERAGE ABSOLUTE COVARIANCE RESIDUALS = 0.0547

AVERAGE OFF-DIAGONAL ABSOLUTE COVARIANCE RESIDUALS = 0.0690

STANDARDIZED RESIDUAL MATRIX:

		ARROW V 1	WORD V 2	LETTER V 3	TTT V 4	VERBAL V 5
ARROW V 1		0.000				
WORD V 2		0.048	0.000			
LETTER V 3		0.052	0.035	-0.008		
TTT V 4		0.144	-0.082	0.098	0.000	
VERBAL V 5		0.097	0.028	0.103	-0.034	0.000
SYLLVER V 7		0.064	0.157	0.014	0.150	0.037
SYLLVIS V 9		-0.022	0.049	0.103	0.058	0.114
LETERR V 10		-0.093	0.018	-0.071	0.184	-0.040

		SYLLVER V 7	SYLLVIS V 9	LETERR V 10
SYLLVER V 7		0.007		
SYLLVIS V 9		0.020	0.020	
LETERR V 10		-0.017	-0.001	0.000

AVERAGE ABSOLUTE STANDARDIZED RESIDUALS = 0.0547

AVERAGE OFF-DIAGONAL ABSOLUTE STANDARDIZED RESIDUALS = 0.0690

LARGEST STANDARDIZED RESIDUALS:

V 10,V 4 0.184 V 7,V 2 0.157 V 7,V 4 0.150 V 4,V 1 0.144 V 9,V 5 0.114

V 9,V 3 0.103 V 5,V 3 0.103 V 4,V 3 0.098 V 5,V 1 0.097 V 10,V 1 -0.093

V 4,V 2 -0.082 V 10,V 3 -0.071 V 7,V 1 0.064 V 9,V 4 0.058 V 3,V 1 0.052

V 9,V 2 0.049 V 2,V 1 0.048 V 10,V 5 -0.040 V 7,V 5 0.037 V 3,V 2 0.035

DISTRIBUTION OF STANDARDIZED RESIDUALS

				RANGE	FREQ	PERCENT
20-	!	!				
	!	!				
	!	!				
	!	!				
	!	!	*			
15-	!	!	*			
	!	!	*	1 -0.5 - --	0	0.00%
	!	!	* *	2 -0.4 - -0.5	0	0.00%
	!	!	* *	3 -0.3 - -0.4	0	0.00%
	!	!	* *	4 -0.2 - -0.3	0	0.00%
10-	!	!	* *	5 -0.1 - -0.2	0	0.00%
	!	!	* *	6 0.0 - -0.1	13	36.11%
	!	!	* *	7 0.1 - 0.0	16	44.44%
	!	!	* * *	8 0.2 - 0.1	7	19.44%
	!	!	* * *	9 0.3 - 0.2	0	0.00%
5-	!	!	* * *	A 0.4 - 0.3	0	0.00%
	!	!	* * *	B 0.5 - 0.4	0	0.00%
	!	!	* * *	C ++ - 0.5	0	0.00%
	!	!	* * *			
	!	!	* * *	TOTAL	36	100.00%

1 2 3 4 5 6 7 8 9 A B C EACH "*" REPRESENTS 1 RESIDUALS

MEASUREMENT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS

$$\text{ARROW} = V1 = -0.573 \cdot F2 + 0.334 \cdot F3 + 1.000 \cdot E1$$

.122 .113
-4.686 2.957

$$\text{WORD} = V2 = 0.480 \cdot F1 + 0.635 \cdot F3 + 1.000 \cdot E2$$

.112 .100
4.297 6.328

$$\text{LETTER} = V3 = -0.386 \cdot F2 + 0.590 \cdot F3 + 1.000 \cdot E3$$

.127 .114
-3.027 5.158

$$\text{TTT} = V4 = -0.096 \cdot F2 + 0.602 \cdot F3 + 1.000 \cdot E4$$

.115 .086
-0.839 7.000

$$\text{VERBAL} = V5 = \begin{matrix} .473 * F1 & + & .638 * F3 & + & 1.000 E5 \\ .111 & & .100 & & \\ 4.252 & & 6.382 & & \end{matrix}$$

$$\text{SYLLVER} = V7 = \begin{matrix} .310 * F1 & - & .518 * F2 & + & .416 * F3 \\ .118 & & .105 & & .120 \\ 2.625 & & -4.919 & & 3.483 \end{matrix}$$

$$+ 1.000 E7$$

$$\text{SYLLVIS} = V9 = \begin{matrix} .637 * F1 & - & .746 * F2 & + & .134 * F3 \\ .155 & & .144 & & .154 \\ 4.096 & & -5.185 & & .871 \end{matrix}$$

$$+ 1.000 E9$$

$$\text{LETTER} = V10 = \begin{matrix} -.438 * F3 & + & 1.000 E10 \\ .113 & & \\ -3.887 & & \end{matrix}$$

VARIANCES OF INDEPENDENT VARIABLES

V	F	
---	---	
I F1 - F1	1.000	I
I		I
I		I
I		I
I F2 - F2	1.000	I
I		I
I		I
I		I
I F3 - F3	1.000	I
I		I
I		I
I		I

VARIANCES OF INDEPENDENT VARIABLES

	E	D	
	---	---	
F1 -ARROW	.560*		I
	.117		I

	4.779	I
		I
E2 - WORD	.366*	I
	.090	I
	4.058	I
		I
E3 -LETTER	.511*	I
	.104	I
	4.922	I
		I
E4 - TTT	.629*	I
	.087	I
	7.231	I
		I
E5 -VERBAL	.369*	I
	.089	I
	4.126	I
		I
E6 -	.000*	I
	.000	I
	:0000000.000	I
		I
E7 -	.454*	I
	.088	I
	5.187	I
		I
E9 -	.000*	I
	.024	I
	.000	I
		I
E10 -	.808*	I
	.129	I
	6.238	I
		I

STANDARDIZED SOLUTION:

R-SQUARED

ARROW	=V1 =	-.573*F2	+ .334*F3	+ .749 E1	.440
WORD	=V2 =	.480*F1	+ .635*F3	+ .605 E2	.634
LETTER	=V3 =	-.384*F2	+ .588*F3	+ .712 E3	.494
TTT	=V4 =	-.096*F2	+ .602*F3	+ .793 E4	.371
VERBAL	=V5 =	.473*F1	+ .638*F3	+ .608 E5	.631
SYLLVER	=V7 =	.312*F1	- .520*F2	+ .418*F3 + .676 E7	.543
SYLLVIS	=V9 =	.644*F1	- .753*F2	+ .135*F3 + .000 E9	1.000
LETTERR	=V10 =	-.438*F3	+ .899 E10		.192

 END OF METHOD

STATISTICS FOR MULTIPLE POPULATION ANALYSIS

ALL EQUALITY CONSTRAINTS WERE CORRECTLY IMPOSED

*** WARNING *** TEST RESULTS MAY NOT BE APPROPRIATE DUE TO
CONDITION CODE

GOODNESS OF FIT SUMMARY

INDEPENDENCE MODEL CHI-SQUARE = 363.893 ON 79 DEGREES OF
FREEDOM

INDEPENDENCE AIC = 205.89278 INDEPENDENCE CAIC = -104.36026
MODEL AIC = -65.73135 MODEL CAIC = -324.93009

CHI-SQUARE = 66.269 BASED ON 66 DEGREES OF FREEDOM
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS 0.46756

BENTLER-BONETT NORMED FIT INDEX= 0.818
BENTLER-BONETT NONNORMED FIT INDEX= 0.999
COMPARATIVE FIT INDEX (CFI) = 0.999

ITERATIVE SUMMARY

PARAMETER			
ITERATION	ABS CHANGE	ALPHA	FUNCTION
1	1.504200	1.00000	6.78130
2	0.804048	1.00000	3.76480
3	0.369785	1.00000	1.89001
4	0.272837	1.00000	1.59240
5	0.367499	0.50000	1.33348
6	0.211601	1.00000	0.77847
7	0.114526	1.00000	0.56235
8	0.057159	1.00000	0.51060
9	0.023945	1.00000	0.49483
10	0.013211	1.00000	0.49204
11	0.006159	1.00000	0.49121
12	0.003578	1.00000	0.49097
13	0.001731	1.00000	0.49090
14	0.000977	1.00000	0.49088

LAGRANGE MULTIPLIER TEST (FOR RELEASING CONSTRAINTS)

CONSTRAINTS TO BE RELEASED ARE:

CONSTRAINTS FROM GROUP 3

CONSTR: 1 (1.V1.F3)-(2.V1.F3)=0;
 CONSTR: 2 (1.V1.F3)-(3.V1.F3)=0;
 CONSTR: 3 (1.V2.F3)-(2.V2.F3)=0;
 CONSTR: 4 (1.V2.F3)-(3.V2.F3)=0;
 CONSTR: 5 (1.V3.F3)-(3.V3.F3)=0;
 CONSTR: 6 (1.V4.F3)-(2.V4.F3)=0;
 CONSTR: 7 (1.V4.F3)-(3.V4.F3)=0;
 CONSTR: 8 (1.V5.F3)-(2.V5.F3)=0;
 CONSTR: 9 (1.V5.F3)-(3.V5.F3)=0;
 CONSTR: 10 (1.V6.F3)-(2.V6.F3)=0;
 CONSTR: 11 (1.V7.F3)-(3.V7.F3)=0;
 CONSTR: 12 (1.V10.F3)-(3.V10.F3)=0;
 CONSTR: 13 (1.V1.F2)-(2.V1.F2)=0;
 CONSTR: 14 (1.V1.F2)-(3.V1.F2)=0;
 CONSTR: 15 (1.V2.F1)-(2.V2.F1)=0;
 CONSTR: 16 (1.V2.F1)-(3.V2.F1)=0;
 CONSTR: 17 (1.V3.F2)-(3.V3.F2)=0;
 CONSTR: 18 (1.V4.F2)-(2.V4.F2)=0;
 CONSTR: 19 (1.V4.F2)-(3.V4.F2)=0;
 CONSTR: 20 (1.V5.F1)-(2.V5.F1)=0;
 CONSTR: 21 (1.V5.F1)-(3.V5.F1)=0;
 CONSTR: 22 (1.V6.F1)-(2.V6.F1)=0;
 CONSTR: 23 (1.V7.F1)-(3.V7.F1)=0;
 CONSTR: 24 (1.V6.F2)-(2.V6.F2)=0;
 CONSTR: 25 (1.V7.F2)-(3.V7.F2)=0;
 CONSTR: 26 (1.E1.E1)-(2.E1.E1)=0;
 CONSTR: 27 (1.E1.E1)-(3.E1.E1)=0;
 CONSTR: 28 (1.E2.E2)-(2.E2.E2)=0;
 CONSTR: 29 (1.E2.E2)-(3.E2.E2)=0;
 CONSTR: 30 (1.E3.E3)-(3.E3.E3)=0;
 CONSTR: 31 (1.E4.E4)-(2.E4.E4)=0;
 CONSTR: 32 (1.E4.E4)-(3.E4.E4)=0;
 CONSTR: 33 (1.E5.E5)-(2.E5.E5)=0;
 CONSTR: 34 (1.E5.E5)-(3.E5.E5)=0;
 CONSTR: 35 (1.E6.E6)-(2.E6.E6)=0;
 CONSTR: 36 (1.E7.E7)-(3.E7.E7)=0;
 CONSTR: 37 (1.E10.E10)-(3.E10.E10)=0;

UNIVARIATE TEST STATISTICS:

NO	CONSTRAINT	CHI-SQUARE	PROBABILITY
--	-----	-----	-----
1	CONSTR: 1	0.263	0.608
2	CONSTR: 2	0.842	0.359
3	CONSTR: 3	0.255	0.613
4	CONSTR: 4	0.003	0.959
5	CONSTR: 5	0.033	0.855
6	CONSTR: 6	0.047	0.829
7	CONSTR: 7	0.092	0.762

8	CONSTR: 8	0.095	0.758
9	CONSTR: 9	0.027	0.870
10	CONSTR: 10	0.027	0.871
11	CONSTR: 11	0.582	0.446
12	CONSTR: 12	0.002	0.963
13	CONSTR: 13	0.005	0.946
14	CONSTR: 14	0.351	0.554
15	CONSTR: 15	0.603	0.437
16	CONSTR: 16	0.022	0.883
17	CONSTR: 17	0.148	0.701
18	CONSTR: 18	0.478	0.489
19	CONSTR: 19	1.375	0.241
20	CONSTR: 20	0.164	0.686
21	CONSTR: 21	0.177	0.674
22	CONSTR: 22	1.659	0.198
23	CONSTR: 23	0.027	0.869
24	CONSTR: 24	3.613	0.057
25	CONSTR: 25	0.514	0.474
26	CONSTR: 26	0.290	0.590
27	CONSTR: 27	0.063	0.801
28	CONSTR: 28	0.007	0.934
29	CONSTR: 29	0.032	0.859
30	CONSTR: 30	0.734	0.392
31	CONSTR: 31	0.075	0.784
32	CONSTR: 32	0.000	0.983
33	CONSTR: 33	0.001	0.980
34	CONSTR: 34	0.384	0.535
35	CONSTR: 35	3.193	0.074
36	CONSTR: 36	0.493	0.483
37	CONSTR: 37	0.000	0.994

CUMULATIVE MULTIVARIATE STATISTICS

UNIVARIATE INCREMENT

STEP PARAMETER CHI-SQUARE D.F. PROBABILITY CHI-SQUARE
PROBABILITY

1	CONSTR: 24	3.613	1	0.057	3.613	0.057
2	CONSTR: 35	7.226	2	0.027	3.613	0.057
3	CONSTR: 22	12.483	3	0.006	5.256	0.022
4	CONSTR: 10	13.834	4	0.008	1.351	0.245
5	CONSTR: 19	15.052	5	0.010	1.219	0.270
6	CONSTR: 2	15.811	6	0.015	0.759	0.384
7	CONSTR: 11	17.082	7	0.017	1.270	0.260
8	CONSTR: 20	17.889	8	0.022	0.807	0.369
9	CONSTR: 25	18.677	9	0.028	0.788	0.375
10	CONSTR: 34	19.228	10	0.037	0.550	0.458
11	CONSTR: 27	19.759	11	0.049	0.532	0.466
12	CONSTR: 6	20.163	12	0.064	0.404	0.525
13	CONSTR: 31	20.654	13	0.080	0.491	0.483
14	CONSTR: 7	21.089	14	0.099	0.435	0.510

15	CONSTR:	32	21.531	15	0.121	0.442	0.506
16	CONSTR:	18	21.908	16	0.146	0.377	0.539
17	CONSTR:	28	22.288	17	0.174	0.379	0.538
18	CONSTR:	17	22.481	18	0.211	0.193	0.660
19	CONSTR:	30	22.847	19	0.244	0.366	0.545
20	CONSTR:	8	23.019	20	0.288	0.172	0.678
21	CONSTR:	9	23.367	21	0.325	0.348	0.555
22	CONSTR:	3	23.698	22	0.363	0.331	0.565
23	CONSTR:	15	23.867	23	0.411	0.169	0.681
24	CONSTR:	14	24.025	24	0.460	0.158	0.691
25	CONSTR:	12	24.076	25	0.515	0.051	0.821
26	CONSTR:	37	24.107	26	0.570	0.031	0.861
27	CONSTR:	16	24.132	27	0.623	0.025	0.874
28	CONSTR:	4	24.157	28	0.673	0.025	0.874
29	CONSTR:	21	24.176	29	0.720	0.019	0.891
30	CONSTR:	23	24.195	30	0.763	0.019	0.890
31	CONSTR:	1	24.208	31	0.802	0.013	0.910
32	CONSTR:	13	24.220	32	0.836	0.012	0.914
33	CONSTR:	26	24.243	33	0.866	0.023	0.879
34	CONSTR:	36	24.251	34	0.892	0.008	0.928
35	CONSTR:	29	24.254	35	0.914	0.003	0.953
36	CONSTR:	5	24.258	36	0.932	0.003	0.954
37	CONSTR:	33	24.258	37	0.947	0.000	0.987

**Appendix B1 . 14: Experiments 4, 5 and 6 Statistics: Multi-Group Factor Analysis,
Four-Factor Model**

EQS. A STRUCTURAL EQUATION PROGRAM MULTIVARIATE SOFTWARE. INC.
COPYRIGHT BY P.M. BENTLER VERSION 5.7b (C) 1985 - 1998.

PROGRAM CONTROL INFORMATION

1 /TITLE
2 Experiment 4 - Four factors
3 /SPECIFICATIONS
4 VARIABLES = 11; CASES = 47; METHOD = ML; MATRIX = CORR;
5 ANALYSIS = CORR; GROUP = 3;
6 /LABELS
7 V1 = ARROW; V2 = WORD; V3 = TTT; V4 = VERBAL; V5 = SIMSPAT;
8 V6 = COMSPAT; V7 = ABSTRACT; V8 = CONDITION; V9 = SYLL;
9 V10 = MULTI; V11 = TEMPORAL;
10 /EQUATIONS
11 V1 = + *F2 + *F3 + E1;
12 V2 = + *F1 + *F3 + E2;
13 V3 = + *F2 + *F3 + E3;
14 V4 = + *F1 + *F3 + E4;
15 V6 = + *F1 + *F2 + *F3 + *F4 + E6;
16 V11 = + *F1 + *F2 + *F3 + *F4 + E11;
17 /VARIANCES
18 F1 = 1.00;
19 F2 = 1.00;
20 F3 = 1.00;
21 F4 = 1.00;
22 E1 = *;
23 E2 = *;
24 E3 = *;
25 E4 = *;
26 E6 = *;
27 E11 = *;
28 /MATRIX
29 1
30 0.38 1
31 0.45 0.29 1
32 0.45 0.71 0.37 1
33 0.00 0.00 0.00 0.00 1
34 0.47 0.31 0.45 0.27 0.00 1
35 0.00 0.00 0.00 0.00 0.00 0.00 1
36 0.00 0.00 0.00 0.00 0.00 0.00 0.00 1
37 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 1
38 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 1
39 0.40 0.32 0.62 0.31 0.00 0.68 0.00 0.00 0.00 0.00 1
40 /STANDARD DEVIATIONS
41 10.10 14.83 9.40 18.73 0.00 3.37 0.00 0.00 0.00 0.00 3.74
42 /TECHNICAL
43 ITR = 80;

44 /END

44 CUMULATED RECORDS OF INPUT MODEL FILE WERE READ (GROUP 1)

PROGRAM CONTROL INFORMATION

45 /TITLE

46 Experiment 5 - Four factors

47 /SPECIFICATIONS

48 VARIABLES = 11; CASES = 51; METHOD = ML; MATRIX = CORR;

49 ANALYSIS = CORR;

50 /LABELS

51 V1 = ARROW; V2 = WORD; V3 = TTT; V4 = VERBAL; V5 = SIMSPAT;

52 V6 = COMSPAT; V7 = ABSTRACT; V8 = CONDITION; V9 = SYLL;

53 V10 = MULTI; V11 = TEMPORAL;

54 /EQUATIONS

55 V1 = + *F2 + *F3 + E1;

56 V2 = + *F1 + *F3 + E2;

57 V5 = + *F2 + *F3 + *F4 + E5;

58 V6 = + *F1 + *F2 + *F3 + *F4 + E6;

59 V7 = + *F1 + *F3 + *F4 + E7;

60 V8 = + *F1 + *F3 + *F4 + E8;

61 V9 = + *F1 + *F2 + *F3 + *F4 + E9;

62 V10 = + *F1 + *F2 + *F3 + *F4 + E10;

63 /VARIANCES

64 F1 = 1.00;

65 F2 = 1.00;

66 F3 = 1.00;

67 F4 = 1.00;

68 E1 = *;

69 E2 = *;

70 E5 = *;

71 E6 = *;

72 E7 = *;

73 E8 = *;

74 E9 = *;

75 E10 = *;

76 /MATRIX

77 1

78 0.17 1

79 0.00 0.00 1

80 0.00 0.00 0.00 1

81 0.33 0.39 0.00 0.00 1

82 0.27 0.08 0.00 0.00 0.30 1

83 0.26 0.32 0.00 0.00 0.45 0.23 1

84 0.19 0.21 0.00 0.00 0.40 0.06 0.45 1

85 0.20 0.04 0.00 0.00 0.28 0.48 0.16 0.34 1

86 0.32 0.23 0.00 0.00 0.53 0.36 0.19 0.26 0.22 1

87 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 1

88 /STANDARD DEVIATIONS

89 12.93 15.08 0.00 0.00 2.99 3.44 1.59 1.80 3.41 1.95 0.00

90 TECHNICAL

91 ITR = 80;
92 /END

92 CUMULATED RECORDS OF INPUT MODEL FILE WERE READ (GROUP 2)

PROGRAM CONTROL INFORMATION

93 /TITLE
94 Experiment 6 - Four factors
95 /SPECIFICATIONS
96 VARIABLES = 11; CASES = 48; METHOD = ML; MATRIX = CORR;
97 ANALYSIS = CORR;
98 /LABELS
99 V1 = ARROW; V2 = WORD; V3 = TTT; V4 = VERBAL; V5 = SIMSPAT;
100 V6 = COMSPAT; V7 = ABSTRACT; V8 = CONDITION; V9 = SYLL;
101 V10 = MULTI; V11 = TEMPORAL;
102 /EQUATIONS
103 $V1 = + *F2 + *F3 + E1;$
104 $V2 = + *F1 + *F3 + E2;$
105 $V3 = + *F2 + *F3 + E3;$
106 $V4 = + *F1 + *F3 + E4;$
107 $V5 = + *F2 + *F3 + *F4 + E5;$
108 $V6 = + *F1 + *F2 + *F3 + *F4 + E6;$
109 $V7 = + *F1 + *F3 + *F4 + E7;$
110 $V8 = + *F1 + *F3 + *F4 + E8;$
111 $V9 = + *F1 + *F2 + *F3 + *F4 + E9;$
112 $V10 = + *F1 + *F2 + *F3 + *F4 + E10;$
113 /VARIANCES
114 $F1 = 1.00;$
115 $F2 = 1.00;$
116 $F3 = 1.00;$
117 $F4 = 1.00;$
118 $E1 = *;$
119 $E2 = *;$
120 $E3 = *;$
121 $E4 = *;$
122 $E5 = *;$
123 $E6 = *;$
124 $E7 = *;$
125 $E8 = *;$
126 $E9 = *;$
127 $E10 = *;$
128 /MATRIX
129 1
130 0.42 1
131 0.31 0.05 1
132 0.33 0.69 -0.01 1
133 0.53 0.39 0.20 0.28 1
134 0.28 0.07 0.11 -0.14 0.44 1
135 0.05 0.24 -0.21 0.14 0.31 0.00 1
136 0.18 0.24 0.16 0.13 0.23 0.16 0.31 1
137 0.13 0.04 0.28 0.21 0.30 0.25 0.19 0.10 1

```

138 0.26 0.00 0.51 -0.14 0.47 0.36 0.08 0.10 0.41 1
139 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 1
140 /STANDARD DEVIATIONS
141 10.46 17.07 7.00 20.52 2.94 3.14 1.97 1.99 3.00 1.92 0.00
142 /TECHNICAL
143 ITR = 80:
144 /LMTEST
145 /CONSTRAINT
146 (1.V1.F2) = (2.V1.F2) = (3.V1.F2):
147 (1.V1.F3) = (2.V1.F3) = (3.V1.F3):
148 (1.V2.F1) = (2.V2.F1) = (3.V2.F1):
149 (1.V2.F3) = (2.V2.F3) = (3.V2.F3):
150 (1.V3.F2) = (3.V3.F2):
151 (1.V3.F3) = (3.V3.F3):
152 (1.V4.F1) = (3.V4.F1):
153 (1.V4.F3) = (3.V4.F3):
154 (2.V5.F2) = (3.V5.F2):
155 (2.V5.F3) = (3.V5.F3):
156 (2.V5.F4) = (3.V5.F4):
157 (1.V6.F1) = (2.V6.F1) = (3.V6.F1);
158 (1.V6.F2) = (2.V6.F2) = (3.V6.F2):
159 (1.V6.F3) = (2.V6.F3) = (3.V6.F3):
160 (1.V6.F4) = (2.V6.F4) = (3.V6.F4);
161 (2.V7.F1) = (3.V7.F1):
162 (2.V7.F3) = (3.V7.F3):
163 (2.V7.F4) = (3.V7.F4):
164 (2.V8.F1) = (3.V8.F1):
165 (2.V8.F3) = (3.V8.F3):
166 (2.V8.F4) = (3.V8.F4):
167 (2.V9.F1) = (3.V9.F1);
168 (2.V9.F2) = (3.V9.F2):
169 (2.V9.F3) = (3.V9.F3):
170 (2.V9.F4) = (3.V9.F4):
171 (2.V10.F1) = (3.V10.F1):
172 (2.V10.F2) = (3.V10.F2):
173 (2.V10.F3) = (3.V10.F3):
174 (2.V10.F4) = (3.V10.F4):
175 (1.E1.E1) = (2.E1.E1) = (3.E1.E1):
176 (1.E2.E2) = (2.E2.E2) = (3.E2.E2):
177 (1.E3.E3) = (3.E3.E3);
178 (1.E4.E4) = (3.E4.E4):
179 (2.E5.E5) = (3.E5.E5):
180 (2.E6.E6) = (3.E6.E6):
181 (2.E7.E7) = (3.E7.E7):
182 (2.E8.E8) = (3.E8.E8):
183 (2.E9.E9) = (3.E9.E9):
184 (2.E10.E10) = (3.E10.E10);
185 /END

```

185 CUMULATED RECORDS OF INPUT MODEL FILE WERE READ (GROUP 3)

CORRELATION MATRIX TO BE ANALYZED: 6 VARIABLES (SELECTED FROM 11 VARIABLES) BASED ON 47 CASES.

*** WARNING *** STATISTICS MAY NOT BE MEANINGFUL
DUE TO ANALYZING CORRELATION MATRIX

		ARROW V 1	WORD V 2	TTT V 3	VERBAL V 4	COMSPAT V 6
ARROW	V 1	1.000				
WORD	V 2	0.380	1.000			
TTT	V 3	0.450	0.290	1.000		
VERBAL	V 4	0.450	0.710	0.370	1.000	
COMSPAT	V 6	0.470	0.310	0.450	0.270	1.000
TEMPORAL	V 11	0.400	0.320	0.620	0.310	0.680

		TEMPORAL V 11
TEMPORAL	V 11	1.000

BENTLER-WEEKS STRUCTURAL REPRESENTATION:

NUMBER OF DEPENDENT VARIABLES = 6
DEPENDENT V'S : 1 2 3 4 6 11

NUMBER OF INDEPENDENT VARIABLES = 10
INDEPENDENT F'S : 1 2 3 4
INDEPENDENT E'S : 1 2 3 4 6 11

NUMBER OF FREE PARAMETERS = 22
NUMBER OF FIXED NONZERO PARAMETERS = 10

CORRELATION MATRIX TO BE ANALYZED: 8 VARIABLES (SELECTED FROM 11 VARIABLES) BASED ON 51 CASES.

*** WARNING *** STATISTICS MAY NOT BE MEANINGFUL
DUE TO ANALYZING CORRELATION MATRIX

		ARROW V 1	WORD V 2	SIMSPAT V 5	COMSPAT V 6	ABSTRACT V 7
ARROW	V 1	1.000				
WORD	V 2	0.170	1.000			
SIMSPAT	V 5	0.330	0.390	1.000		
COMSPAT	V 6	0.270	0.080	0.300	1.000	
ABSTRACT	V 7	0.260	0.320	0.450	0.230	1.000
CONDITIO	V 8	0.190	0.210	0.400	0.060	0.450
SYLL	V 9	0.200	0.040	0.280	0.480	0.160
MULTI	V 10	0.320	0.230	0.530	0.360	0.190

		CONDITIO	SYLL	MULTI
		V 8	V 9	V 10
CONDITIO	V 8	1.000		
SYLL	V 9	0.340	1.000	
MULTI	V 10	0.260	0.220	1.000

BENTLER-WEEKS STRUCTURAL REPRESENTATION:

NUMBER OF DEPENDENT VARIABLES = 8

DEPENDENT V'S : 1 2 5 6 7 8 9 10

NUMBER OF INDEPENDENT VARIABLES = 12

INDEPENDENT F'S : 1 2 3 4

INDEPENDENT E'S : 1 2 5 6 7 8 9 10

NUMBER OF FREE PARAMETERS = 33

NUMBER OF FIXED NONZERO PARAMETERS = 12

CORRELATION MATRIX TO BE ANALYZED: 10 VARIABLES (SELECTED FROM 11 VARIABLES) BASED ON 48 CASES.

*** WARNING *** STATISTICS MAY NOT BE MEANINGFUL
DUE TO ANALYZING CORRELATION MATRIX

		ARROW	WORD	TTT	VERBAL	SIMSPAT
		V 1	V 2	V 3	V 4	V 5
ARROW	V 1	1.000				
WORD	V 2	0.420	1.000			
TTT	V 3	0.310	0.050	1.000		
VERBAL	V 4	0.330	0.690	-0.010	1.000	
SIMSPAT	V 5	0.530	0.390	0.200	0.280	1.000
COMSPAT	V 6	0.280	0.070	0.110	-0.140	0.440
ABSTRACT	V 7	0.050	0.240	-0.210	0.140	0.310
CONDITIO	V 8	0.180	0.240	0.160	0.130	0.230
SYLL	V 9	0.130	0.040	0.280	0.210	0.300
MULTI	V 10	0.260	0.000	0.510	-0.140	0.470

		COMSPAT	ABSTRACT	CONDITIO	SYLL	MULTI
		V 6	V 7	V 8	V 9	V 10
COMSPAT	V 6	1.000				
ABSTRACT	V 7	0.000	1.000			
CONDITIO	V 8	0.160	0.310	1.000		
SYLL	V 9	0.250	0.190	0.100	1.000	
MULTI	V 10	0.360	0.080	0.100	0.410	1.000

BENTLER-WEEKS STRUCTURAL REPRESENTATION:

NUMBER OF DEPENDENT VARIABLES = 10

DEPENDENT V'S : 1 2 3 4 5 6 7 8 9 10

NUMBER OF INDEPENDENT VARIABLES = 14

INDEPENDENT F'S : 1 2 3 4

INDEPENDENT E'S : 1 2 3 4 5 6 7 8 9 10

NUMBER OF FREE PARAMETERS = 39

NUMBER OF FIXED NONZERO PARAMETERS = 14

3RD STAGE OF COMPUTATION REQUIRED 36326 WORDS OF MEMORY.
PROGRAM ALLOCATED 100000 WORDS

DETERMINANT OF INPUT MATRIX IN GROUP 1 IS 0.84680E-01

DETERMINANT OF INPUT MATRIX IN GROUP 2 IS 0.15352E+00

DETERMINANT OF INPUT MATRIX IN GROUP 3 IS 0.37202E-01

MULTIPLE POPULATION ANALYSIS, INFORMATION IN GROUP 1

MAXIMUM LIKELIHOOD SOLUTION (NORMAL DISTRIBUTION THEORY)

PARAMETER	CONDITION CODE
E4,E4	CONSTRAINED AT LOWER BOUND
E11,E11	CONSTRAINED AT LOWER BOUND

RESIDUAL COVARIANCE MATRIX (S-SIGMA) :

	ARROW	WORD	TTT	VERBAL	COMSPAT
	V 1	V 2	V 3	V 4	V 6
ARROW V 1	0.000				
WORD V 2	0.055	0.000			
TTT V 3	0.145	0.144	0.023		
VERBAL V 4	0.170	0.038	0.245	0.033	
COMSPAT V 6	0.184	0.183	0.087	0.204	0.116
TEMPORAL V 11	0.048	0.188	0.112	0.193	0.109

	TEMPORAL
	V 11
TEMPORAL V 11	0.093

AVERAGE ABSOLUTE COVARIANCE RESIDUALS = 0.1129

AVERAGE OFF-DIAGONAL ABSOLUTE COVARIANCE RESIDUALS = 0.1404

STANDARDIZED RESIDUAL MATRIX:

		ARROW V 1	WORD V 2	TTT V 3	VERBAL V 4	COMSPAT V 6
ARROW	V 1	0.000				
WORD	V 2	0.055	0.000			
TTT	V 3	0.145	0.144	0.023		
VERBAL	V 4	0.170	0.038	0.245	0.033	
COMSPAT	V 6	0.184	0.183	0.087	0.204	0.116
TEMPORAL	V 11	0.048	0.188	0.112	0.193	0.109

TEMPORAL
V 11
TEMPORAL V 11 0.093

AVERAGE ABSOLUTE STANDARDIZED RESIDUALS = 0.1129

AVERAGE OFF-DIAGONAL ABSOLUTE STANDARDIZED RESIDUALS = 0.1404

LARGEST STANDARDIZED RESIDUALS:

V 4,V 3 0.245
V 6,V 4 0.204
V 11,V 4 0.193
V 11,V 2 0.188
V 6,V 1 0.184

V 6,V 2 0.183
V 4,V 1 0.170
V 3,V 1 0.145
V 3,V 2 0.144
V 6,V 6 0.116

V 11,V 3 0.112
V 11,V 6 0.109
V 11,V 11 0.093
V 6,V 3 0.087
V 2,V 1 0.055

V 11,V 1 0.048
V 4,V 2 0.038
V 4,V 4 0.033
V 3,V 3 0.023
V 2,V 2 0.000

DISTRIBUTION OF STANDARDIZED RESIDUALS

		RANGE	FREQ	PERCENT
20-	!	1 -0.5 - --	0	0.00%
!	!	2 -0.4 - -0.5	0	0.00%
!	!	3 -0.3 - -0.4	0	0.00%
!	!	4 -0.2 - -0.3	0	0.00%
10-	* !	5 -0.1 - -0.2	0	0.00%
!	* !	6 0.0 - -0.1	1	4.76%
!	* * !	7 0.1 - 0.0	8	38.10%
!	* * !	8 0.2 - 0.1	10	47.62%
!	* * !	9 0.3 - 0.2	2	9.52%
5-	* * -	A 0.4 - 0.3	0	0.00%
!	* * !	B 0.5 - 0.4	0	0.00%
!	* * !	C ++ - 0.5	0	0.00%
!	* * * !			
!	* * * * !	TOTAL	21	100.00%

1 2 3 4 5 6 7 8 9 A B C EACH "*" REPRESENTS 1 RESIDUALS

MEASUREMENT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS

$$\text{ARROW} = V1 = .336 * F2 + .479 * F3 + 1.000 E1$$

.104 .100
3.216 4.766

$$\text{WORD} = V2 = .348 * F1 + .679 * F3 + 1.000 E2$$

.116 .097
3.010 7.009

$$\text{TTT} = V3 = .603 * F2 + .214 * F3 + 1.000 E3$$

.121 .129
4.982 1.661

$$\text{VERBAL} = V4 = .791 * F1 + .584 * F3 + 1.000 E4$$

.101 .141
7.826 4.135

$$\text{COMSPAT} = V6 = -.089 * F1 + .520 * F2 + .233 * F3 + .238 * F4 + 1.000 E6$$

	.094	.099	.118	.124	
	-.949	5.275	1.974	1.927	
TEMPORAL=V11 =	.006*F1	+ .775*F2	+ .191*F3	+ .520*F4	+1.000 E11
	.146	.184	.187	.209	
	.042	4.221	1.025	2.484	

VARIANCES OF INDEPENDENT VARIABLES

V	F	
---	---	
I F1 - F1	1.000	I
I		I
I		I
I		I
I F2 - F2	1.000	I
I		I
I		I
I		I
I F3 - F3	1.000	I
I		I
I		I
I		I
I F4 - F4	1.000	I
I		I
I		I
I		I

VARIANCES OF INDEPENDENT VARIABLES

E	D	
---	---	
E1 -ARROW	.658*	I
	.092	I
	7.127	I
		I
E2 - WORD	.418*	I
	.068	I
	6.103	I
		I
E3 - TTT	.568*	I
	.120	I
	4.722	I
		I
E4 -VERBAL	.000*	I
	.048	I
	.000	I
		I

E6 -COMSPAT	.495*	I
	.111	I
	4.472	I
		I
E11 -TEMPORAL	.000*	I
	.013	I
	.000	I
		I

STANDARDIZED SOLUTION:

R-SQUARED

ARROW	=V1 =	.336*F2	+ .479*F3	+ .811 E1	.342
WORD	=V2 =	.348*F1	+ .679*F3	+ .646 E2	.582
TTT	=V3 =	.610*F2	+ .217*F3	+ .762 E3	.419
VERBAL	=V4 =	.804*F1	+ .594*F3	+ .000 E4	1.000
COMSPAT	=V6 =	-.095*F1	+ .553*F2	+ .248*F3	+ .254*F4
	+ .748 E6				.441
TEMPORAL	=V11 =	.006*F1	+ .814*F2	+ .201*F3	+ .546*F4
	+ .000 E11				1.000

 END OF METHOD

MULTIPLE POPULATION ANALYSIS, INFORMATION IN GROUP 2

MAXIMUM LIKELIHOOD SOLUTION (NORMAL DISTRIBUTION THEORY)

PARAMETER CONDITION CODE
 E9,E9 CONSTRAINED AT LOWER BOUND

RESIDUAL COVARIANCE MATRIX (S-SIGMA) :

		ARROW	WORD	SIMSPAT	COMSPAT	ABSTRACT
		V 1	V 2	V 5	V 6	V 7
ARROW	V 1	0.000				
WORD	V 2	-0.155	0.000			
SIMSPAT	V 5	-0.129	-0.064	-0.076		
COMSPAT	V 6	-0.016	-0.047	-0.136	-0.080	
ABSTRACT	V 7	0.050	0.038	0.006	-0.009	-0.014
CONDITIO	V 8	0.025	-0.026	0.057	-0.118	0.070
SYLL	V 9	0.016	-0.053	-0.104	0.040	-0.116
MULTI	V 10	-0.064	0.075	0.000	0.104	-0.033
		CONDITIO	SYLL	MULTI		
		V 8	V 9	V 10		
CONDITIO	V 8	-0.010				

SYLL	V 9	0.120	-0.072	
MULTI	V 10	0.097	-0.149	-0.031

AVERAGE ABSOLUTE COVARIANCE RESIDUALS = 0.0611
 AVERAGE OFF-DIAGONAL ABSOLUTE COVARIANCE RESIDUALS = 0.0685

STANDARDIZED RESIDUAL MATRIX:

		ARROW	WORD	SIMSPAT	COMSPAT	ABSTRACT
		V 1	V 2	V 5	V 6	V 7
ARROW	V 1	0.000				
WORD	V 2	-0.155	0.000			
SIMSPAT	V 5	-0.129	-0.064	-0.076		
COMSPAT	V 6	-0.016	-0.047	-0.136	-0.080	
ABSTRACT	V 7	0.050	0.038	0.006	-0.009	-0.014
CONDITIO	V 8	0.025	-0.026	0.057	-0.118	0.070
SYLL	V 9	0.016	-0.053	-0.104	0.040	-0.116
MULTI	V 10	-0.064	0.075	0.000	-0.104	-0.033

		CONDITIO	SYLL	MULTI
		V 8	V 9	V 10
CONDITIO	V 8	-0.010		
SYLL	V 9	0.120	-0.072	
MULTI	V 10	0.097	-0.149	-0.031

AVERAGE ABSOLUTE STANDARDIZED RESIDUALS = 0.0611
 AVERAGE OFF-DIAGONAL ABSOLUTE STANDARDIZED RESIDUALS = 0.0685

LARGEST STANDARDIZED RESIDUALS:

V 2,V 1	V 10,V 9	V 6,V 5	V 5,V 1	V 9,V 8
-0.155	-0.149	-0.136	-0.129	0.120

V 8,V 6	V 9,V 7	V 9,V 5	V 10,V 6	V 10,V 8
-0.118	-0.116	-0.104	-0.104	0.097

V 6,V 6	V 5,V 5	V 10,V 2	V 9,V 9	V 8,V 7
-0.080	-0.076	0.075	-0.072	0.070

V 5,V 2	V 10,V 1	V 8,V 5	V 9,V 2	V 7,V 1
-0.064	-0.064	0.057	-0.053	0.050

DISTRIBUTION OF STANDARDIZED RESIDUALS

			RANGE	FREQ	PERCENT
20-	!	!			
	!	!			
	!	!			
	!	!			
	!	!			
15-	*	-			
	*	!	1 -0.5 - --	0	0.00%
	*	!	2 -0.4 - -0.5	0	0.00%
	* *	!	3 -0.3 - -0.4	0	0.00%
	* *	!	4 -0.2 - -0.3	0	0.00%
10-	* *	-	5 -0.1 - -0.2	8	22.22%
	* *	!	6 0.0 - -0.1	15	41.67%
	* * *	!	7 0.1 - 0.0	12	33.33%
	* * *	!	8 0.2 - 0.1	1	2.78%
	* * *	!	9 0.3 - 0.2	0	0.00%
5-	* * *	-	A 0.4 - 0.3	0	0.00%
	* * *	!	B 0.5 - 0.4	0	0.00%
	* * *	!	C ++ - 0.5	0	0.00%
	* * *	!			
	* * * *	!	TOTAL	36	100.00%

1 2 3 4 5 6 7 8 9 A B C EACH "*" REPRESENTS 1 RESIDUALS

MEASUREMENT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS

$$\text{ARROW} = V1 = .336 * F2 + .479 * F3 + 1.000 E1$$

.104 .100
3.216 4.766

$$\text{WORD} = V2 = .348 * F1 + .679 * F3 + 1.000 E2$$

.116 .097
3.010 7.009

$$\text{SIMSPAT} = V5 = .413 * F2 + .669 * F3 + .271 * F4 + 1.000 E5$$

.112 .114 .132
3.694 5.891 2.048

$$\text{COMSPAT} = V6 = -.089 * F1 + .520 * F2 + .233 * F3 + .238 * F4 + 1.000 E6$$

.094 .099 .118 .124
-.949 5.275 1.974 1.927

$$\text{ABSTRACT=V7} = -.045*F1 + .439*F3 + .556*F4 + 1.000 E7$$

.144	.138	.159
-.314	3.187	3.503

$$\text{CONDITIO=V8} = .006*F1 + .345*F3 + .412*F4 + 1.000 E8$$

.133	.127	.134
.042	2.724	3.078

$$\text{SYLL =V9} = .461*F1 + .689*F2 - .099*F3 + .611*F4 + 1.000 E9$$

.151	.163	.139	.181
3.054	4.232	-.711	3.375

$$\text{MULTI =V10} = -.234*F1 + .646*F2 + .349*F3 + .107*F4 + 1.000 E10$$

.109	.122	.142	.140
-2.138	5.317	2.450	.766

VARIANCES OF INDEPENDENT VARIABLES

V	F	
---	---	
I F1 - F1	1.000	I
I		I
I		I
I		I
I F2 - F2	1.000	I
I		I
I		I
I		I
I F3 - F3	1.000	I
I		I
I		I
I		I
I F4 - F4	1.000	I
I		I
I		I
I		I

VARIANCES OF INDEPENDENT VARIABLES

	E	D	
	---	---	
E1 -ARROW	.658*		I
	.092		I

	7.127	I
		I
E2 - WORD	.418*	I
	.068	I
	6.103	I
		I
E5 -SIMSPAT	.383*	I
	.085	I
	4.492	I
		I
E6 -COMSPAT	.690*	I
	.111	I
	6.234	I
		I
E7 -ABSTRACT	.510*	I
	.163	I
	3.127	I
		I
E8 -CONDITIO	.722*	I
	.132	I
	5.480	I
		I
E9 - SYLL	.000*	I
	.014	I
	.000	I
		I
E10 -MULTI	.425*	I
	.114	I
	3.714	I
		I

STANDARDIZED SOLUTION:

R-SQUARED

ARROW	=V1 =	.336*F2	+ .479*F3	+ .811 E1	.342	
WORD	=V2 =	.348*F1	+ .679*F3	+ .646 E2	.582	
SIMSPAT	=V5 =	.399*F2	+ .645*F3	+ .262*F4	+ .597 E5	.644
COMSPAT	=V6 =	-.086*F1	+ .501*F2	+ .224*F3	+ .229*F4	
		+ .799 E6				.361
ABSTRACT	=V7 =	-.045*F1	+ .435*F3	+ .552*F4	+ .709 E7	.497
CONDITIO	=V8 =	.006*F1	+ .343*F3	+ .410*F4	+ .845 E8	.286
SYLL	=V9 =	.446*F1	+ .666*F2	-.096*F3	+ .591*F4	
		+ .000 E9				1.000
MULTI	=V10 =	-.231*F1	+ .637*F2	+ .344*F3	+ .106*F4	
		+ .642 E10				.588

 END OF METHOD

MULTIPLE POPULATION ANALYSIS. INFORMATION IN GROUP 3

MAXIMUM LIKELIHOOD SOLUTION (NORMAL DISTRIBUTION THEORY)

PARAMETER CONDITION CODE
 E4,E4 CONSTRAINED AT LOWER BOUND

ALL EQUALITY CONSTRAINTS WERE CORRECTLY IMPOSED

NOTE: 1 CONSTRAINTS APPEAR TO BE LINEARLY DEPENDENT.
 THE DEGREES OF FREEDOM HAVE BEEN ADJUSTED.

E9,E9 VARIANCE OF PARAMETER ESTIMATE IS SET TO ZERO.

RESIDUAL COVARIANCE MATRIX (S-SIGMA) :

		ARROW V 1	WORD V 2	TTT V 3	VERBAL V 4	SIMSPAT V 5
ARROW	V 1	0.000				
WORD	V 2	0.095	0.000			
TTT	V 3	0.005	-0.096	0.023		
VERBAL	V 4	0.050	0.018	-0.135	0.033	
SIMSPAT	V 5	0.071	-0.064	-0.193	-0.111	-0.076
COMSPAT	V 6	-0.006	-0.057	-0.253	-0.206	0.004
ABSTRACT	V 7	-0.160	-0.042	-0.304	-0.080	-0.134
CONDITIO	V 8	0.015	0.004	0.086	-0.076	-0.113
SYLL	V 9	-0.054	-0.053	-0.114	-0.097	-0.084
MULTI	V 10	-0.124	-0.155	0.046	-0.159	-0.060

		COMSPAT V 6	ABSTRACT V 7	CONDITIO V 8	SYLL V 9	MULTI V 10
COMSPAT	V 6	-0.080				
ABSTRACT	V 7	-0.239	-0.014			
CONDITIO	V 8	-0.018	-0.070	-0.010		
SYLL	V 9	-0.190	-0.086	-0.120	-0.072	
MULTI	V 10	-0.104	-0.143	-0.063	0.041	-0.031

AVERAGE ABSOLUTE COVARIANCE RESIDUALS = 0.0862
 AVERAGE OFF-DIAGONAL ABSOLUTE COVARIANCE RESIDUALS = 0.0978

STANDARDIZED RESIDUAL MATRIX:

	ARROW V 1	WORD V 2	TTT V 3	VERBAL V 4	SIMSPAT V 5
--	--------------	-------------	------------	---------------	----------------

ARROW	V 1	0.000					
WORD	V 2	0.095	0.000				
TTT	V 3	0.005	-0.096	0.023			
VERBAL	V 4	0.050	0.018	-0.135	0.033		
SIMSPAT	V 5	0.071	-0.064	-0.193	-0.111		-0.076
COMSPAT	V 6	-0.006	-0.057	-0.253	-0.206	0.004	
ABSTRACT	V 7	-0.160	-0.042	-0.304	-0.080	-0.134	
CONDITIO	V 8	0.015	0.004	0.086	-0.076	-0.113	
SYLL	V 9	-0.054	-0.053	-0.114	-0.097	-0.084	
MULTI	V 10	-0.124	-0.155	0.046	-0.159	-0.060	

	COMSPAT V 6	ABSTRACT V 7	CONDITIO V 8	SYLL V 9	MULTI V 10
COMSPAT V 6	-0.080				
ABSTRACT V 7	-0.239	-0.014			
CONDITIO V 8	-0.018	-0.070	-0.010		
SYLL V 9	-0.190	-0.086	-0.120	-0.072	
MULTI V 10	-0.104	-0.143	-0.063	0.041	-0.031

AVERAGE ABSOLUTE STANDARDIZED RESIDUALS = 0.0862

AVERAGE OFF-DIAGONAL ABSOLUTE STANDARDIZED RESIDUALS = 0.0978

LARGEST STANDARDIZED RESIDUALS:

V 7,V 3 V 6,V 3 V 7,V 6 V 6,V 4 V 5,V 3
-0.304 -0.253 -0.239 -0.206 -0.193

V 9,V 6 V 7,V 1 V 10,V 4 V 10,V 2 V 10,V 7
-0.190 -0.160 -0.159 -0.155 -0.143

V 4,V 3 V 7,V 5 V 10,V 1 V 9,V 8 V 9,V 3
-0.135 -0.134 -0.124 -0.120 -0.114

V 8,V 5 V 5,V 4 V 10,V 6 V 9,V 4 V 3,V 2
-0.113 -0.111 -0.104 -0.097 -0.096

DISTRIBUTION OF STANDARDIZED RESIDUALS

			RANGE	FREQ	PERCENT
!	!				
40-	-				
!	!				
!	!				
!	!				
!	!				
30-	-				
!	!		1 -0.5 - --	0	0.00%
!	!		2 -0.4 - -0.5	0	0.00%
!	!	*	3 -0.3 - -0.4	1	1.82%
!	!	*	4 -0.2 - -0.3	3	5.45%
20-	-	*	5 -0.1 - -0.2	14	25.45%
!	!	*	6 0.0 - -0.1	23	41.82%
!	!	*	7 0.1 - 0.0	14	25.45%
!	!	* * *	8 0.2 - 0.1	0	0.00%
!	!	* * *	9 0.3 - 0.2	0	0.00%
10-	-	* * *	A 0.4 - 0.3	0	0.00%
!	!	* * *	B 0.5 - 0.4	0	0.00%
!	!	* * *	C ++ - 0.5	0	0.00%
!	!	* * * *			
!	!	* * * * *			
			TOTAL	55	100.00%

1 2 3 4 5 6 7 8 9 A B C EACH "*" REPRESENTS 2 RESIDUALS

MEASUREMENT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS

$$\text{ARROW} = V1 = .336 * F2 + .479 * F3 + 1.000 E1$$

.104 .100
3.216 4.766

$$\text{WORD} = V2 = .348 * F1 + .679 * F3 + 1.000 E2$$

.116 .097
3.010 7.009

$$\text{TTT} = V3 = .603 * F2 + .214 * F3 + 1.000 E3$$

.121 .129
4.982 1.661

$$\text{VERBAL} = V4 = .791 * F1 + .584 * F3 + 1.000 E4$$

.101 .141
7.826 4.135

$$\text{SIMSPAT} = V5 = .413 * F2 + .669 * F3 + .271 * F4$$

.112	.114	.132
3.694	5.891	2.048

+ 1.000 E5

COMSPAT=V6 = -.089*F1 + .520*F2 + .233*F3
 .094 .099 .118
 -.949 5.275 1.974

+ .238*F4 + 1.000 E6
 .124
 1.927

ABSTRACT=V7 = -.045*F1 + .439*F3 + .556*F4
 .144 .138 .159
 -.314 3.187 3.503

+ 1.000 E7

CONDITIO=V8 = .006*F1 + .345*F3 + .412*F4
 .133 .127 .134
 .042 2.724 3.078

+ 1.000 E8

SYLL =V9 = .461*F1 + .689*F2 - .099*F3
 .151 .163 .139
 3.054 4.232 -.711

+ .611*F4 + 1.000 E9
 .181
 3.375

MULTI =V10 = -.234*F1 + .646*F2 + .349*F3
 .109 .122 .142
 -2.138 5.317 2.450

+ .107*F4 + 1.000 E10
 .140
 .766

VARIANCES OF INDEPENDENT VARIABLES

V		F	
---		---	
I F1 -	F1	1.000	I
I			I
I			I
I			I
I F2 -	F2	1.000	I
I			I
I			I
I			I
I F3 -	F3	1.000	I
I			I
I			I
I			I
I F4 -	F4	1.000	I
I			I
I			I
I			I

VARIANCES OF INDEPENDENT VARIABLES

E		D	
---		---	
E1 -ARROW	.658*		I
	.092		I
	7.127		I
			I
E2 - WORD	.418*		I
	.068		I
	6.103		I
			I
E3 - TTT	.568*		I
	.120		I
	4.722		I
			I
E4 -VERBAL	.000*		I
	.056		I
	.000		I
			I
E5 -SIMSPAT	.383*		I
	.085		I
	4.492		I
			I
E6 -COMSPAT	.690*		I
	.111		I
	6.234		I
			I
E7 -ABSTRACT	.510*		I

	.163	I
	3.127	I
E8 -CONDITIO	.722*	I
	.132	I
	5.480	I
E9 - SYLL	.000*	I
	.000	I
	:0000000.000	I
E10 -MULTI	.425*	I
	.114	I
	3.714	I
		I

STANDARDIZED SOLUTION:

R-SQUARED

ARROW	=V1 =	.336*F2	+ .479*F3	+ .811 E1		.342
WORD	=V2 =	.348*F1	+ .679*F3	+ .646 E2		.582
TTT	=V3 =	.610*F2	+ .217*F3	+ .762 E3		.419
VERBAL	=V4 =	.804*F1	+ .594*F3	+ .000 E4		1.000
SIMSPAT	=V5 =	.399*F2	+ .645*F3	+ .262*F4	+ .597 E5	.644
COMSPAT	=V6 =	-.086*F1	+ .501*F2	+ .224*F3	+ .229*F4	
		+ .799 E6				.361
ABSTRACT	=V7 =	-.045*F1	+ .435*F3	+ .552*F4	+ .709 E7	.497
CONDITIO	=V8 =	.006*F1	+ .343*F3	+ .410*F4	+ .845 E8	.286
SYLL	=V9 =	.446*F1	+ .666*F2	-.096*F3	+ .591*F4	
		+ .000 E9				1.000
MULTI	=V10 =	-.231*F1	+ .637*F2	+ .344*F3	+ .106*F4	
		+ .642 E10				.588

 END OF METHOD

STATISTICS FOR MULTIPLE POPULATION ANALYSIS

ALL EQUALITY CONSTRAINTS WERE CORRECTLY IMPOSED

NOTE: CONSTRAINTS APPEAR TO BE LINEARILY DEPENDENT, 1 DEGREE(S)
 OF FREEDOM ADJUSTED

*** WARNING *** TEST RESULTS MAY NOT BE APPROPRIATE DUE TO
 CONDITION CODE

GOODNESS OF FIT SUMMARY

INDEPENDENCE MODEL CHI-SQUARE = 361.960 ON 88 DEGREES OF FREEDOM

INDEPENDENCE AIC = 185.95973 INDEPENDENCE CAIC = -164.59765
MODEL AIC = -74.19283 MODEL CAIC = -337.11087

CHI-SQUARE = 57.807 BASED ON 66 DEGREES OF FREEDOM
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS 0.75372

BENTLER-BONETT NORMED FIT INDEX= 0.840
BENTLER-BONETT NONNORMED FIT INDEX= 1.040
COMPARATIVE FIT INDEX (CFI) = 1.000

ITERATIVE SUMMARY

ITERATION	PARAMETER ABS CHANGE	ALPHA	FUNCTION
1	0.431299	1.00000	1.64948
2	0.200842	1.00000	1.01298
3	0.113402	1.00000	0.59798
4	0.062814	1.00000	0.52026
5	0.097965	0.12500	0.51803
6	0.039787	1.00000	0.48510
7	0.113527	0.25000	0.47997
8	0.058886	1.00000	0.43002
9	0.021710	1.00000	0.40769
10	0.009288	1.00000	0.40464
11	0.005250	1.00000	0.40436
12	0.003167	1.00000	0.40428
13	0.001424	1.00000	0.40426
14	0.000883	1.00000	0.40425

LAGRANGE MULTIPLIER TEST (FOR RELEASING CONSTRAINTS)

CONSTRAINTS TO BE RELEASED ARE:

CONSTRAINTS FROM GROUP 3

CONSTR: 1 (1.V1.F2)-(2.V1.F2)=0;
CONSTR: 2 (1.V1.F2)-(3.V1.F2)=0;
CONSTR: 3 (1.V1.F3)-(2.V1.F3)=0;
CONSTR: 4 (1.V1.F3)-(3.V1.F3)=0;
CONSTR: 5 (1.V2.F1)-(2.V2.F1)=0;
CONSTR: 6 (1.V2.F1)-(3.V2.F1)=0;
CONSTR: 7 (1.V2.F3)-(2.V2.F3)=0;
CONSTR: 8 (1.V2.F3)-(3.V2.F3)=0;
CONSTR: 9 (1.V3.F2)-(3.V3.F2)=0;
CONSTR: 10 (1.V3.F3)-(3.V3.F3)=0;
CONSTR: 11 (1.V4.F1)-(3.V4.F1)=0;

CONSTR: 12 (1,V4.F3)-(3,V4.F3)=0:
 CONSTR: 13 (2,V5.F2)-(3,V5.F2)=0:
 CONSTR: 14 (2,V5.F3)-(3,V5.F3)=0:
 CONSTR: 15 (2,V5.F4)-(3,V5.F4)=0:
 CONSTR: 16 (1,V6.F1)-(2,V6.F1)=0:
 CONSTR: 17 (1,V6.F1)-(3,V6.F1)=0:
 CONSTR: 18 (1,V6.F2)-(2,V6.F2)=0:
 CONSTR: 19 (1,V6.F2)-(3,V6.F2)=0:
 CONSTR: 20 (1,V6.F3)-(2,V6.F3)=0:
 CONSTR: 21 (1,V6.F3)-(3,V6.F3)=0:
 CONSTR: 22 (1,V6.F4)-(2,V6.F4)=0;
 CONSTR: 23 (1,V6.F4)-(3,V6.F4)=0:
 CONSTR: 24 (2,V7.F1)-(3,V7.F1)=0;
 CONSTR: 25 (2,V7.F3)-(3,V7.F3)=0:
 CONSTR: 26 (2,V7.F4)-(3,V7.F4)=0:
 CONSTR: 27 (2,V8.F1)-(3,V8.F1)=0:
 CONSTR: 28 (2,V8.F3)-(3,V8.F3)=0;
 CONSTR: 29 (2,V8.F4)-(3,V8.F4)=0:
 CONSTR: 30 (2,V9.F1)-(3,V9.F1)=0:
 CONSTR: 31 (2,V9.F2)-(3,V9.F2)=0:
 CONSTR: 32 (2,V9.F3)-(3,V9.F3)=0:
 CONSTR: 33 (2,V9.F4)-(3,V9.F4)=0:
 CONSTR: 34 (2,V10.F1)-(3,V10.F1)=0:
 CONSTR: 35 (2,V10.F2)-(3,V10.F2)=0:
 CONSTR: 36 (2,V10.F3)-(3,V10.F3)=0;
 CONSTR: 37 (2,V10.F4)-(3,V10.F4)=0:
 CONSTR: 38 (1,E1.E1)-(2,E1.E1)=0:
 CONSTR: 39 (1,E1.E1)-(3,E1.E1)=0:
 CONSTR: 40 (1,E2.E2)-(2,E2.E2)=0:
 CONSTR: 41 (1,E2.E2)-(3,E2.E2)=0;
 CONSTR: 42 (1,E3.E3)-(3,E3.E3)=0:
 CONSTR: 43 (1,E4.E4)-(3,E4.E4)=0:
 CONSTR: 44 (2,E5.E5)-(3,E5.E5)=0;
 CONSTR: 45 (2,E6.E6)-(3,E6.E6)=0:
 CONSTR: 46 (2,E7.E7)-(3,E7.E7)=0:
 CONSTR: 47 (2,E8.E8)-(3,E8.E8)=0;
 CONSTR: 48 (2,E9.E9)-(3,E9.E9)=0;
 CONSTR: 49 (2,E10.E10)-(3,E10.E10)=0:

UNIVARIATE TEST STATISTICS:

NO	CONSTRAINT	CHI-SQUARE	PROBABILITY
--	-----	-----	-----
1	CONSTR: 1	0.071	0.789
2	CONSTR: 2	0.000	0.992
3	CONSTR: 3	1.729	0.188
4	CONSTR: 4	1.551	0.213
5	CONSTR: 5	0.055	0.815
6	CONSTR: 6	0.003	0.956
7	CONSTR: 7	0.013	0.908

8	CONSTR: 8	0.218	0.641
9	CONSTR: 9	0.031	0.859
10	CONSTR: 10	0.957	0.328
11	CONSTR: 11	0.271	0.602
12	CONSTR: 12	1.099	0.294
13	CONSTR: 13	0.257	0.612
14	CONSTR: 14	0.104	0.747
15	CONSTR: 15	0.002	0.965
16	CONSTR: 16	1.798	0.180
17	CONSTR: 17	2.319	0.128
18	CONSTR: 18	0.270	0.603
19	CONSTR: 19	1.200	0.273
20	CONSTR: 20	1.728	0.189
21	CONSTR: 21	0.007	0.934
22	CONSTR: 22	0.212	0.645
23	CONSTR: 23	0.280	0.597
24	CONSTR: 24	0.273	0.602
25	CONSTR: 25	1.556	0.212
26	CONSTR: 26	0.853	0.356
27	CONSTR: 27	0.752	0.386
28	CONSTR: 28	0.098	0.754
29	CONSTR: 29	1.706	0.192
30	CONSTR: 30	0.011	0.916
31	CONSTR: 31	0.230	0.631
32	CONSTR: 32	0.184	0.668
33	CONSTR: 33	0.091	0.763
34	CONSTR: 34	0.167	0.683
35	CONSTR: 35	0.932	0.334
36	CONSTR: 36	3.590	0.058
37	CONSTR: 37	1.277	0.258
38	CONSTR: 38	1.269	0.260
39	CONSTR: 39	0.292	0.589
40	CONSTR: 40	0.596	0.440
41	CONSTR: 41	0.114	0.735
42	CONSTR: 42	1.070	0.301
43	CONSTR: 43	NOT TESTED DUE TO NUMERICAL PROBLEMS.	
44	CONSTR: 44	0.028	0.866
45	CONSTR: 45	0.151	0.698
46	CONSTR: 46	0.187	0.666
47	CONSTR: 47	0.511	0.475
48	CONSTR: 48	0.001	0.975
49	CONSTR: 49	0.215	0.643

CUMULATIVE MULTIVARIATE STATISTICS
INCREMENT

UNIVARIATE

STEP PARAMETER CHI-SQUARE D.F. PROBABILITY
PROBABILITY

CHI-SQUARE

STEP	PARAMETER	CHI-SQUARE	D.F.	PROBABILITY	CHI-SQUARE	PROBABILITY
1	CONSTR: 36	3.590	1	0.058	3.590	0.058
2	CONSTR: 17	5.869	2	0.053	2.279	0.131
3	CONSTR: 29	7.665	3	0.053	1.796	0.180
4	CONSTR: 10	9.406	4	0.052	1.742	0.187
5	CONSTR: 47	10.926	5	0.053	1.519	0.218
6	CONSTR: 38	12.312	6	0.055	1.386	0.239
7	CONSTR: 3	14.234	7	0.047	1.923	0.166
8	CONSTR: 19	15.398	8	0.052	1.164	0.281
9	CONSTR: 25	16.447	9	0.058	1.049	0.306
10	CONSTR: 37	17.591	10	0.062	1.143	0.285
11	CONSTR: 42	18.492	11	0.071	0.901	0.342
12	CONSTR: 20	19.399	12	0.079	0.907	0.341
13	CONSTR: 21	21.243	13	0.068	1.844	0.174
14	CONSTR: 26	22.040	14	0.078	0.797	0.372
15	CONSTR: 16	22.792	15	0.089	0.752	0.386
16	CONSTR: 12	23.237	16	0.108	0.444	0.505
17	CONSTR: 11	24.059	17	0.118	0.822	0.365
18	CONSTR: 41	24.849	18	0.129	0.790	0.374
19	CONSTR: 8	25.250	19	0.152	0.401	0.526
20	CONSTR: 35	25.627	20	0.178	0.376	0.540
21	CONSTR: 13	26.054	21	0.204	0.427	0.513
22	CONSTR: 2	26.481	22	0.232	0.427	0.514
23	CONSTR: 30	26.690	23	0.269	0.209	0.647
24	CONSTR: 4	26.888	24	0.310	0.198	0.657
25	CONSTR: 1	27.153	25	0.348	0.266	0.606
26	CONSTR: 15	27.275	26	0.395	0.122	0.727
27	CONSTR: 22	27.447	27	0.440	0.171	0.679
28	CONSTR: 18	27.633	28	0.484	0.186	0.666
29	CONSTR: 23	27.794	29	0.529	0.161	0.689
30	CONSTR: 9	27.953	30	0.573	0.160	0.689
31	CONSTR: 24	28.093	31	0.616	0.140	0.708
32	CONSTR: 32	28.192	32	0.660	0.099	0.753
33	CONSTR: 14	28.439	33	0.694	0.248	0.619
34	CONSTR: 31	28.705	34	0.725	0.265	0.607
35	CONSTR: 28	28.877	35	0.758	0.173	0.678
36	CONSTR: 34	28.960	36	0.791	0.083	0.774
37	CONSTR: 6	29.120	37	0.819	0.160	0.689
38	CONSTR: 45	29.181	38	0.847	0.061	0.804
39	CONSTR: 27	29.235	39	0.872	0.054	0.816
40	CONSTR: 39	29.288	40	0.894	0.053	0.818
41	CONSTR: 33	29.364	41	0.912	0.076	0.783
42	CONSTR: 40	29.397	42	0.929	0.033	0.857
43	CONSTR: 7	29.649	43	0.939	0.253	0.615
44	CONSTR: 5	29.760	44	0.950	0.111	0.739

45	CONSTR:	44	29.816	45	0.960	0.056	0.813
46	CONSTR:	49	29.840	46	0.969	0.024	0.877
47	CONSTR:	46	29.857	47	0.976	0.017	0.898
48	CONSTR:	48	29.863	48	0.981	0.006	0.936