
Craig P. Speelman. B.Sc. (Hons)

Being a dissertation submitted as a requirement for the Degree of Doctor of Philosophy within the University of Western Australia, Department of Psychology

January, 1991
Abstract

There were two main aims of the research described in this thesis. The first was to investigate the determinants of transfer of training. The second was to examine the effects of transfer on the shape of learning functions. In parallel with these aims was the aim to evaluate and extend certain features of Anderson's ACT* theory of skill acquisition.

Part 1 of the thesis was concerned with the determinants of transfer. This issue was investigated by comparing two accounts of a transfer phenomenon called the contextual interference effect. In brief, this effect typically involves comparing performance in two presentation conditions: (1) random presentation, where various tasks are performed in a random order; and (2) blocked presentation, where a block of trials of one task are performed together, followed by a block of another task, and so on. Performance in blocked training is usually faster than during random training. However, this difference is usually reversed when presentation is switched to random only. One account of this effect is provided by the ACT* theory. This theory holds that skilled performance is based on hierarchies of condition-action pairs called productions. These productions involve a condition that needs to be satisfied before a particular action can proceed. The ACT* theory predicts that developing a set of productions to perform one task will benefit performance on another task to the extent that the two tasks require the same productions. In other words, transfer of training is a function of the number of "old" productions that are useful in a new situation. The ACT* account of the contextual interference effect basically suggests that blocked training does not encourage development of productions which are required with random transfer items. This account of the effect was contrasted with one termed the intratrial processing account. This account suggests that the important feature
of random training which provides the advantage during transfer is that subjects in this condition have extensive practice at loading new solution methods into working memory. This practice is not provided with blocked training but is important for efficient performance with random transfer items. These two accounts of the contextual interference effect were compared in Experiments 1, 2 and 3. Subjects were given extensive practice with syllogistic reasoning problems that varied with respect to several presentation features. These features were shown to affect the type of processing strategies that subjects adopted to solve the problems. In turn, the various strategies were shown to vary in their appropriateness for problems with different presentation features. The occurrence of these strategy differences was accounted for by the ACT* theory but not the intratrial processing account. The ACT* account was not only able to predict these differences during training, but was able to relate these to performance differences during transfer on the basis of the extent to which the various strategies shared identical productions. Thus it was concluded that the ACT* theory provided a superior account of the contextual interference effect compared to the one provided by the intratrial processing account.

The second part of the thesis is concerned with the rate at which skills are acquired. ACT* states that learning proceeds in accordance with the Power Law of Practice. This law says that the logarithm of performance time is a straight-line function of the logarithm of the amount of practice at a task. The learning rate is indicated by the gradient of this straight line. This law was extended into the situation where learning a task involved continued practice at old skills that were useful for performance of the task, and the development of new skills to complement the old skills as required by the new task. An extended power function was developed to describe this situation. This new function involved the combination of two simple power functions which
described the separate improvement of the old and new task components. On the basis of an assumption that learning proceeds at a relatively constant rate for each person, certain predictions could be derived from the extended power function concerning the rate at which learning would appear to proceed in any given situation. The basic prediction was that when old and new skills are combined to perform a new task, the rate at which performance of this task will improve will be slower than the rate at which the old skills were learned. This attenuation is in spite of the fact that the old and new skills are improving at the same constant rate. The amount by which learning rate is attenuated is moderated by the relative amounts of practice each set of skills had prior to their combination, and the ratio of the number of old to new skills involved in the new task. Experiments 4 - 8 all provided evidence in support of this view of the combination of skills that differ with respect to amount of practice. However the amount by which learning rate was observed to be attenuated was greater than predicted by the extended power function. Furthermore, the combination of old and new skills was also found to be associated with an increase in performance asymptote. A new version of the extended power function was derived which incorporated this change in asymptote. This function was able to account for not only the increase in asymptote, but also the greater attenuation of learning rate. The major conclusion of this section of the thesis is that the combination of old and new skills can be described by a combination of two power functions representing the two components. This combination of two functions accounts for the attenuation of learning rate. However, the cause of the increase in asymptote was unknown. Various possibilities are discussed.

The main conclusion of the thesis is that the proposed model of transfer provides a useful account in that it is able to predict the rate at which particular tasks are learned. The assumption that each person has some intrinsic learning
rate that is relatively constant was shown to be a reasonable one although various issues concerning the measurement of learning rate and the effects of task conditions on learning rate were discussed in relation to the difficulty of effectively evaluating this assumption.
Sources Statement

The present thesis describes original research I have undertaken in the Department of Psychology, at the University of Western Australia. Any theories and techniques that are not my own have been acknowledged. The theoretical contributions in this thesis are my original work and have not been submitted for any other degree.

Signed
Acknowledgements

I would like to express my gratitude to Kim Kirsner for his supervision of this thesis. Kim provided invaluable advice, direction and support throughout the whole of this project. I am also grateful to Peter Standen for essential programming assistance, and to Murray Maybery for valuable editorial comment.

This thesis is dedicated to my wife Marisa, whose love and support, understanding and patience through all stages of this project, particularly the difficult ones, enabled me to persevere.
## Contents

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Title Page</strong></td>
<td>(i)</td>
</tr>
<tr>
<td></td>
<td><strong>Abstract</strong></td>
<td>(ii)</td>
</tr>
<tr>
<td></td>
<td><strong>Sources Statement</strong></td>
<td>(vi)</td>
</tr>
<tr>
<td></td>
<td><strong>Acknowledgements</strong></td>
<td>(vii)</td>
</tr>
<tr>
<td></td>
<td><strong>Contents</strong></td>
<td>(viii)</td>
</tr>
<tr>
<td></td>
<td><strong>Figures</strong></td>
<td>(ix)</td>
</tr>
<tr>
<td></td>
<td><strong>Tables</strong></td>
<td>(xiii)</td>
</tr>
<tr>
<td>Chapter 1</td>
<td><strong>PART 1 Transfer of Training and the Contextual Interference Effect</strong></td>
<td></td>
</tr>
<tr>
<td>Chapter 2</td>
<td><em><em>ACT</em> and intratrial processing accounts of the contextual interference effect</em>*</td>
<td>36</td>
</tr>
<tr>
<td>Chapter 3</td>
<td><strong>Chapter 3 Experiments 1, 2 and 3</strong></td>
<td>56</td>
</tr>
<tr>
<td>PART 2</td>
<td><strong>PART 2 The Shape of Learning Functions following Transfer</strong></td>
<td>130</td>
</tr>
<tr>
<td>Chapter 4</td>
<td><strong>Chapter 4 Development of a model describing changes in the shape of learning functions following partial transfer</strong></td>
<td>130</td>
</tr>
<tr>
<td>Chapter 5</td>
<td><strong>Chapter 5 Experiments 4, 5 and 6</strong></td>
<td>170</td>
</tr>
<tr>
<td>Chapter 6</td>
<td><strong>Chapter 6 Experiment 7</strong></td>
<td>242</td>
</tr>
<tr>
<td>Chapter 7</td>
<td><strong>Chapter 7 Experiment 8</strong></td>
<td>304</td>
</tr>
<tr>
<td>Chapter 8</td>
<td><strong>Chapter 8 General Discussion</strong></td>
<td>338</td>
</tr>
<tr>
<td></td>
<td><strong>References</strong></td>
<td>362</td>
</tr>
</tbody>
</table>
Figures

Figure 3.1  Goal structure underlying solution of syllogisms with random presentation  
Page 59

Figure 3.2  A summary of the design of Training trials in Experiment 1  
Page 75

Figure 3.3  Mean Premise and Conclusion RTs in Experiment 1  
Page 78

Figure 3.4  Mean Premise and Conclusion RTs for combined Random and Blocked Training conditions in Experiment 1 plotted on log-log axes  
Page 80

Figure 3.5  Serial position curves of mean Premise and Conclusion RTs during Training in Experiment 1  
Page 84

Figure 3.6  Mean Premise and Conclusion RTs during Training (Random) and Transfer (Random) phases of Experiment 1  
Page 89

Figure 3.7  Mean Premise and Conclusion RTs during Training (Blocked) and Transfer (Blocked) phases of Experiment 1  
Page 90

Figure 3.8  Mean Premise and Conclusion RTs during Training (Random) and Transfer (Blocked) phases of Experiment 1  
Page 91

Figure 3.9  Mean Premise and Conclusion RTs during Training (Blocked) and Transfer (Random) phases of Experiment 1  
Page 92

Figure 3.10  Mean Premise and Conclusion RTs during Training (Alternating) and Transfer (Random) phases of Experiment 2  
Page 102

Figure 3.11  Mean Premise and Conclusion RTs during Training (Highlight) and Transfer (Random) phases of Experiment 3  
Page 111

Figure 3.12  Mean Premise and Conclusion RTs during Training and Transfer phases of Experiments 1, 2 and 3  
Page 117

Figure 4.1  Learning rate as a function of the ratio of the number of steps in old skills vs. the number of steps in new skills (as predicted by a version of Equation 6)  
Page 141
Figure 4.2 Demonstration that a practice function with a 'slow' learning rate may be the tail-end of a function with a faster rate

Figure 4.3 Learning rate as a function of the amount of extra practice of old skills in comparison to new skills (as predicted by a version of Equation 6)

Figure 4.4 S-R mappings in the four conditions used in the experiment reported in Duncan (1977)

Figure 4.5 Mean RTs reported in Duncan (1977)

Figure 4.6 Data taken from an unpublished study by Fitts and Switzer, reported in Fitts (1964)

Figure 4.7 Data reported in Woltz (1988)

Figure 4.8 Data reported in Snyder & Pronko (1952)

Figure 5.1 Mean Premise and Conclusion RTs for both Experimental and Control groups, during Training and Transfer phases of Experiment 4

Figure 5.2 Mean Premise and Conclusion RTs for the Control group during Training (Highlight) and Transfer (Highlight) phases of Experiment 4

Figure 5.3 Mean Premise and Conclusion RTs for the Experimental group during Training (Highlight) and Transfer (Random) phases of Experiment 4

Figure 5.4 Mean Premise and Conclusion RTs for H.S. during Training and Transfer phases of Experiment 5

Figure 5.5 Mean Premise and Conclusion RTs for M.S. during Training and Transfer phases of Experiment 5

Figure 5.6 Mean Total RTs for H.S. and M.S. during Training and Transfer phases of Experiment 5

Figure 5.7 Mean Premise and Conclusion RTs for both Experimental and Control groups, during Training and Transfer phases of Experiment 6

Figure 5.8 Mean Premise and Conclusion RTs for the Control group during Training (Alternating) and Transfer (Alternating) phases of Experiment 6
Figure 5.9  Mean Premise and Conclusion RTs for the Experimental group during Training (Alternating) and Transfer (Random) phases of Experiment 6

Figure 5.10  Mean Premise RTs for H.S. during Training and Transfer phases of Experiment 5

Figure 5.11  Mean Total RTs for H.S. during Training and Transfer phases of Experiment 5

Figure 5.12  Mean Premise RTs for M.S. during Training and Transfer phases of Experiment 5

Figure 5.13  Mean Total RTs for M.S. during Training and Transfer phases of Experiment 5

Figure 5.14  Mean Premise RTs for the Experimental group during Training (Highlight) and Transfer (Random) phases of Experiment 4

Figure 6.1  Basic stimulus configuration presented in Experiment 7

Figure 6.2  Goal structure underlying solution of simple version of tank task

Figure 6.3  Decision tree underlying goal 8 in simple version of the tank task

Figure 6.4  Display sequence in the complex version of the tank task presented during the Transfer phase of Experiment 7

Figure 6.5  Goal structure underlying solution of complex version of tank task

Figure 6.6  Example of stimulus configuration presented in the simple version of the tank task in Experiment 7

Figure 6.7  Example of stimulus configuration presented in the complex version of the tank task in Experiment 7

Figure 6.8  Mean Flow-Route RTs during Training and Transfer phases of Experiment 7 (with outlier)

Figure 6.9  Mean Flow-Route RTs during Training and Transfer phases of Experiment 7 (without outlier)

Figure 6.10  Mean 3-tank RTs during Transfer phase of Experiment 7

Figure 6.11  Mean 1-tank RTs during Transfer phase of Experiment 7
Figure 6.12  Mean Conclusion RTs during Training and Transfer phases of Experiment 7  

Figure 6.13  Mean Total RTs during Training and Transfer phases of Experiment 7  

Figure 7.1  Data reported in Snyder and Pronko (1952)  

Figure 7.2  Mean Premise and Conclusion RTs during Training, Transfer 1 and Transfer 2 phases of Experiment 8  

Figure 7.3  Mean Premise RTs during Training, Transfer 1 and Transfer 2 phases of Experiment 8  

Figure 7.4  Mean Conclusion RTs during Training, Transfer 1 and Transfer 2 phases of Experiment 8  

Figure 7.5  Mean Total RTs during Training, Transfer 1 and Transfer 2 phases of Experiment 8
### Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1.1</td>
<td>Example of a set of domain-specific productions for solving algebra problems</td>
<td>20</td>
</tr>
<tr>
<td>Table 3.1</td>
<td>Summary of predicted results for Experiment 1 based on the ACT* and intratrial processing accounts of the contextual interference effect</td>
<td>62</td>
</tr>
<tr>
<td>Table 3.2</td>
<td>Sample syllogism used in Experiment 1</td>
<td>72</td>
</tr>
<tr>
<td>Table 3.3</td>
<td>Sample syllogism used in Experiment 3</td>
<td>109</td>
</tr>
<tr>
<td>Table 4.1</td>
<td>Power functions describing performance with three text-editors during Training and Transfer phases of an experiment reported in Singley &amp; Anderson (1985)</td>
<td>148</td>
</tr>
<tr>
<td>Table 4.2</td>
<td>Equations of power functions fitted to data reported in Duncan (1977)</td>
<td>155</td>
</tr>
<tr>
<td>Table 4.3</td>
<td>Equations of power functions fitted to data from an unpublished study by Fitts and Switzer and reported in Fitts (1964)</td>
<td>158</td>
</tr>
<tr>
<td>Table 5.1</td>
<td>Parameters of power functions fitted to Premise and Conclusion RTs during Training and Transfer phases of Experiment 4</td>
<td>189</td>
</tr>
<tr>
<td>Table 5.2</td>
<td>Parameters of power functions fitted to Premise, Conclusion and Total RTs of H.S. during Training and Transfer phases of Experiment 5</td>
<td>203</td>
</tr>
<tr>
<td>Table 5.3</td>
<td>Parameters of power functions fitted to Premise, Conclusion and Total RTs of M.S. during Training and Transfer phases of Experiment 5</td>
<td>205</td>
</tr>
<tr>
<td>Table 5.4</td>
<td>Parameters of power functions fitted to Premise and Conclusion RTs during Training and Transfer phases of Experiment 6</td>
<td>225</td>
</tr>
<tr>
<td>Table 5.5</td>
<td>Parameters of power functions fitted to Premise and Total RTs of H.S. and M.S. during Training and Transfer phases of Experiment 5</td>
<td>232</td>
</tr>
<tr>
<td>Table 5.6</td>
<td>Parameters of power functions fitted to Premise RTs during Training and Transfer phases of Experiment 4</td>
<td>240</td>
</tr>
<tr>
<td>Table 6.1</td>
<td>Summary table of predictions for performance during Transfer phase of Experiment 7</td>
<td>261</td>
</tr>
<tr>
<td>Table</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>-------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>6.2</td>
<td>Parameters of power functions fitted to Flow-Route RTs during Training and Transfer phases of Experiment 7</td>
<td>276</td>
</tr>
<tr>
<td>6.3</td>
<td>Parameters of power functions fitted to 3-tank RTs during Transfer phase of Experiment 7</td>
<td>284</td>
</tr>
<tr>
<td>6.4</td>
<td>Parameters of power functions fitted to 1-tank RTs during Transfer phase of Experiment 7</td>
<td>289</td>
</tr>
<tr>
<td>6.5</td>
<td>Parameters of power functions fitted to Conclusion RTs during Training and Transfer phases of Experiment 7</td>
<td>294</td>
</tr>
<tr>
<td>6.6</td>
<td>Parameters of power functions fitted to Total RTs during Training and Transfer phases of Experiment 7</td>
<td>298</td>
</tr>
<tr>
<td>7.1</td>
<td>Parameters of power functions fitted to Premise RTs during Training, Transfer 1 and Transfer 2 phases of Experiment 8</td>
<td>319</td>
</tr>
<tr>
<td>7.2</td>
<td>Parameters of power functions fitted to Conclusion RTs during Training, Transfer 1 and Transfer 2 phases of Experiment 8</td>
<td>326</td>
</tr>
<tr>
<td>7.3</td>
<td>Parameters of power functions fitted to Total RTs during Training, Transfer 1 and Transfer 2 phases of Experiment 8</td>
<td>333</td>
</tr>
</tbody>
</table>
# Chapter 1

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Preface</td>
<td>2</td>
</tr>
<tr>
<td>1.2 Historical Summary</td>
<td>4</td>
</tr>
<tr>
<td>1.2.1 Plateaus</td>
<td>4</td>
</tr>
<tr>
<td>1.2.2 Part-Whole Training</td>
<td>5</td>
</tr>
<tr>
<td>1.2.3 Massed versus Distributed Practice</td>
<td>6</td>
</tr>
<tr>
<td>1.2.4 Knowledge of Results</td>
<td>6</td>
</tr>
<tr>
<td>1.2.5 Transfer of Training</td>
<td>7</td>
</tr>
<tr>
<td>1.2.6 The Status of Cognitive Skill</td>
<td>8</td>
</tr>
<tr>
<td>1.3 Fitts and the Phases of Skill Acquisition</td>
<td>9</td>
</tr>
<tr>
<td>1.4 The Cognitive Architecture Proposed by ACT*</td>
<td>12</td>
</tr>
<tr>
<td>1.5 The ACT* Theory of Skill Acquisition</td>
<td>14</td>
</tr>
<tr>
<td>1.6 ACT* and Fitts’ Phases of Skill Acquisition</td>
<td>25</td>
</tr>
<tr>
<td>1.7 The Acquisition of Expertise</td>
<td>26</td>
</tr>
<tr>
<td>1.7.1 Experts Possess Compiled Productions</td>
<td>27</td>
</tr>
<tr>
<td>1.7.2 Experts Use Hierarchical Strategies</td>
<td>33</td>
</tr>
</tbody>
</table>
1.1 Preface

The experimental investigations reported in this thesis had two aims. The first aim was to examine the determinants of the transfer of cognitive skill. The second aim was to investigate the effect of transfer on the shape of learning functions. The thesis is divided into two parts which focus on these two aims separately, although much of what is achieved in the first part is used as a platform for discussion and experimentation in the second part.

J.R. Anderson's ACT* theory of skill acquisition is used as a theoretical focus for the thesis. In particular the experiments were aimed at examining the ACT* account of transfer and the shape of learning functions.

The main reason for focussing on the ACT* theory is that, at present, this theory is the most developed theory of skill acquisition. No other theory is as comprehensive in terms of specifying the processes underlying the acquisition of cognitive skill. The comprehensiveness of the ACT* theory of skill acquisition stems from the fact that it forms a subset of the ACT* theory of cognition (Anderson, 1983). In Newell's terms (e.g., Newell, 1989), ACT* is a unified theory of cognition, in that it proposes a cognitive architecture with the aim of accounting for all known cognitive processes. The only other theory of skill acquisition that approaches the development of the ACT* theory is SOAR (Laird, Newell & Rosenbloom, 1987). The reason why ACT* was chosen as the main theoretical focus of this thesis in preference to SOAR is the greater ease in extracting predictions from ACT* concerning skilled performance. The SOAR theory has developed more from computer simulations of human behaviour than from experimental investigations. As a result it is difficult to relate theory and behaviour to the degree that is necessary in order to make predictions about performance. In contrast, ACT*
has developed mainly from experimental results informing simulations and so theory is tied more closely to behaviour. In any case, it is not obvious where the two theories would make different predictions. Even Newell (1989) was unable to identify any substantial differences in this sense between the two theories.

The ACT* theory makes strong claims about the relationship between performance and underlying cognitive processes. Two of these claims are evaluated in this thesis. The first claim concerns transfer of training. ACT* states that there will be transfer between tasks to the extent that performance of the tasks relies on common underlying cognitive structures. The second claim concerns the pattern of improvement in a task that comes with practice. ACT* provides an account of the improvement pattern that is typically observed. This account states that performance time is a function of the number of cognitive structures underlying performance. These cognitive structures are the same ones that ACT* suggests underlie transfer.

In this chapter the ACT* theory of skill acquisition will be described in relation to previous theories of skilled performance, both motor and cognitive, and the phenomena ACT* can account for. Following the introductory chapter will be Part 1 of the thesis. The main focus of this section will be the ACT* account of the transfer of cognitive skill. This account will be evaluated with respect to a transfer phenomenon known as the contextual interference effect. Experiments will be described that form a major portion of this evaluation. In these experiments the ACT* account of the contextual interference effect will be contrasted with the intratrial processing account (Carlson & Yaure, 1990). Part 2 of the thesis will be concerned with the shape of learning functions. The ACT* theory can account for the ubiquitous finding that learning functions are described by power functions.
(e.g., Newell & Rosenbloom, 1981). However some of the ACT* account is
deficient in terms of the effects of particular variables on the parameters of
these power functions. Using the ACT* account as a starting point, a
hypothesis will be developed that the learning rate of skills, and therefore the
shape of learning functions, is affected to a large extent by the combination of
old and new skills. Experiments will be described that examine this
hypothesis.

1.2 Historical Summary

From the late 1800's to the early 1960's research into skilled performance and
skill acquisition was largely devoid of any clear direction, theory or results.
The research was mainly applied in nature and concerned motor skills almost
exclusively. The focus was typically on discovering the best methods for
training motor skills, where the best methods were those that enabled the
fastest learning and enabled the greatest transfer to different situations and
tasks. The major areas of research during this period will be summarised
below. A more comprehensive treatment of this research can be found in

1.2.1 Plateaus

Research on complex skills really began with the work of Bryan and Harter
(1899). They trained subjects in the sending and receiving of Morse code
signals and examined the learning curves of these two tasks. The most
interesting result of this research was the observation of plateaus in the
learning curves of the receiving task. These plateaus represented periods
during training where subjects' performance did not improve. However
further training usually resulted in further improvement. Similar plateaus were
also reported during the training of typewriting (Book, 1925). The observation of plateaus led Bryan and Harter (1899) to propose that skill learning involved the acquisition of a hierarchy of habits. In a statement that has no doubt influenced today's theories of skill acquisition, Bryan and Harter described plateaus as periods where "lower-order habits are approaching their maximum development, but are not yet sufficiently automatic to leave the attention free to attack the higher-order habits" (p.357).

The concept of plateaus in learning curves has not enjoyed wide support. Since Book's research there has been little replication of the original findings (Adams, 1987; McGeoch, 1931, 1942). In addition, the extent to which plateaus can constrain theories of skill acquisition is questionable, since any variable that retards learning will produce them (Hunter, 1929). However, as indicated above, the hierarchical view of skilled behaviour is fundamental to all modern theories of skill acquisition.

1.2.2 Part-Whole Training

The aim of research in this area was to determine whether training on parts of a task and then combining these parts could be more efficient than training on the whole task. The benefits of such a training strategy are obvious, especially where the criterion task involves a large degree of cost or danger. For example, there is obvious value in a training method that would allow a substantial portion of pilot training to be achieved on the ground. Unfortunately, generalisations are not easy to arrive at in this area (McGeoch, 1952). The best approximation to a definitive answer to this research issue is that the relative benefits of part versus whole training is dependent on the task. Welford (1968) suggests that whole task training is the most efficient means of learning to perform tasks that involve highly interrelated activities,
such as flying an aircraft. In contrast, those tasks that involve components which are performed in a fixed order and are largely independent of each other appear to benefit most from part training (Welford, 1968). At present, no theory of skill acquisition purports to account for this relationship between task type and the most efficient means of training.

1.2.3 Massed versus Distributed Practice

The question of concern in this area was similar to that examined with the part-whole issue. That is, which type of practice is the most efficient training method: continuous practice in one long session (massed) or spaced practice in a number of sessions separated by time intervals of a certain duration (distributed)? All of the research in this area is beset with the problem of defining training efficiency when the cost of training is dependent on both the time spent training on the task and the length of time from the beginning of training to the final testing. Despite this problem though the most popular generalisation is that distributed training is the most efficient method (McGeoch, 1931; Welford, 1968). However, Adams (1987) has recently rejected this conclusion. After reviewing 100 years of research on this issue, Adams concluded that distributed practice does not improve learning relative to massed practice, but instead improves the momentary level of performance only. For example, subjects trained with massed practice and then examined under distributed conditions perform equally as well as subjects performing under distributed conditions throughout (Adams & Reynolds, 1954).

1.2.4 Knowledge of Results

The important concern in this area of research is whether performance improves without knowledge of results. Bartlett (1948) stated an answer to
this question: practice without knowledge of results does not improve performance. This generalisation has been widely accepted, although explanations of this effect are subject to debate (Adams, 1987), ranging from motivational to associative.

The effects of withdrawing and delaying feedback have been examined extensively and some clear results have emerged. Welford (1968) reports a number of these: (1) When knowledge of results in motor tasks is delayed, learning is slowed. However movement accuracy on most trials is affected to a small extent only. The slowing of learning is due to a greater proportion of trials which involve relatively larger movement errors. (2) When subjects perform some other activity between a trial and knowledge of the results of that trial, learning is slowed. (3) Learning is slowed as the gap between feedback and the following trial is lengthened. (4) Increasing the precision of feedback increases the accuracy of motor performance. (5) If knowledge of results is provided during training and then subsequently removed, performance deteriorates.

1.2.5 Transfer of Training

As with the part versus whole training issue, research examining transfer of training has always had an applied perspective. The general aim has been to determine when skills learned in one task can be transferred to performance of another task, with the hope of devising more efficient training methods. Efficiency in this context refers to some notion of the cost benefits of training, where training on a less expensive and time-consuming task may reduce the training time required on the criterion task. However, the most influential theory of the principles underlying transfer suggested that there were no free rides in skill acquisition. Thorndike (e.g., 1906) proposed that transfer
between tasks would occur to the extent that they shared identical elements. These elements have been interpreted as stimulus-response pairs (e.g., Singley & Anderson, 1989). This interpretation suggests that to perform a task that requires certain stimulus-response pairings, training must involve those same elements. Therefore the training task must involve the same elements as the criterion task which precludes any obvious economies of training. Thorndike's theory received some empirical support but has usually been criticised for being too restrictive in the specification of what can be transferred between tasks, that is, the nature of the identical elements (e.g., Singley & Anderson, 1989). This issue will be discussed in more detail in Chapter 2.

1.2.6 The Status of Cognitive Skill

The above summary of results and theories was mainly concerned with motor skills. In fact, until the 1960's, skills research was almost exclusively concerned with motor tasks. The most influential definition of what constitutes skilled performance was restricted to muscular performance (Pear, 1948). It was not until the cognitive revolution during the 1960's that cognitive performance began to be considered under the skill heading. The most important early discussion of skill acquisition as involving cognitive processes was by Fitts (1964). This work is important in the sense that it specified the phenomena that a theory of skill acquisition should explain. The influence of Fitts on the ACT* theory will be clear in the description of ACT* below.
1.3 Fitts and the Phases of Skill Acquisition

Although Fitts did not propose a theory of skill acquisition, his descriptions of the sequence of events involved in developing a skill were the first steps towards such a theory. One of his descriptions is worth quoting because it identifies the processes that need elucidation by theory:

An adult, or even a child of a few years of age, never begins the acquisition of a new form of skilled behaviour except from the background of many already existing, highly developed, both general and specific skills. Thus the initial state of our model is not that of a random model, but an already highly organised system processing language skills, concepts, and many efficient subroutines... The number of such identifiable highly developed skills in an adult is certainly in the hundreds, each having its own executive program and library of subroutines, many of the subroutines being shared with other skills.

The actual sequence of behaviour processes employed early in learning varies with the type of activity, of course, but might be somewhat as follows: The S observes or samples certain aspects of the environment, puts this information into short-term storage after some recoding, makes a decision such as selecting an appropriate subroutine which sets up a response pattern, executes a short behaviour sequence..., samples the internal and external feedback from this response plus additional stimulus information from the environment, recodes and stores this new information (in the process losing some of the information already in short-term storage), makes another decision which might be to use a different subroutine, and so on. As learning progresses, the subroutine
becomes longer, the executive routine or overall strategy is perfected, the stimulus sampling becomes less frequent and the coding more efficient, and different aspects of the activity become integrated or coordinated... As learning continues, overall performance may come to resemble more and more closely a continuous process. The overall program having now been perfected, frequent changes no longer need to be made in it. However, subroutines may continue slowly to become more efficient, and the S to become increasingly able to carry on the entire behaviour process while engaged simultaneously in other activities, with little or no interference between the two. (Fitts, 1964, p. 260)

Fitts (1964) suggested that skill acquisition involved three phases. The early phase was termed the "cognitive stage" by Fitts. This stage lasts for only a few trials while the subject comes to terms with instructions and develops performance strategies. According to Fitts, these strategies develop from general "sets" and strategies developed with previously learned tasks. Refinement of the performance strategy comes in the intermediate phase - "the associative stage." Features of the previously learned strategies that are appropriate to the new situation are strengthened on the basis of feedback, whereas inappropriate features are weakened. This process forms new associations between specific stimulus cues and appropriate responses. In the end phase - "the autonomous stage" - the components of the performance strategy slowly become more autonomous so that they are less subject to cognitive control or external interference. As a result skilled performance of the task requires increasingly less processing, which means that more processing resources can be used for other activities. During this phase skills continue to become faster and more efficient although the rate of improvement slows with practice.
Fitts provided no theoretical accounts of the processes he identified in the three phases, although he did point out where he thought existing theories were useful in this scheme. For instance, the selection of previously learned general sets and strategies for incorporation into new strategies draws on Crossman's (1959) general probability learning model. In this model, subjects are presumed to possess a repertoire of methods for performing a task. Each method is selected at random and the probability of its subsequent selection is dependent on its performance speed - the faster ones being more likely to be selected. This process predicts the typical power function speed-up found in skill acquisition (e.g., Newell & Rosenbloom, 1981). However, other features of Fitts' description of skill development are not specified to the same extent. The development from the initial stages of task performance of the "executive program", the "subroutines" and the relationship between these is not given a clear process description. However, the suggestion that these skills have a hierarchical structure that is goal-driven is a large step towards such a description.

Fitts' work did not lead to a large increase of research into cognitive skills. Although there was certainly a growing interest in cognitive processes during this time, research was dominated by the advent of the information processing approach. Most researchers were interested in performance questions, where performance was examined in a limited context. As a result the field was soon well-supplied with very specific theories concerned with isolated and disparate processes (e.g., Chase, 1973). There did not appear to be much interest in higher cognitive processes, that is, processes responsible for changes in performance.

Eventually, though, cognitive scientists began to be interested in more unified theories of cognition, attempting to describe cognitive architectures that could
account for a wide variety of phenomena (e.g., Newell & Simon, 1972; Minsky, 1975). It was in this context that Anderson proposed the precursors to the ACT* theory: ACTE (Anderson, 1976) - a successor to the HAM model of memory (Anderson & Bower, 1973) - and ACTF (Anderson, Kline & Beasley, 1979, 1980). The ACT* theory (Anderson, 1982, 1983) is first and foremost a general theory of cognition that describes an architecture which underlies all cognitive processes. The ACT* theory, therefore, qualifies as a candidate for a unified theory of cognition, as advocated by Newell (e.g., 1989, pp 404-405). It is from this general theory of cognition that the ACT* theory of skill acquisition emerges. Features of the cognitive architecture proposed by ACT* which are relevant to skill acquisition will be described below. This will be followed by a description of the processes involved in skill acquisition.

1.4 The Cognitive Architecture Proposed by ACT*

ACT* proposes that the architecture underlying cognition is a production system. Such a system involves the application of production rules which are activated by the contents of a working memory. In this sense production systems can be considered "cognitive S-R theories" (Anderson, 1983, p. 6). Production systems are not peculiar to the ACT* theory and have the property of being computationally universal. That is, they can be used to model all cognitive processes. However, the use of a production system in ACT* is not simply to redescribe behaviour but is sensitive to psychological and empirical constraints. As a result, testable predictions can be made on the basis of this production system, as will be illustrated in the following chapters.

Basic to ACT* is the distinction between declarative and procedural knowledge. Declarative knowledge can be considered to be the representation
of facts (e.g., "A red traffic light is a signal to stop."). Procedural knowledge is basically the representation of what to do in particular situations. Thus actions are contingent on certain conditions being present. In ACT* procedural knowledge is embodied as production rules - condition-action pairs which associate the presence of a particular data pattern in working memory (condition) with the performance of a certain action. Therefore when the condition of a production rule is satisfied, the production can apply and the action will follow (e.g., "IF traffic light is red, THEN stop.").

Particular features of the ACT* production system constrain the activation of productions. These are as follows: (1) Productions can only be activated by knowledge that is currently active in working memory. (2) The speed at which the condition of a production is matched to data in working memory is a function of the strength of the production. In ACT* productions gain strength with successful application.

In addition to the above constraints on production activation, ACT* includes three rules of conflict resolution. These rules determine which production will apply when the conditions of more than one production are matched by the data in working memory. (1) Refractoriness prevents the same production from applying to the same data in the same way more than once. As a result, the production cannot repeat itself over again. (2) When two or more productions can apply, the production with the more specific condition will apply. (3) Specificity and strength interact in the application of productions. If a production with a general condition is selected and applies before a more specific production can apply, then the general production will be the one that actually applies. Therefore more specific productions can only take precedence if they have sufficient strength to ensure faster selection and
application times than more general productions. This strength comes only with successful application (i.e., practice).

Individual productions do not usually function in a vacuum. Instead sets of productions are organised with hierarchical goal-structures. This organisation accounts for the hierarchical nature of human behaviour identified by Fitts (1964). Productions that underly the performance of a particular behaviour are organised around the satisfaction of goals and sub-goals. This results in serial processing where only one goal can be attended to at a time (Anderson, 1983, p. 33). This goal-driven structure also has the function of biasing pattern-matching processes towards matching structures involving the current goal. Considering the specificity constraint described above, this means that productions that refer to the current goal are more likely to apply and apply more rapidly than productions that do not refer to this goal.

The above description of the production system underlying the ACT* theory of cognition is only a brief one, touching upon those features that are important for understanding the ACT* theory of skill acquisition. A more complete account is available in Anderson (1983).

1.5 The ACT* Theory of Skill Acquisition

Singley & Anderson, 1985, 1989), calculus (Singley & Anderson, 1989),
Anderson, Kline & Beasley, 1981), schemata (Anderson, Kline & Beasley,
1979), and creating computer simulations of the processes underlying skill
development in each of these areas. The computer simulations relied on
similar architectures to that specified by ACT* and so provided a useful
method of assessing the adequacy of the ACT* account of skill acquisition.

In all of the work cited above involving Anderson, the basic theory of skill
acquisition has remained consistent, varying little from the first appearance of
the ACT* theory. However some details have been modified recently and
these will be described below. The general sequence of events in skill
acquisition is suggested to be as follows: Knowledge relevant to the
performance of a skill begins in declarative form. This knowledge is
interpreted by general productions called weak problem-solving methods.
These methods are termed "weak" because they are domain-general, that is,
their operation is not specific to any particular type of task (e.g., analogy).
Domain-specific productions are created by a process called compilation. This
process involves two sub-processes. The first is procedularisation which
describes the creation of domain-specific productions as a by-product of the
interpretation of declarative knowledge via weak problem-solving methods.
These new productions perform the goal behaviour without the need to
consult declarative knowledge. Composition is the second compilation
process, and describes the formation of efficient productions by collapsing
sequences of productions into single productions which have the effect of the
series. The likelihood of a production being applied in a particular situation,
and the speed at which the production will be executed, are both functions of
the production's strength. Productions accumulate strength depending on
their history of success. Stronger productions are matched and applied faster.
Therefore highly practised productions are executed faster than newly formed productions. All of the above processes will now be described in more detail. Following this description will be a discussion of how well ACT* accounts for the most common observations of novices gaining expertise at a task with practice.

The processes underlying skill acquisition will be illustrated with respect to a student learning to solve algebra problems. The strategy described is not necessarily how people solve such problems but is useful in illustrating the changes that can occur with practice.

Imagine that a teacher is describing an algebra solution method to a student. The teacher may start with a problem like $79 = 3x + 4$ and tell the student that the goal is to solve for $x$. To achieve this goal requires achieving a number of sub-goals. For example, the teacher may tell the student that the first step in realising the overall goal is to isolate the $x$ term on the right-hand side of the equation. This will mean eliminating the '4' from this side of the equation. The teacher will then demonstrate how this is done, by adding '-4' to both sides of the equation:

\[
79 + (-4) = 3x + 4 + (-4)
\]
\[
=> 75 = 3x
\]

Having achieved this sub-goal of isolating the $x$ term on the right-hand side of the equation, the teacher may then describe the second sub-goal, which is to eliminate the coefficient of the $x$ term, which is 3. This is done by dividing both sides of the equation by 3:
\[
\frac{75}{3} = \frac{3x}{3} = 25 = x
\]

This is then the solution to the problem.

The student's memory of these instructions for how to solve the problem can be considered declarative knowledge. It represents knowledge about how to solve that particular problem, but cannot be used alone to solve other problems. For this knowledge to be useful in solving other problems requires processes that can interpret this knowledge and translate it into action. In ACT* these processes are weak problem solving methods that can be useful in a wide range of domains. According to Anderson (e.g., Singley & Anderson, 1989), humans develop these at an early age and by adulthood these methods are well-developed. These problem-solving methods include analogy, means-end analysis, hill climbing, and pure forward search.

In the algebra example, if the student was presented with another problem to solve, such as \(85 = 4x + 5\), analogy would be the most likely method to apply. Analogy would function to enable the student to mimic the previous solution. This process will only apply, of course, if the student notices the usefulness of the previous solution (e.g., Gick & Holyoak, 1983; Holyoak, 1985). The process can be illustrated with the following plausible imitation of a talk-aloud protocol (for real examples of such protocols, see Anderson, 1983).

"This problem \((85 = 4x + 5)\) looks similar to the teacher's example, so maybe if I try the method that the teacher described I'll solve the problem. The
teacher started by isolating the 'x' term on the right-hand side of the equation by adding the negative of the left-over term (4) to both sides of the equation. In this new problem the left-over term is 5, so I should add -5 to both sides of the equation:

\[
85 + (-5) = 4x + 5 + (-5) \\
=> 80 = 4x
\]

"After doing this, the teacher eliminated the number in front of the 'x' term by dividing both sides of the equation by that number (3). In this new problem the number in front of the 'x' term is 4 so I should divide both sides of the equation by 4:

\[
\frac{80}{4} = \frac{4x}{4} \\
=> 20 = x
\]

So, the solution must be 20."

Although this is a purely fictitious protocol, it captures the essence of protocols reported by Anderson (1983). This example was designed to illustrate how general interpretive methods can be used to translate declarative knowledge into action, given the limitations of declarative memory (i.e., that the student can remember all that the teacher did and why) and that a previous solution will be noticed as useful (see Singley & Anderson, 1989, p. 34).

A by-product of the application of weak problem-solving methods to interpret declarative knowledge and achieve a solution is the formation of new
productions. In contrast to the domain-general weak methods, these new productions are domain-specific. That is, their application is peculiar to the domain of the problem, operating only on particular features of that domain. This development of productions from the application of general methods is called compilation and, as indicated earlier, involves two processes: proceduralisation and composition.

Proceduralisation eliminates the reference to declarative knowledge by building into productions the effect of that reference. In the algebra example, proceduralisation of the problem solution would mean that the student no longer needs to refer to the memory of the teacher's instructions to solve further problems. Instead the student will have developed a set of productions that will solve such problems directly. These productions are domain-specific - they operate only in the domain of solving such algebra problems. This contrasts with the weak methods which will apply in a large variety of domains. An example of a set of domain-specific productions for solving the algebra problems is presented in Table 1.1.

The development of productions such as those in Table 1.1 precludes the need to hold declarative information (i.e., the teacher's instructions) in working memory and use analogy at each step of the problem. Thus to solve new algebra problems, the student does not need to continually refer back to previous solutions for directions. This prediction is supported by the drop-out of verbal rehearsal of problem-solving steps that characterises early performance on this type of task (e.g., Anderson, 1983). As the need to refer to declarative knowledge is reduced with proceduralisation, so should the load on working memory be similarly reduced. This prediction is also consistent with observation (e.g., Woltz, 1988).
P1  IF  goal is to solve for x in equation of form $a = bx + c$
    THEN  set as sub-goal to isolate x on RHS of equation

P2  IF  goal is to isolate x on RHS of equation
    THEN  set as sub-goals
           to eliminate c from RHS of equation
           and then to eliminate b from RHS of equation

P3  IF  goal is to eliminate c from RHS of equation
    THEN  add $-c$ to both sides of equation

P4  IF  goal is to eliminate b from RHS of equation
    THEN  divide both sides of equation by b

P5  IF  goal is to solve for x in equation
      and x has been isolated on RHS of equation
    THEN  LHS of equation is solution for x

Table 1.1: Example of a set of domain-specific productions for solving algebra problems.
An important feature to note of the set of productions in Table 1.1 is that they have a hierarchical structure that matches the goal structure implicit in the solution of such problems. This is fundamental to the ACT* description of skills and skill acquisition and underlies the form of the production sets that develop. The second process involved in compilation - composition - is determined by the goal structure of problems. Composition collapses several productions into a single production. These productions must occur in a sequence and share the same overall goal. The new single production does the work of the sequence but in less steps. For example, productions P2, P3 and P4 in Table 1.1 would be composed to:

P6 IF goal is to isolate x on RHS of equation
THEN add -c to both sides of equation
and then divide both sides of equation by b

With further practice, productions P1, P6 and P5 would be composed to:

P7 IF goal is to solve for x in equation of form
a = bx + c
THEN add -c to both sides of equation
then divide both sides of equation by b
and result is solution

An algebra 'expert' (i.e., someone with many years of experience solving such problems) should be able to recognise this solution immediately upon observation of the problem. The expert would be unlikely to consider the intermediate steps that the novice needs to perform.
Compilation predicts a speed-up in performance for a number of reasons. Firstly, Anderson (1982) suggests that the time to perform a task is a function of the number of steps involved. Therefore, since composition reduces the number of steps (productions) required to perform a task, with practice performance will rely on fewer productions and so will take less time. A more significant reduction in performance time comes with proceduralisation. Performing a task on the basis of a set of productions that execute the task directly should take a lot less time than having to interpret declarative knowledge for procedural directions. This accounts for the dramatic improvements in performance time observed by Singley and Anderson (1989) after only one trial of learning. Singley and Anderson report that subjects showed a 50\% improvement in the time to produce a LISP function from the first trial to the second trial.

The combined speed-up in performance predicted by composition and proceduralisation is not sufficient to account for all the speed-up that is observed in skill acquisition (Anderson, 1982). ACT* includes a tuning mechanism - strengthening - that results in further improvements in performance time with practice. Productions are strengthened with successful practice. That is, each time a production is applied successfully it gains strength. Conversely, if a production is applied inappropriately it loses strength. In ACT* computer simulations of skill development, strength gain is additive whereas strength loss is multiplicative, which results in negative reinforcement having greater effect than positive reinforcement (Anderson, 1982). The stronger a production is, the faster it is to apply. So the combination of compilation and strengthening predicts a speed-up in performance that continues with practice.
Anderson (1982) demonstrates how the compilation and strengthening mechanisms account for the power-law of learning (e.g., Newell & Rosenbloom, 1981). This law describes the observation that when the logarithm of performance time is plotted against the logarithm of amount of practice, a straight line will be observed. Thus learning functions are generally best described by power functions. The basic ACT* version of such a power function is:

\[ T = N \cdot P^c \]

where

- \( T \) = time on trial 1, related to original No. of productions
- \( N \) = amount of practice
- \( P \) = rate of learning, \( c < 0 \)
- \( c \) = \( f + g \)
- \( f \) = the fraction by which the # of steps is reduced by composition.
- \( g \) = related to memory decay

The ACT* account of the power-law of learning will be examined in Chapter 4.

Early versions of the ACT* theory of skill acquisition (e.g., Anderson, 1982, 1983) included additional tuning mechanisms to strengthening - generalisation and discrimination. These mechanisms were suggested to be automatic induction processes which refined the productions developed by compilation. Generalisation is a process whereby general productions are generated from more specific ones. An example from Anderson (1987, p. 205) illustrates this
process with respect to language acquisition. If a child has developed the following two productions:

*IF* the goal is to generate the present tense of KICK

*THEN* say KICK + s

*IF* the goal is to generate the present tense of HUG

*THEN* say HUG + s

the generalisation mechanism would develop a more general rule that would be applicable in the above situations and others:

*IF* the goal is to generate the present tense of 'verb'

*THEN* say 'verb' + s

Discrimination has the effect of restricting the range of such general rules. Thus, in the above example, the general rule is overly general, and in certain circumstances is not appropriate. As a result the discrimination mechanism would generate new rules appropriate for these circumstances. For example,

*IF* the goal is to generate the present tense of 'verb'

and the subject of the sentence is singular

*THEN* say 'verb' + s

*IF* the goal is to generate the present tense of 'verb'

and the subject of the sentence is plural

*THEN* say 'verb'
Despite the utility of generalisation and discrimination in accounting for various phenomena in language acquisition (e.g., Anderson, 1983), the more recent versions of ACT* (Anderson, 1986, 1987, 1989a; Singley & Anderson, 1989) have suggested that these two tuning mechanisms are unnecessary in a theory of skill acquisition. In fact, the effects of these mechanisms can be implemented by the same general problem-solving methods that initiate proceduralisation (e.g., analogy) (Anderson, 1987). In addition, by having generalisation and discrimination implemented by these general methods, the induction of more refined rules is more sensitive to semantic and strategic factors than the automatic processes of the generalisation and discrimination mechanisms (Anderson, 1987).

In addition to the observations of cognitive skill acquisition that were mentioned above, the ACT* theory of skill acquisition provides useful accounts of phenomena in two other major areas of skills research. The first is the three phases of skill development described by Fitts (1964). The second is the growing field of research into expert/novice differences.

1.6 ACT* and Fitts' Phases of Skill Acquisition

The early phase of skill acquisition identified by Fitts (1964) corresponds in ACT* to the application of general problem-solving methods to declarative knowledge and to the initial development of productions. Fitts suggested that this phase only lasts for a short time, which is consistent with Anderson's reports of one-trial learning (e.g., Singley & Anderson, 1989). Fitts describes this phase as the cognitive stage, where most of the thinking about a task is performed. Anderson (1983) claims that it is natural to equate this stage with the interpretive application of knowledge. Certainly there has been considerable evidence recently that higher order cognitive activities, such as
comprehension of task requirements and planning, are more prevalent early in skill development than later (e.g., Ackerman, 1988; Woltz, 1988). In addition, processing during this stage is more error-prone and deliberate than in subsequent stages (Ackerman, 1988; Woltz, 1988) as working memory resources are stretched by the interpretation of declarative knowledge.

The intermediate phase identified by Fitts describes the formation of specific associations between stimulus cues and appropriate responses. The similarities between such associations and production rules are obvious. In ACT* this stage corresponds to the drop-out of verbal rehearsal of instructions and the associated reduction of working memory load.

The end phase was described by Fitts and Posner (1967) as the stage where "component processes become increasingly autonomous (p. 14)." During this stage, skills are less reliant on working memory resources and become faster with practice (Fitts & Posner, 1967). The ACT* theory suggests that productions gain strength with practice and this results in faster application of the productions. In accordance with the power-law of learning, the effect of strengthening on improvement becomes increasingly small with practice. Eventually, after many thousands of trials (Anderson, 1989b), no further improvement will be observed. At this point performance may appear to be automatic (e.g., Shiffrin & Schneider, 1977): when the conditions of a production are satisfied, the action will follow automatically.

1.7 The Acquisition of Expertise

The ACT* interpretation of the three stages of skill acquisition is also a feature of the ACT* account of expert/novice differences. ACT* predicts a number of performance characteristics of people with considerable experience
in a particular domain (experts) compared to those with little or no experience (novices). Two such characteristics will be discussed below:

1. Experts possess compiled productions that control behaviour in their domain of expertise, whereas novices do not.

2. Experts' domain-specific productions are organised around efficient strategies which are driven by a hierarchical goal-structure. Novices perform on the basis of domain-general productions and so do not possess efficient strategies.

1.7.1 Experts Possess Compiled Productions

The ACT* theory of skill acquisition states that consistent practice of a task will result in the compilation of domain-specific productions. Therefore an expert in a particular domain should possess productions for efficiently performing a task in that domain. A novice to the domain should not possess such productions.

One area of research that illustrates this distinction between experts and novices indicates that experts virtually 'see' a different problem to novices. For example, Chi, Feltovich and Glaser (1981) asked expert and novice physicists to sort physics problems on the basis of similar solution methods. The experts sorted the problems according to the physics principles that were involved in the problems (e.g., Newton's Second Law). In contrast, the novices sorted the problems according to similarities in peripheral information or key words specified in the problem statement, such as ramps or pulleys. Further evidence of this nature was obtained by Schoenfeld and Herrmann (1982), who controlled for the possible confounds of age and ability in the
Chi et al. study. Students' perceptions of the structure of mathematical problems were examined before and after a month-long intensive course on mathematical problem-solving. Initially these subjects perceived maths problems on the basis of surface structure (i.e., words or objects described in the problem statement). After the course, the students perceived problem relatedness more like a group of experts who were also studied, that is, according to principles or methods relevant for problem solution.

A slightly different source of similar evidence is the study of chess masters. Chase and Simon (1973) asked their subjects to recall the positions of chess pieces from mid-game configurations. All subjects tended to recall pieces in clusters. However, the clusters recalled by chess masters typically consisted of pieces which formed attack or defense configurations. The clusters recalled by novice chess players were more likely to be related to the proximity of pieces on the board, and therefore unrelated to the individual functions of the pieces.

These three studies suggest that experts 'see' problems in their area of expertise in a different way to novices. This is as would be expected if experts perform on the basis of compiled productions. From experience, experts will have learned that certain characteristics of problems are important with respect to deriving their solutions, and others are not. In physics and maths problems the underlying principles are much more important than surface information. For instance, not all pulley problems involve the same physics principle. Therefore the productions developed by experts to solve such problems should be principle-based. In other words, the conditions of these productions will be based on 'principle' features of a problem. As a result, it is the principle underlying a problem, not surface information in the problem description, that will activate an expert's productions. This is likely
to result in a processing bias in experts towards the principle features of a problem, and away from the surface information.

In contrast to experts, novices should have no domain-specific productions. They will rely only on their declarative knowledge of the problems, if they possess any, and weak problem-solving methods. This suggests why novices are particularly affected by surface information in problems. The probability of reasoning by analogy, a weak problem-solving method, has been shown to be affected by the surface similarities between problems (e.g., Holyoak & Koh, 1987), especially in novice performance (Novick, 1988). This is not to say that analogy is necessarily an inefficient form of reasoning. Surface similarities between problems usually indicate important information. It is only when such similarities are not useful that analogy causes errors, and novices appear to be less well-equipped to notice this type of situation than experts.

If experts' strategies in such problems are indeed principle-based, then experts might be expected to approach problems in their domain of expertise in different ways to novices, especially if the principle underlying a problem is not immediately obvious. Indeed, this has been shown to be the case. More research on physics expertise (Chi, Glaser & Rees, 1982; Larkin, McDermott, Simon & Simon, 1980; McCloskey, Caramazza & Green, 1980) indicates that experts spend a lot of their problem-solving time constructing a representation of the problem in terms of basic physics principles. Only when these representations, either mental or physical (i.e., drawings), have been constructed do they embark on a solution path. Novices, however, quickly start trying to retrieve an equation that will take them from the quantities given to the quantities needed to answer the problem. Therefore surface features of the problem tend to invoke equations for novices and they begin on the
algebra quickly. In contrast, experts concentrate first on understanding the problem.

Interestingly, then, it seems that through experience, the compilation process has developed in experts productions that are principle-based. This has resulted in strategies that to some extent involve a 'principle-search.' That is, experts automatically approach problems by trying to understand the underlying principle.

Having compiled productions also accounts for why most errors in the acquisition of skill are made, not by novices or experts, but by trainees with an intermediate amount of experience (Anderson, 1982). For example, Lesgold (1988) reported such an observation in a study of radiologists making diagnoses on the basis of X-Ray images. There were some cases where third and fourth year residents performed worse than new residents, who performed almost as well as senior staff. In fact, there were some X-Ray images where a resident had made the correct diagnosis in their first year but made an incorrect diagnosis in their third year. Lesgold pointed out that these particular images depicted classic examples of particular diseases, similar to those that new residents may have studied in text-books. However, Lesgold also suggested that the obvious features that indicate one disease may also be consistent with other diseases. When faced with such films, new residents made the obvious choice as they seemed unable to assess less likely alternatives. After further training though, these residents were aware of other potential diagnoses but were not capable of reliably choosing between them. Thus in some cases less likely alternatives were chosen because not enough was known to rule them out. This appears to be a phenomenon that characterises medical education in general (Lesgold, 1988). On the basis of a small set of symptoms, intermediate-level trainees are likely to entertain a
greater variety of diagnoses than either novices or experts and therefore are more likely to make the wrong choice.

The distinction between experts and novices and the middle-level medical trainees can be considered in ACT* terms. Beginners are trained with classic cases and so develop productions that are executed when the features of classic cases are perceived. Middle-level trainees develop more productions through encountering other cases. This may lead to competition between productions in terms of which will be executed if the conditions of a number of productions match the features of a case. Thus inappropriate productions may be executed sometimes. The additional experience of experts results in productions with more specific conditions. This enables experts to make finer discriminations between symptoms or features of X-Ray images.

There is also evidence that experts do not perform as well as novices on some tasks and that this may be a result of experts' compiled productions. Myles-Worsley, Johnston and Simons (1988) presented faces and X-Ray images to observers of four different levels of radiology expertise. Half of the X-Ray images featured clinically significant abnormalities, whereas the other half did not. In a recognition memory test of these faces and X-Ray images three basic results were observed: (1) Recognition memory for faces was high across all levels of expertise. (2) Memory for abnormal X-Rays improved with expertise to the point where it was equal to memory for faces in the most experienced radiologists. (3) Memory for normal X-Rays declined with expertise, from above chance in the novices to below chance level in the most experienced radiologists. Myles-Worsely et al. interpreted these results as suggesting that developing radiological expertise is associated with increased selective processing of features relevant to clinical abnormalities. Expert radiologists appear to process X-Ray images in a similar way to the
processing of faces, that is, by quickly focussing on features that distinguish one image from another. However, this selective processing appears to be characteristic only of X-Ray images that contain clinically abnormal features. Radiologists apparently lose the ability to notice variations in normal features of X-Ray images as they develop the ability to notice abnormal features.

An ACT* interpretation of Myles-Worsely et al.'s results, which follows their own interpretation, is that expert radiologists have developed two basic types of productions. One set enables the experts to quickly assess clinically normal features of X-Ray images. The second set processes those features that are characteristic of abnormalities and results in swift detection of such features. Together these two types of productions can account for the basic results reported by Myles-Worsley et al. The experts focus most of their attention on detecting abnormalities in X-Ray images. As a result they are more likely to recognise features that distinguish particular abnormal images from others. However, because most of the expert's attention is focussed on detecting clinically relevant abnormal features, little attention is paid to detecting irrelevant abnormalities that less-experienced observers may use as recognition cues. Reflecting on this interpretation, Myles-Worsley et al. (1988, p. 557) make the interesting point that "expertise in a particular domain is likely to be a two-edged sword: It can bias perception toward some classes of stimuli in that domain and away from others." This impression will be reiterated in the experiments reported in Part 1 of this thesis.

In summary, the studies just described are consistent with experts operating with compiled productions. These are efficient productions that are executed when certain stimulus conditions are present. It was shown that this creates a processing bias in some experts in some areas: experts 'see' a different problem to novices. It was also shown that compiled productions lead to
experts exhibiting poorer performance in some medical situations than novices. A similar result has been observed in computer programming (Adelson, 1984).

1.7.2 Experts Use Hierarchical Strategies

As was described in section 1.5, ACT* states that productions develop with a structure that matches the hierarchical nature of tasks. This structure is organised around the satisfaction of goals and sub-goals. This suggests that experts should perform with more efficient strategies than novices because this goal-structuring apparatus should be more developed in experts. Some evidence does exist for this strategy distinction.

In the previous section it was shown that experts tend to approach problems in a different way to novices. That is, they begin by constructing representations of the problems based on the principles inherent in them. Simon and Simon (1978) found that once these representations have been built, experts and novices diverge in another way, this time in their method of solution. Experts tend to use a "working forward" strategy where they generate equations from the givens of the problem to the goal. Novices, however, are more likely to use a "working backwards" strategy, where they work backward toward the givens from the goal. New equations are chosen to fill the gaps from previous equations. Once equations allowing a solution are obtained, novices reverse the process and work towards the goal. This strategy corresponds to a means-end analysis, one of the weak problem-solving methods Anderson suggests play an important role in the early stages of skill acquisition. This method requires less specific knowledge than the experts' strategy but requires more information to be retained in working memory (Lesgold, 1984). Thus novices appear to be searching their
knowledge of equations in the hope of finding some that will enable
calculation of particular unknowns. Experts on the other hand spend time
constructing representations of problems based on principles and these
representations automatically reveal solution paths. This corresponds to
satisfying conditions in productions which results in their execution. In turn
this leads to the activation and execution of other productions relating to the
same goal, which is observed as an efficient problem-solving strategy.

Similar strategy differences have been found between experts and novices in
bridge, chess and radiology. Chamess (1989) found that certain card hands
automatically suggested lines of play to bridge experts. Experts were shown
to exhibit sophisticated planning and goal-setting in response to various
features of play. Novices exhibited little planning if any. Similarly, chess
configurations immediately suggest move sequences to chess masters (Chase
& Simon, 1973). In radiology, when experts look at an X-Ray image they
quickly mention a diagnostic category (Lesgold, 1988). This in turn appears
to trigger a plan of attack which is aimed at evaluating the probability of this
diagnosis. The image is then examined in the context of this diagnosis, with
experts identifying more features relevant to it and fewer that are irrelevant to
it than novice radiologists.

In summary, experts appear to organise their knowledge around plans of
attack. These plans typically involve goal-setting and usually have a
hierarchical structure. In contrast, novices tend to have a lot less structure and
sophistication in terms of efficiency associated with their solution methods.

In conclusion, the ACT* predictions of processing changes associated with
skill acquisition appear consistent with representative results of expertise
research. In this sense ACT* provides a realistic account of skill acquisition.
However, this account has only provided a *post hoc* explanation of the phenomenon typical of the acquisition of expertise. A great deal of this account was guided by the phenomenon, and so does not provide a predictive account. If ACT* is to provide a useful account of skill acquisition it must be able to predict changes in behaviour that result from practice. This issue forms the basis of the following chapter, where the ACT* account of transfer of training is examined with respect to the contextual interference effect. Predictions based on the ACT* account are contrasted with those from an alternate account of this effect.
PART 1 Transfer of Training and the Contextual Interference Effect

Chapter 2 ACT* and intratrial processing accounts of the contextual interference effect

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 ACT* and Transfer of Training</td>
<td>37</td>
</tr>
<tr>
<td>2.1.1 Positive Transfer</td>
<td>39</td>
</tr>
<tr>
<td>2.1.2 Negative Transfer</td>
<td>41</td>
</tr>
<tr>
<td>2.1.3 Zero Transfer</td>
<td>42</td>
</tr>
<tr>
<td>2.2 The Contextual Interference Effect</td>
<td>44</td>
</tr>
<tr>
<td>2.3 Overview of Experiments: Aims and Predictions</td>
<td>52</td>
</tr>
</tbody>
</table>
2.1 ACT* and Transfer of Training

Transfer of training research has typically had an applied focus. The implicit motive in this research is usually to demonstrate that training in one situation will improve performance in another situation, with experimental manipulation aimed at either increasing or decreasing the extent of this transfer. An obvious example is research that looks at the relationship between training in a flight-simulator and subsequent performance in a real aircraft. Historically, the challenge in such research has been to investigate what features in the training and transfer situations determine the amount of transfer that occurs. In other words, what is transferred from one situation to another that can either benefit or impede performance in the second situation?

The most influential theory of transfer has been the identical elements theory of Thorndike (Thorndike, 1906; Thorndike & Woodworth, 1901). This theory states that transfer between two tasks is determined by the extent to which the tasks share the same content. Thus the more that is common between tasks in terms of stimuli, responses or stimulus-response pairs, the greater will be the transfer. The theory has enjoyed wide-spread support in both verbal learning (e.g., McGeoch, 1952; Osgood, 1949) and motor learning (e.g., Adams, 1987; Duncan, 1958) research.

The identical elements theory was developed in the early days of behaviourism and so was couched in stimulus-response terms rather than cognitive terms. As a result criticism was levelled at the theory's inability to account for transfer that apparently was not related to commonalities between tasks but was associated with more cognitive features of the tasks (e.g., Orata, 1928). A second, more telling, criticism of the identical elements theory concerned the assumption that stimulus-response pairs are the basis of
transfer. This assumption was criticised as being so restrictive as to rule out transfer altogether (Meiklejohn, 1908). Unless a new task involved the same responses to the same stimuli as an original training task, there would be no transfer between the tasks. Thus learning to drive a red car would be of no benefit for driving a blue car. Therefore stimulus-response pairs were not sufficiently abstract to be identified as the identical elements of Thorndike's theory.

Little was achieved in the years that followed Thorndike's work to provide a suitable representation for skill. Research into transfer was more concerned with the effect of various training conditions on transfer performance. However Briggs (1969) suggested that analysing the relationship between common features of tasks and the extent of transfer between them could only provide a preliminary understanding of transfer and that it was "important to determine what is learned...during training...for a more complete understanding " (p. 217). Hence the progress of skills research was restricted by the lack of suitable concepts for describing the "what" of learning.

The cognitive representation of knowledge became a popular topic of research and theory in the 1960's resulting in many suggested modes of representation, such as semantic networks (Collins & Quillian, 1969, 1972), productions (Newell & Simon, 1972), schemas (Minsky, 1975), and, more recently, mental models (Johnson-Laird, 1983). As a result, theoretical tools not available to Thorndike were developed, enabling more abstract discussions of the processes underlying skills and transfer.

Anderson (1987; Singley & Anderson, 1989) has recently resurrected the identical elements theory by identifying production rules as the elements of knowledge underlying transfer. This proposal is now an explicit and
fundamental feature of the ACT* theory of skill acquisition, although originally it was only implicit. According to this proposal transfer between two skills is determined by the extent to which productions underlying performance in one skill are useful in performing the second. The greater the production overlap, the greater the transfer.

Productions have an immediate advantage over stimulus-response pairs with respect to representing the processing operations underlying skilled performance: productions are abstract cognitive representations. As described in the ACT* theory, productions are formed as the by-product of an interpretive process which compares two declarative representations. For example, analogy compares the representation of a previous solution to the representation of the current situation and extracts common features. The productions that result from this process are generalisations and, therefore, are necessarily abstract (Singley & Anderson, 1989).

Three forms of transfer are traditionally considered in any discussion of transfer results: positive, negative, and zero transfer. The identification of identical productions as the basis for transfer leads to strong predictions concerning each of these situations. These will be described below, along with studies which examine these predictions.

2.1.1 Positive Transfer

As a task is practised, a set of productions is developed which underlies the performance of this task. If the performance of a second task can utilise these same productions then positive transfer will result between the two tasks. Positive transfer in this situation refers to the fact that knowledge developed in one situation is transferred to another situation. This has a positive effect in
the sense that there is no need to develop productions from scratch to perform operations that existing productions already perform. Therefore the extent of this positive transfer is determined by the number of productions developed in the context of the first task that can be used in performance of the second task.

Singley and Anderson (1985, 1989) were able to provide support for the ACT* account of transfer in a detailed study of text-editing skill. Two basic types of editors were examined: line editors which allowed examination of only one line of text at a time, and screen editors which enabled viewing of whole screens of text. Singley and Anderson predicted that there would be almost complete positive transfer between two line editors but only partial transfer from line editors to screen editors. The transfer predictions were based on models of the productions underlying performance with each of the editors. Despite the fact that the two line editors shared few commands, Singley and Anderson identified considerable production overlap in these editors, mainly concerned with the more abstract planning operations. This was shown to result in almost complete transfer between the two line editors: two days of practice with one of these editors was almost equivalent to two days of practice with the other in terms of preparation for further performance with this second editor. Much less production overlap was identified between the line editors and the screen editor and this was shown to result in only partial transfer from training with the line editors to performance with the screen editor.

Kieras and Bovair (1986) have also reported success at predicting the extent of transfer between tasks on the basis of the number of shared productions. Detailed predictions of the training time to reach a performance criterion were shown to be accurate when the task was analysed in terms of old and new
productions. Old productions had been learned in the context of another task and so did not require learning. New productions were those that needed to be developed. Kieras and Bovair predicted and observed that the greater the number of old productions that underlied performance of a new task, the smaller the training time required on this task.

2.1.2 Negative Transfer

Negative transfer refers to a situation where performance of a task is worse than if a preceding task had never been performed. One interpretation of such a result is that the procedure for performing the second task is performed less effectively as a result of another procedure having been learned in the context of the first task (Anderson, 1987). However, the ACT* account of transfer predicts that negative transfer in this sense does not exist (Singley & Anderson, 1989). The worst case that can be expected in a transfer situation is that two tasks share no productions, resulting in zero transfer (this situation will be described below). ACT* proposes that results which might indicate negative transfer in fact indicate positive transfer. However, what has been transferred are non-optimal methods for performing the task. Therefore productions developed in the context of one task are transferred to the performance of another task. These productions do not interfere with the execution of productions relating to the second task (i.e., negative transfer). Instead, when these productions are used to perform the second task, performance is less efficient than if it relied on productions that were developed in the context of the second task only. This situation will occur when the stimulus conditions of the second task match the conditions of productions developed in the context of the first task. These productions will execute in response to these stimulus conditions, although the processing strategy they embody will not be optimal for the second task.
Some evidence consistent with the ACT* account of negative transfer exists, the most famous being the demonstration of the *Einstellung* phenomenon by Luchins (1942). Subjects were required to solve a number of problems which involved determining the best method of combining three water jugs (A, B & C) to make up designated volumes of water. The problem sequence was such that subjects were presented with five problems that each had the same solution (i.e., combine the jugs according to B - A - 2C). Following these problems, the subjects were presented with another two problems that could either be solved with this same method or with a more direct method (A + C). Luchins found that almost all of the subjects solved the two transfer problems with the method discovered during training. In contrast, control subjects who were first presented with the transfer problems almost always solved them with the easier method. Therefore the training problems encouraged the use of a non-optimal solution procedure with the transfer problems. The stimulus conditions of the transfer problems were obviously sufficiently similar to those of the training problems to trigger the execution of this non-optimal method. This meant that subjects could utilise a previous solution rather than develop a new strategy specific to the transfer problems. Further evidence that negative transfer is in fact the positive transfer of non-optimal methods has been reported with text-editing (Singley & Anderson, 1989) and computer programming (Kessler & Anderson, 1986).

2.1.3 Zero Transfer

The ACT* account of transfer predicts that training with one task will provide no benefit for performance with a second task if the two tasks do not share common productions. More specifically, transfer between tasks is restricted to common productions and is not related to the abstract knowledge underlying the productions. Therefore there will be no transfer between skills
that use the same knowledge in different ways. This proposal leads to some counter-intuitive predictions. For example, there should be no transfer between the comprehension and generation of language. Singley and Anderson (1989) reviewed the small amount of research that reflects upon this issue and concluded that there was some evidence that comprehension and generation involved separate systems. In addition, Singley and Anderson reported an experiment that examined the training of subjects to solve calculus problems. Results of this experiment suggested that translating written problems into equations was completely independent of the solution of these equations. These processes shared common abstract knowledge about the function of calculus operators but used it in different ways. Similar evidence was provided by McKendree and Anderson (1987), who found that there was little transfer between evaluating and generating LISP code. There also appeared to be little transfer between evaluating certain combinations of LISP commands when the same components were presented in one combination during training and another during transfer. Thus although the same primitive knowledge was used in both cases, apparently the productions developed in one situation were not appropriate for the other situation. These studies demonstrate that the acquisition of skills is specific to their use.

In conclusion, by specifying productions as the elements underlying transfer, the ACT* account of transfer provides an advance over Thorndike's identical elements theory. Productions appear to be more appropriate representations of procedural knowledge than stimulus-response pairs. In addition, there is some evidence that transfer between tasks can be predicted on the basis of identical productions underlying performance of the tasks. In the next section, an interesting transfer phenomenon - the contextual interference effect - will be examined in light of the ACT* account of transfer and an alternative
account of this effect. The contrast between the two accounts forms the basis for the experiments reported in the following chapter.

2.2 The Contextual Interference Effect

The contextual interference effect has been investigated mainly in relation to the training of motor tasks (e.g., Shea & Morgan, 1979). However this effect was originally noted in verbal learning research. Battig (1966) reported that the difficulty of learning paired associates was increased by learning the list of pairs in the context of other already well-learned pairs. However, this intra-task interference led to maximal facilitation in learning a related second list of pairs. Battig (1972) accounted for this result by suggesting that high contextual interference during training leads to multiple processing strategies and this increases the probability of transfer to other situations compared to training with low contextual interference. This hypothesis suggests that transfer is determined by the extent to which tasks require similar processing strategies.

Explanations of the contextual interference effect have been extended considerably in relation to the learning of motor skills. Shea and Morgan (1979) had subjects learn three motor tasks under a blocked (low contextual interference) or a random (high contextual interference) sequence of presentation. The results showed that subjects in the random condition performed slower during training than subjects in the blocked condition but were faster during transfer to other similar tasks. The transfer difference between the two groups increased as the complexity of the transfer tasks increased. Shea and Morgan interpreted these results as supporting Battig’s (1972) theory of contextual interference. The random sequence of presentation increased contextual interference between the three tasks and this
led to multiple processing strategies during training. "Blocked" subjects did not need to develop multiple strategies and so were not as well equipped to handle new tasks as were the random subjects. Similar results and interpretations have been reported by Del Ray, Wughalter and Whitehurst (1982), Lee and Magill (1983), and Shea and Zimny (1983).

Shea and Zimny (1983), in further articulating the contextual interference effect, emphasise that a motor act is represented in memory in terms of the operations used by the subject during performance, as opposed to a representation of the sensory attributes of the task (including the context). In addition, these operations can be more applicable to performance in some tasks, than in others. The implication then is that transfer will not be determined by sensory aspects of tasks, but will be influenced by the operations that are induced by the context of the performance situation. Thus transfer between two tasks will occur to the extent that the two tasks require common processing operations. In this way the Shea and Zimny (1983) account of transfer in relation to the contextual interference effect is quite similar to the ACT* account of transfer. However, where Shea and Zimny specify common processing operations as the basis of transfer, ACT* specifies production rules. In fact, Anderson (1989b) has applied the identical productions explanation of transfer to account for the contextual interference effect.

The ACT* account of the contextual interference effect is as follows: The subjects in the Shea and Morgan (1979) study who were trained with a blocked sequence of presentation were able to compile the productions necessary for performance of the first task in one block of several trials. Following this task the subjects compiled another set of productions for the next task, and so on for the third task. Thus these subjects did not need to
develop productions for identifying the task at hand. The "random" subjects on the other hand would have found it necessary to compile not only productions specific to each of the tasks, but also productions for identifying the particular task they were to perform on any one trial. Only after identification was made could performance proceed. Therefore, in the transfer phase of the experiment, where all tasks were presented in a random order, the "blocked" subjects were not equipped to perform efficiently. That is, they did not have efficient productions available which would function to identify the task they were to perform. As a result, subjects would have experienced difficulty utilising the productions that they had developed. In the more complex transfer tasks both groups of subjects were required to develop more productions to cope with the new tasks, but the random subjects were still faster because they had more appropriate productions to work with (i.e., those relating to task identification). The basic difference between this explanation and that provided by Shea and Morgan, and others, is that where they suggested that random presentation leads to multiple processing strategies, the ACT* explanation identifies productions as the basis of these strategies or operations.

An alternative explanation of the contextual interference effect emphasises the difference between intertrial and intratrial processing (Lee & Magill, 1983; Carlson & Yaure, 1990). Carlson and Yaure (1990) suggest that in explanations like the ACT* account the locus of the contextual interference effect is in the ability to contrast different tasks in order to develop appropriate processing strategies. Random presentation, therefore, enables subjects to contrast different tasks and so develop the skill of distinguishing one type of task from another. On the other hand, presenting similar types of tasks in blocks reduces the ability to contrast the different types of tasks. As a result
"blocked" subjects do not develop the appropriate skills for later random presentation.

Carlson and Yaure (1990) distinguish accounts based on intertrial processing, such as the one provided by the ACT* theory, with an account of the contextual interference effect based on intratrial processing. They suggest that in random practice subjects have to load a solution to the present problem into working memory. On most trials this will not be the same solution as on the previous trial. In contrast, with blocked practice the same solution can be retained in working memory for the duration of the trials in a block. Carlson and Yaure propose that because random training affords subjects more practice at loading solutions, subjects become more fluent at accessing and using the component skills underlying these solutions. The increased efficiency of memory processes is therefore seen as the basis of the contextual interference effect, rather than the development of productions that determine which skill to use.

Some evidence exists that supports the hypothesis that intratrial processing accounts for the contextual interference effect. Lee and Magill (1983) included a serial presentation condition in their study in addition to random and blocked presentation. In this serial condition subjects performed three tasks in a fixed repeating order. Thus there was perfect predictability of task type on each trial, as in the blocked condition, but on each trial a new solution procedure was required, approximating the extent of solution switching in random presentation. The predictability of task type on each trial in the serial condition reduced the need to develop the skill of identifying which task to perform. According to the ACT* account of the contextual interference effect, this should result in subjects being ill-equipped to perform with random presentation during transfer in comparison with subjects trained with random
presentation. However, the fact that different tasks were performed on each trial would mean that new solutions would be loaded into working memory on every trial, thus improving the fluency of memory processes. Therefore, according to the intratrial processing account of the contextual interference effect the serial condition should have a similar effect on training and transfer performance as the random condition. That is, the performance of serial trained subjects will be slower than blocked subjects during training, but this difference will be reversed during transfer when presentation is random. Lee and Magill reported results consistent with this latter prediction, supporting the intratrial processing account.

Further evidence in support of the intratrial processing account was reported by Carlson and Yaure (1990). Subjects performed tasks that intervened between successive trials in a blocked presentation condition. This was shown to result in a similar degree of transfer to a problem-solving task compared to that which was observed from a random training condition. Subjects in these two groups solved the problems faster than subjects who were trained with a standard blocked presentation. Carlson and Yaure interpreted these results as suggesting that the intervening tasks had a similar effect to random presentation because both conditions required new solutions to be loaded into working memory on each trial. Carlson and Yaure proposed this as support for the hypothesis that the efficiency of working memory processes underlies the contextual interference effect.

The Carlson and Yaure (1990) and Lee and Magill (1983) studies do not necessarily disconfirm the ACT* account of the contextual interference effect. As Carlson and Yaure point out, the intratrial processing account can not explain the slower performance during random training compared to blocked training that is characteristic of the contextual interference effect. Initially
though, this account does seem to explain this training difference. If blocked trained subjects do retain the same solution method in working memory for a number of trials, then there should be some speed advantage in performing without the need to load a new solution method on each trial. Random trained subjects, on the other hand, should be slower because they will need to load new solution methods more often, a time-consuming process. However, the intratrial processing account cannot propose this suggestion as an explanation for the training differences as well as claim the results of the intervening-task condition as support. The reason is that the training performance of subjects in the intervening-task condition was as fast as the training performance of subjects in a standard blocked condition, although the transfer performance of the former group was more like random trained subjects. Thus, having to load a new solution method into working memory on each trial in the intervening-task condition was able to provide the transfer advantage typically provided by random training, but it was not able to reduce the advantage during training associated with blocked presentation. Therefore, given that Carlson and Yaure clearly state that the intratrial processing account cannot explain the training differences in the contextual interference effect, the facility to retain a solution method in working memory for a number of trials provided by blocked training is not the cause of the speed advantage associated with this training condition.

The ACT* account of the contextual interference training difference concerns the relationship between the number of steps involved in performing a task and the time to perform that task. Implicit in the ACT* description of the power law of learning is the proposal that performance time is a function of the number of productions underlying performance. As described earlier, ACT* identifies the main difference between random and blocked training as the development of skills concerned with identifying task type: random
training encourages the development of these skills whereas performance during blocked training can proceed without them. The implication of this account is that during training, more productions will underlie the performance of random subjects than the performance of blocked subjects, and so the former group will have longer performance times.

Although Carlson and Yaure did not consider the ACT* explanation of the training differences, they did suggest that both intertrial and intratrial processing accounts may be required to account for all of the contextual interference effect.

The major aim of the experiments reported in the first part of this thesis was to examine further the contextual interference effect and the two different explanations for this effect. The aim was primarily to determine whether one of these explanations is sufficient or that both are required. A number of issues reflect upon this general issue and so determined the types of experiments that were conducted. These are described below:

(1) All of the studies that have reported the contextual interference effect examined the acquisition of motor skills, except the Carlson and Yaure (1990) study and an earlier study involving Carlson (Carlson, Sullivan & Schneider, 1989). The two Carlson studies both examined the acquisition of a cognitive skill - making judgements about digital logic gates. Most importantly, the effect of a serial presentation condition has only been reported in relation to performance on a motor task (Lee & Magill, 1983). Therefore, one of the aims of the present study was to continue the extension of the contextual interference effect to cognitive tasks, and include a similar condition to Lee and Magill's serial presentation condition.
(2) Carlson and Yaure's (1990) study was not designed to compare the intertrial and intratrial processing accounts. The experiments in the present study, however, have this as an explicit aim. This will be achieved by examining whether the ACT* theory can provide as good an account of the contextual interference effect as the intratrial processing account, something that has not been attempted in previous investigations of the effect. In particular, an explicit test will be made of whether training differences associated with the contextual interference effect can be accounted for by the ACT* explanation. This explanation states that differences in performance time in the different training conditions are related to the number of productions that underlie performance in these conditions. Furthermore, transfer differences will be explicitly examined in relation to the transfer of productions.

(3) The third issue is related to the second. In all previous studies of the contextual interference effect the measures used to examine transfer have not been suitable for assessing all of the ACT* predictions. For example, ACT* predicts that blocked trained subjects will perform random transfer trials slower than random trained subjects. This is the basic contextual interference effect. However, ACT* also predicts that the performance of blocked trained subjects during random transfer will rely not only on the productions that were developed during transfer but also on new productions for identifying task type. As a result, training performance will not be a good indicator of transfer performance because it will not include the development of new productions. In contrast, random trained subjects will be able to perform during random transfer solely on the basis of previously developed productions. Thus transfer performance in this condition will simply involve continued improvement of the productions developed during training. Therefore training performance will be a much better indicator of transfer
performance than in the previous condition. The only way to examine this prediction is to compare transfer performance with performance predicted on the basis of training performance. Any differences between observed and predicted performance will indicate whether transfer performance is based on the execution of 'old' productions only or both 'old' and 'new' productions. No previous studies of the contextual interference effect have examined this issue although it is necessary in order to evaluate whether the transfer of identical productions underlies the effect. The present experiments examine transfer with both of the above methods (details of the second method will be described in the next chapter).

2.3 Overview of Experiments: Aims and Predictions

A reasoning task was chosen which had two basic forms (details to be described in Chapter 3). Performance in this task required identifying problem type and then implementing the appropriate solution method. Three experiments were conducted which involved substantial practice on this task during training and a third as many trials during transfer.

The aim of Experiment 1 was to replicate the contextual interference effect with the reasoning task. Subjects were divided into two training groups. One group was presented with the two types of problem in a random order. The other group was presented with a block of trials of one problem type, followed by a block of the other type, then another block of the first type, and so on. Transfer trials were also presented in either a random or a blocked order. The factorial combination of training and transfer condition resulted in four groups of subjects.
The predicted results for Experiment 1 were those that are characteristic of the contextual interference effect: during training, performance with random presentation would be slower than with blocked presentation. Only the ACT* theory can account for this training difference. However, during random transfer, this difference would be reversed: random trained subjects would solve these problems faster than blocked trained subjects. Both the ACT* theory and the intratrial processing account predict this transfer difference.

Experiment 2 was concerned with the effect of a presentation condition similar to Lee and Magill's (1983) serial condition. In this experiment, subjects were presented with the two types of problem in a fixed alternating order throughout training. In the transfer phase the presentation order was random. The ACT* account and the intratrial processing account predict different amounts of transfer from alternating to random presentation. The absolute predictability of problem type that results from alternating presentation means that subjects need not develop the skill of identifying problem type. According to the ACT* theory, if this skill is not developed then performance with random presentation will not be as efficient as observed with alternating presentation. Therefore transfer will be positive but will only be partial. In other words, performance in this condition will be similar to that of subjects trained with blocked presentation. However, the alternating presentation means that new solution methods are loaded into working memory on each trial. According to the intratrial account, this practice at loading solution methods results in efficient memory processes and so performance with alternating presentation will be as efficient as with random presentation. Therefore, transfer will be positive and complete. In other words, performance in this condition will be similar to that of subjects trained with random presentation.
In Experiment 3 the two problem types were presented in a random order during both the training and transfer phases. However, during training, the identification of problem type was made easier by highlighting the feature that this identification was based on. Subjects could then rely on this feature rather than learning to make the discrimination that is necessary when this cue is not present. According to ACT*, if subjects do not practise this discrimination then they will not be able to perform it when it is necessary, as in random presentation without the highlighted feature. Therefore subjects will lack some of the appropriate productions for efficient performance during random transfer and so will exhibit partial transfer similar to that exhibited by blocked-trained subjects. On the other hand, if intratrial processing underlies the contextual interference effect, then the fact that presentation during training in this experiment is random means that subjects will have practice at loading solution methods into working memory. This will then result in complete transfer to random presentation, similar to that exhibited by subjects trained with random presentation without the identification cue.

Results from all three experiments will be combined to examine further the possibility that the training differences that have been observed between random and blocked presentation are a result of the execution of different numbers of productions. If transfer between two tasks is determined by the extent to which the tasks share identical productions, then performance differences between the training conditions should predict transfer differences. For example, consider the case where there are two training conditions (random and blocked) and only one transfer condition (random). If performance during blocked training was shown to be faster than during random training, then this suggests that a different number of productions are being executed in each of these conditions. Blocked training, being the fastest, would be associated with the smaller number of productions. If it is
the case that the productions developed with random training are necessary to perform efficiently with random presentation, then the productions developed with blocked training will not be sufficient for such presentation. The development of new productions will be necessary for subjects in this condition. This would result in blocked-trained subjects performing slower with random presentation because these subjects would be required to learn new, and therefore slow, productions. In sum, ACT* predicts that training differences will be good indicators of transfer differences. No such predictions can be made on the basis of the intratrial processing account.

All of the above predictions will be presented in greater detail in the following chapter.
# Chapter 3  Experiments 1, 2 and 3

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Overview</td>
<td>57</td>
</tr>
<tr>
<td>3.2 Experiment 1</td>
<td>61</td>
</tr>
<tr>
<td>3.2.1 Introduction</td>
<td>61</td>
</tr>
<tr>
<td>3.2.1.1 ACT* Predictions</td>
<td>61</td>
</tr>
<tr>
<td>3.2.1.2 Intratrial Processing Account Predictions</td>
<td>67</td>
</tr>
<tr>
<td>3.2.1.3 Measures of Transfer</td>
<td>69</td>
</tr>
<tr>
<td>3.2.2 Method</td>
<td>71</td>
</tr>
<tr>
<td>3.2.2.1 Subjects</td>
<td>71</td>
</tr>
<tr>
<td>3.2.2.2 Materials</td>
<td>71</td>
</tr>
<tr>
<td>3.2.2.3 Design</td>
<td>73</td>
</tr>
<tr>
<td>3.2.2.4 Apparatus and Procedure</td>
<td>74</td>
</tr>
<tr>
<td>3.2.3 Results and Discussion</td>
<td>77</td>
</tr>
<tr>
<td>3.2.3.1 Training</td>
<td>77</td>
</tr>
<tr>
<td>3.2.3.2 Transfer</td>
<td>86</td>
</tr>
<tr>
<td>3.3 Experiment 2</td>
<td>97</td>
</tr>
<tr>
<td>3.3.1 Introduction</td>
<td>97</td>
</tr>
<tr>
<td>3.3.2 Method</td>
<td>100</td>
</tr>
<tr>
<td>3.3.2.1 Subjects</td>
<td>100</td>
</tr>
<tr>
<td>3.3.2.2 Materials</td>
<td>100</td>
</tr>
<tr>
<td>3.3.2.3 Design</td>
<td>100</td>
</tr>
<tr>
<td>3.3.2.4 Apparatus and Procedure</td>
<td>100</td>
</tr>
<tr>
<td>3.3.3 Results and Discussion</td>
<td>101</td>
</tr>
<tr>
<td>3.4 Experiment 3</td>
<td>104</td>
</tr>
<tr>
<td>3.4.1 Introduction</td>
<td>104</td>
</tr>
<tr>
<td>3.4.2 Method</td>
<td>108</td>
</tr>
<tr>
<td>3.4.2.1 Subjects</td>
<td>108</td>
</tr>
<tr>
<td>3.4.2.2 Materials</td>
<td>108</td>
</tr>
<tr>
<td>3.4.2.3 Design</td>
<td>108</td>
</tr>
<tr>
<td>3.4.2.4 Apparatus and Procedure</td>
<td>110</td>
</tr>
<tr>
<td>3.4.3 Results and Discussion</td>
<td>110</td>
</tr>
<tr>
<td>3.5 Combined Analyses of Experiments 1, 2 and 3</td>
<td>114</td>
</tr>
<tr>
<td>3.5.1 Introduction</td>
<td>114</td>
</tr>
<tr>
<td>3.5.2 Subjects</td>
<td>115</td>
</tr>
<tr>
<td>3.5.3 Results and Discussion</td>
<td>116</td>
</tr>
<tr>
<td>3.5.3.1 Training</td>
<td>116</td>
</tr>
<tr>
<td>3.5.3.2 Transfer</td>
<td>123</td>
</tr>
<tr>
<td>3.6 Conclusions</td>
<td>126</td>
</tr>
</tbody>
</table>
3.1 Overview

In Experiments 1, 2 and 3 subjects were taught how to solve syllogisms. They were given extensive practice with feedback in order to develop skilled performance, that is, to be able to assess the truth of the syllogisms quickly and accurately. In all three experiments the syllogisms were categorical with both premises being of the universal affirmative type. For example, "All of the artists are beekeepers. All of the beekeepers are chemists. (Therefore) All of the artists are chemists." Two forms of these syllogisms were used. The example just given will be described as being of the ABBC type (the letters corresponding to the order of the elements in the premises, i.e., A for artist, B for beekeeper etc.). The second type of syllogism will be described as BCAB, which is similar to the ABBC type except that the order of the premises is reversed. Both of these forms have equivalent conclusions.

Syllogisms were chosen as stimulus materials because they have a fixed structure leading to a fixed conclusion. The structure can be considered as having variable slots that are filled with different words, though the logic of the syllogism remains the same. Thus these types of syllogisms have a fixed stimulus-response relationship. For instance, the true conclusion for both types of syllogisms was always of the form "All of the A's are C's."

The method of presenting the syllogisms to the subjects was the same throughout the three experiments reported in this chapter. The two premises of each syllogism were presented for study first. When the subjects finished studying them, they were presented with a conclusion which they had to judge as either TRUE or FALSE with reference to the implication of the premises. Premise study times and conclusion reaction times and accuracy were recorded in order to dissociate components of the solution process.
There are many theories concerned with the types of strategies that people use to solve syllogisms and similar problems such as linear series problems (e.g., Clark, 1969a, 1969b; Erickson, 1974; Johnson-Laird, 1983; Maybery, 1987; Sternberg, 1980). However all of these theories describe behaviour in situations where people with little or no experience with these types of problems attempt to solve only a small number of problems. Therefore these theories do not attempt to describe strategies that develop with practice. In fact all of the strategies suggested in the above theories would be inefficient in a situation where subjects were required to solve quickly a large number of syllogisms of the same structure (cf. Charney & Reder, 1987).

Quinton and Fellows (1975) found that practice at solving three-term linear series problems led to subjects changing their solution strategies. Initially subjects were apparently using a "thinking" strategy based on syllogistic logic, but with practice strategies took on a perceptual character based on the invariant structure of the problems. Quinton and Fellows also reported that these perceptual strategies enabled faster and more efficient performance. It was reasonable to expect then, that subjects in the present study would develop perceptual strategies also. An efficient perceptual strategy for solving the syllogisms used in the present study would centre on the fact that the conclusion for each syllogism can be associated with the form of the syllogism. Therefore the strategy should include a component that is aimed at identifying syllogism type. To identify syllogism type a subject could use the perceptual strategy of noting the positions of the common elements in the premise pairs. Figure 3.1 illustrates the goal-structure inherent in such a strategy.

An important feature of the strategy depicted in Figure 3.1 is that it is hierarchical in nature. The goal of solving a syllogism is segmented into a
Figure 3.1: Goal structure underlying solution of syllogisms with random presentation. Flow of control indicated by arrows.
number of sub-goals and some of these sub-goals have further sub-goals underlying them. This follows the characterisation of production systems described by Anderson in the ACT* theory. As is evident from the figure, some goals need to be satisfied before others for a solution to result. For example, in order to derive a conclusion from a premise pair (goal 2 in Figure 3.1), it is necessary to identify syllogism type (goal 3). Syllogism type is associated with the position of the common elements in the premises and so can be determined by locating the position of these elements (goal 4). Goal 5 involves processing the uncommon elements of the premises in terms of what they are and orienting them in accordance with the syllogism type to form a true conclusion. For example, if a syllogism is of the BCAB type, then the uncommon elements are the second element in the first premise and the first element in the second premise. The processes involved in goal 5 will orient these elements so that the expected true conclusion will be "All of the [first element in second premise] are [second element of first premise]."

Undoubtedly the strategy depicted in Figure 3.1 does not represent the most efficient strategy for solving syllogisms in the experimental tasks of this study. The only conclusions that were presented were true ones as described above and false ones which were the converse of the true conclusions (i.e., if true conclusion = All of the A's are C's, then false conclusion = All of the C's are A's). Thus, one very efficient strategy would be to examine only the first element in each premise. Then, if either of these matched the first element in the presented conclusion then the conclusion is true. Otherwise the conclusion must be false. Thus the form of the syllogisms would not be important in such a strategy, in contrast to the strategy portrayed in Figure 3.1. However, the results suggest that, in general, subjects did not use such an efficient strategy and indeed relied on identifying syllogism type.
Chapter 3

3.2 Experiment 1

3.2.1 Introduction

This experiment was designed to replicate the contextual interference effect with a cognitive task involving syllogistic inference. As with the Shea and Morgan (1979) study, subjects were trained with either a random or a blocked order of presentation of items. Similarly, subjects could then be tested with a transfer task that involved either random or blocked presentation. During Training, subjects were presented with almost 300 trials, one at a time. Only a third of this number of trials were presented during the Transfer phase of the experiment. Premise study times and conclusion accuracy and reaction times enabled a test of the predictions derived from the ACT* and intratrial processing accounts of the contextual interference effect. These predictions are summarised in Table 3.1 and will be discussed below.

3.2.1.1 ACT* Predictions

The first prediction that can be made from the ACT* account is based on the assumption that Random and Blocked Training will lead to different sets of productions. Random training will lead to productions with a goal-structure similar to the one described in Figure 3.1. Blocked Training, on the other hand, will enable subjects to perform without the need to develop some of the productions that identify syllogism type. As a result subjects will not need to scan the premise pairs for the position of the common elements. Instead, within blocks, true conclusions will always consist of the same elements from the premises (i.e., if the present block is all ABBC, then the order of elements in the true conclusion will be the first element in the first premise followed by the second element in the second premise; if the present block is all BCAB,
<table>
<thead>
<tr>
<th>Condition</th>
<th>ACT*</th>
<th>Intratrial Processing</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Training</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random vs Blocked</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conclusion RT's</td>
<td>No difference in Conclusion RT's because the same set of productions will be developed in Blocked and Random Training</td>
<td>No Prediction</td>
</tr>
<tr>
<td>Premise RT's</td>
<td>Blocked Premise RT's will be faster than Random Premise RT's because less productions will be executed in the former</td>
<td>No Prediction</td>
</tr>
<tr>
<td><strong>Transfer</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Conclusions</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-R</td>
<td>Complete Transfer in all conditions because the same productions will be developed in Blocked and Random Training and used in Blocked and Random Transfer</td>
<td>Complete Transfer in all conditions because presentation of true and false conclusions is random in all conditions</td>
</tr>
<tr>
<td>R-B</td>
<td></td>
<td>Complete Transfer</td>
</tr>
<tr>
<td>B-R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B-B</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Premises</strong> (Random)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-R</td>
<td>Complete Transfer because the same productions will be used in both phases</td>
<td>Complete Transfer</td>
</tr>
<tr>
<td>B-R</td>
<td>Partial Transfer because subjects will not have had sufficient practice at identifying syllogism type</td>
<td>No Prediction</td>
</tr>
</tbody>
</table>

Continued Overleaf
R-R vs B-R

R-R faster than B-R because R-R subjects have well-practised productions; B-R lack important productions.

R-R faster than B-R because R-R subjects have had a lot of practice at loading solution methods into working memory; B-R subjects lack this practice.

Premises
(Blocked)

R-B

Either Complete Transfer because the same productions will be used in both phases
Or subjects will be faster during Transfer than is predicted from Training because less productions will be used in the second phase

No clear prediction

B-B

Complete Transfer because the same productions will be used in Transfer as were developed in Training

Complete Transfer

R-B vs B-B

Either no difference because both Training conditions will elicit appropriate productions for performance in Blocked presentation
Or, if Random Trained subjects do not notice Blocked presentation and continue to check for syllogism type on each trial, then they will be slower during Transfer

No clear prediction

Table 3.1: Summary of predicted results for Experiment 1 based on the ACT* and intratrial processing accounts of the contextual interference effect. Complete Transfer indicates that Transfer performance is consistent with that predicted on the basis of Training performance as described by a power function (details in §3.4 Results). Partial Transfer indicates that Transfer performance is significantly slower than that predicted on the basis of Training performance as described by a power function. Note: R-R = Random Training - Random Transfer; B-R = Blocked Training - Random Transfer; R-B = Random Training - Blocked Transfer; B-B = Blocked Training - Blocked Transfer.
then the elements in the true conclusion will be the first element in the second premise and the second element in the first premise). Only with each change of block will it be necessary for subjects to 'think' about the difference in syllogism type. Certainly subjects who are trained with Blocked items will not have nearly as much practice with this characteristic of the task as subjects trained with Random items. As a result the Blocked subjects will not become skilled at identifying syllogism type. With respect to the strategy described in Figure 3.1, Blocked Trained subjects may disregard goal 4 and only attempt goal 3 on the first trial of each block. Achieving goal 5 will rely on the result of achieving goal 3 being carried over to all trials in a block. Retaining this information for all of the trials within a block eliminates the need for a sub-goal such as goal 4.

The above analysis only applies to productions concerned with the premises of each syllogism. Productions dealing with conclusions should be identical following Random or Blocked Training. At the conclusion stage of the task subjects simply have to check the presented statement on the screen with their expected true conclusion. If the two statements match they should respond "true", otherwise "false." Therefore strategies dealing with this part of the task should not differ across the two conditions.

The design of Experiment 1 involved the factorial combination of Random versus Blocked Training and Random versus Blocked Transfer. Thus four groups of subjects were used. This enabled the determination of which Training condition allowed the best Transfer performance, and whether this transfer was dependent on Transfer condition.

The ACT* account enables detailed predictions of transfer in the above conditions given that certain assumptions are made. The first is that the time
to perform a task is determined by the number of steps (productions) involved in the task (Anderson, 1983; Staszewski, 1988). Therefore the more complex a task or the strategy for performing the task, the longer it will take to perform. The second assumption is that transfer between tasks is determined by the number of productions that are common to performance of the tasks (Anderson, 1987). The greater the number of common productions, the greater the amount of transfer.

The first prediction concerns performance times during Training. As was suggested above, Blocked presentation during Training should lead to production sets that have fewer steps to solution than those developed with Random Training. The first prediction based on this assumption is that premise times will be faster during Blocked Training than during Random Training. In addition, because the difference in the number of steps is only a feature of the premises part of the production sets, there should not be any difference in conclusion reaction times between the two Training conditions.

The second prediction based on the assumption that Random and Blocked Training lead to different sets of productions concerns transfer. Transfer predictions will be classified in terms of complete and partial transfer. Empirical definitions of these terms will be detailed below. However, in simple terms, complete transfer refers to a situation where performance during Transfer is at least as fast as was observed during Training. Partial transfer refers to a situation where performance during Transfer is slower than observed during Training.

Consistent with the contextual interference effect, subjects trained with a random presentation order will show better transfer to other randomly presented syllogisms than subjects trained with Blocked syllogisms. This is
because subjects with Blocked Training will not have developed and practised productions for identifying the form of syllogisms. Hence they will be ill-equipped for randomly presented items. As a result transfer will be partial. In other words, their performance will be slower than it was during Training and also slower than that of subjects trained with a random presentation. However, this will only be the case with premise study times, where the identification of syllogism form is important. Conclusion reaction times will depend on similar productions for both the Random Trained group of subjects and the Blocked Trained group. Thus both groups will show complete transfer with conclusion reaction times. That is, Transfer performance will be at least as fast as performance during Training. Only the group trained with Random presentation will show complete transfer with premise study times.

It is not clear what will happen with subjects who are trained with Random presentation and then are tested with Blocked presentation. One possibility, albeit an unlikely one, is that the subjects may not notice the difference between the Training phase and the Transfer phase and so will continue on as if the Blocked presentation order is a Random one. As a result there would be complete transfer, as with subjects trained and tested with Random presentation. Another possibility is that subjects will notice the Blocked order and so realise that they do not need to identify syllogism form. This would mean that they could use fewer productions with each item and so will have faster premise times than they did during Training. Conclusion reaction times, on the other hand, will not be affected by the shift from Random to Blocked presentation. Thus complete transfer will be exhibited with conclusion reaction times in this condition.

The last condition in Experiment 1 involves subjects trained with Blocked presentation being tested with another phase of Blocked presentation. As with
the Random Training to Random Transfer condition, productions developed during Training will be perfectly adequate for the Transfer phase. Hence subjects should show complete transfer in all measures.

3.2.1.2 Intratrial Processing Account Predictions

Some of the results predicted by the ACT* account for the conditions examined in Experiment 1 are similar to the results predicted by the intratrial account. However the two accounts arrive at the same predictions for different reasons. Implicit in the strategy depicted in Figure 3.1 is the assumption that the two syllogism types require different solution methods: The solution to an ABBC syllogism will be comprised of different elements to the solution of a BCAB syllogism. Random presentation of the two syllogism types during Training will give subjects practice at loading these different solution methods into working memory. In contrast, Blocked Training will enable subjects to retain one solution method in working memory for many trials before a new method is required. Therefore Blocked Trained subjects will not have had the same amount of practice at loading solution methods into working memory as Random subjects. As a result, the Blocked Trained subjects should show poorer transfer to Random presentation than Random Trained subjects.

It is unclear what prediction the intratrial processing account would make for subjects trained with Random presentation and then given Blocked Transfer items. As described above, Random Training will provide practice at loading solution methods into working memory. Whether or not this will be a benefit during Blocked presentation, where the loading of solution methods into working memory is not as frequent, is not clear.
Subjects trained with Blocked presentation, according to the intratrial processing account, should be adequately prepared for more Blocked items during Transfer. These subjects would have loaded solution methods into working memory during Training with the same frequency as would be required during Transfer. Hence there is no reason to expect performance to be affected.

The prediction of transfer differences in the four conditions based on the intratrial processing account should only apply to premise study times. In all conditions, the presentation of True and False conclusions was random. Therefore, with respect to the loading of solutions into working memory for the processing of conclusions, stimulus conditions are unchanged from Training to Transfer. Thus, conclusion reaction times should be unaffected and so complete transfer should be observed in all conditions.

Where the intratrial processing account differs from the ACT* account is in the prediction of differences in performance times during Training. As described above, the ACT* account predicts that Training performance times will depend on the number of productions underlying performance. Thus performance during Blocked Training will be faster than during Random Training because performance in the former condition is controlled by fewer productions than in the latter condition. More specifically, this difference in the number of productions is predicted to be restricted to the processing of premises. As a result, only premise study times will be affected by the different Training conditions. In contrast, the intratrial processing account makes no predictions about how performances during Random and Blocked Training will compare (Carlson & Yaure, 1990).
No predictions are obvious from either the ACT* or intratrial processing accounts concerning accuracy of performance. Accuracy, like response times, was assumed to reflect the level of skill at which subjects were performing. However, subjects received a substantial amount of training with a simple task, and were required to perform at an accuracy level of greater than 75% in the last half of the Training trials. As a result, it was expected that they would be performing at or near to ceiling accuracy when the Transfer trials were presented. Therefore the accuracy measure is possibly not sufficiently sensitive to test some predictions.

3.2.1.3 Measures of Transfer

The measures of transfer used in this study need further explanation. Two measures were used in order to be able to fully evaluate the predictions described above. The first measure involved direct comparisons of reaction times during the Transfer phases of each condition. For example, in two of the conditions, subjects were trained with either Random or Blocked presentation and then performed with Random presentation during the Transfer phase. The first measure of transfer was a direct comparison of the reaction times of these two groups of subjects on the Random Transfer items. If one group was faster on these items than the other, this would indicate that the Training conditions of the first group provided better preparation for performance in this Transfer condition than those of the second group.

The second measure of transfer used in this study involved within-group comparisons of performance during Training and Transfer. The purpose of this measure was to evaluate whether changes in presentation condition from Training to Transfer affected the number of steps involved in performing the task. As described above, the ACT* predictions were based on the
assumption that performance with Blocked presentation would rely on less productions per problem than with Random presentation. The intratrial account predicts that Random and Blocked presentation differ only with respect to the amount of practice afforded one particular component of the task - the loading of solution methods into working memory. The second measure of transfer was designed to be sensitive to this distinction between the two accounts.

There are a number of assumptions that underlie the second measure of transfer, though these are reasonable assumptions considering previous research. The first is that performance times in the syllogistic reasoning task will improve according to the power law of learning. That is, the logarithm of performance time, when plotted against the logarithm of amount of practice, will approximate a straight line (e.g., Newell & Rosenbloom, 1981). Secondly, if stimulus conditions do not change from Training to Transfer, performance during Transfer should continue to improve according to the power function observed during Training. Anderson's (1982) account of the power law of learning (see Chapter 4 for details) includes the assumption that performance time is a function of the number of productions involved in performing the task, such that the more productions executed the longer the performance time (see Anderson, 1982, for a description of a study that supports this assumption). This account suggests that if the change in stimulus conditions from Training to Transfer results in a change in the number of productions executed, then performance time should deviate from the power function observed during Training in a predictable fashion. For instance, the ACT* account predicts that Blocked Trained subjects will need to execute more productions per problem than was the case during Training to cope with Random presentation during Transfer. Therefore these subjects should show an increase in performance time from Training to Transfer. In
other words, the performance times of these subjects should show an upward deviation from the power function extrapolated from Training. Similarly if the productions executed during Training are the only ones applied during Transfer, then the learning function should not show any substantial deviation from Training to Transfer.

3.2.2 Method

3.2.2.1 Subjects

One hundred and thirty-seven volunteers from the University of Western Australia Psychology Department participated in the experiment for course credit or $4. Nine subjects failed to reach the learning criterion and their data were not further analysed. The learning criterion was defined as an error rate not in excess of 25% in the last half of the Training trials. Although an arbitrary value, this cut-off point clearly distinguished subjects who mastered the task and became faster throughout Training from subjects who obviously did not understand the task. The latter subjects usually responded at or little better than chance (error rate = 50%) and their reaction times did not improve with practice. This was found to be the case in Experiments 2 and 3 also. The exclusion of these subjects left 128 subjects in the experiment, 32 per condition.

3.2.2.2 Materials

Three hundred and eighty-four syllogisms were constructed. They were all of a categorical form (see Table 3.2). The premises were constructed so that one term denoted an occupation and two terms denoted preoccupations, interests or roles (e.g., father). This constraint minimised semantic relations within
<table>
<thead>
<tr>
<th>Premises</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ABBC</td>
<td>BCAB</td>
</tr>
<tr>
<td>All of the acrobats are butchers.</td>
<td>All of the butchers are cricketers.</td>
</tr>
<tr>
<td>All of the butchers are cricketers.</td>
<td>All of the acrobats are butchers.</td>
</tr>
<tr>
<td>response -&gt;</td>
<td>READY</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conclusion</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(True)</td>
<td>All of the acrobats are cricketers.</td>
</tr>
<tr>
<td>(False)</td>
<td>All of the cricketers are acrobats.</td>
</tr>
<tr>
<td>response -&gt;</td>
<td>TRUE or FALSE</td>
</tr>
</tbody>
</table>

Table 3.2: Sample syllogism used in Experiment 1 illustrating the two forms of premise pairs (ABBC and BCAB), the two forms of conclusions (TRUE and FALSE) and the appropriate responses to each part of the syllogism.
premise pairs without sacrificing the plausibility of the premises or conclusions, valid or invalid (Johnson-Laird & Steedman, 1978). Premise pairs were avoided that led to contradictory or confusing conclusions (e.g., All of the fathers are aunties.).

The first 96 syllogisms were constructed using 288 different terms. The following 96 items were constructed on the basis of the first set by using a different permutation of terms across items. The next 96 items were also a different permutation of the set of terms. These three sets of 96 items were used as Training items. Thus each of the 288 terms was presented three times during Training, combined with different terms each time, and also in a different position of the syllogism each time. An additional set of 96 items was constructed by deriving another permutation of the first 96 items. These additional items were used as Transfer items. A further 2 syllogisms were constructed and presented as practice items.

Half of the syllogisms were always presented in ABBC form, the other half were always in BCAB form (see Table 3.2). Across subjects each item was presented half the time with a True conclusion and half with a False conclusion (see Table 3.2).

3.2.2.3 Design

The experiment had four conditions which resulted from the factorial combination of the Training conditions (Random vs. Blocked) with the Transfer conditions (Random vs. Blocked).

Training: Items were presented in sets of eight though these were not apparent to the subjects. Within each set the items were presented in a new random
order for each subject. In the Random condition each set contained four ABBC syllogisms and four BCAB syllogisms. In the Blocked condition the first three sets of items were always ABBC, the next three sets were all BCAB, and these alternated so that subjects were presented with blocks of 24 items of the same type. Thus every 48 trials subjects in both the Random and Blocked conditions saw equal numbers of ABBC and BCAB problems (see Figure 3.2 for a summary of the design of Training trials).

**Transfer:** The Transfer trials were presented in a similar order to the Training trials. In the Random condition 12 sets of eight items were presented, with four ABBC problems and four BCAB problems in each set. The order of presentation was random within each set. In the Blocked condition four blocks of 24 items were presented in the following order: ABBC(24), BCAB(24), ABBC(24), BCAB(24). Thus again every 48 trials subjects in both conditions were presented with equal numbers of ABBC and BCAB type problems.

### 3.2.2.4 Apparatus and Procedure

The experiment was controlled by a PDP 11/73 computer. All experimental sessions were conducted in sound proof booths. Syllogisms were presented on a CRT and subjects made responses on a keyboard.

Subjects were tested in one one hour session. They were instructed that they would be presented with a number of small problems to solve. Each problem would consist of three statements. The first two statements (premises) would describe particular types of people. The third statement (conclusion) could be a conclusion based on what the first two statements said. The subjects' task
Figure 3.2: A summary of the design of Training trials in Experiment 1. Within each set of eight trials items were presented in a random order. This design sequence was repeated every 48 trials with a new set of items each time. Thus every 48 trials subjects in both Random and Blocked conditions saw the same number of ABBC and BCAB items.
was to decide whether the third statement was true or false of what the first two statements described.

Each item was presented as follows (see Table 3.2): The premises were presented on the screen. Subjects were instructed to press the "READY" button (space bar of the keyboard) when they had read the premises. Premises were visible for a maximum of ten seconds. When a subject pressed "READY", or did not respond within ten seconds, the premises disappeared. The conclusion then appeared on the screen below where the premises had been. Subjects were then to press "TRUE" (the 'z' key of the keyboard) or "FALSE" (the '/' key of the keyboard) dependent on how they felt the statement corresponded to the preceding premise pair. Again subjects were given a maximum time of ten seconds to respond. When a subject pressed "TRUE" or "FALSE", or did not respond within ten seconds, the conclusion disappeared. Following this subjects were provided with feedback on the screen for two seconds. This consisted of a sentence of the form: "Correct/Incorrect, conclusion was True/False," dependent on the subjects' response. No feedback was provided if subjects did not press "TRUE" or "FALSE". After the feedback disappeared from the screen the next item was presented automatically.

Subjects were given two practice trials at the beginning of the experiment. They were presented in the same fashion as the experimental trials, were both of the ABBC type, the first one was presented with a True conclusion, the second with a False conclusion, and subjects were provided with feedback after each trial. Following these practice trials subjects were presented with the 288 Training trials. At the end of the Training phase subjects were given one minute rest during which they were instructed on the screen that the next phase of the experiment (Transfer phase) would be similar to the previous
phase though much shorter. Subjects were not forewarned when Training and Transfer conditions were different (i.e., when one condition was Random and the other was Blocked). After the rest period, the computer automatically presented the 96 trials in the Transfer phase.

3.2.3 Results and Discussion

The data were analysed in blocks of 48 trials. Thus in all conditions performance measures in each block were based on 24 ABBC trials and 24 BCAB trials. Mean Premise and Conclusion reaction times based on correct trials only were calculated for each block. These times for all conditions are presented in Figure 3.3. Both the Premise and Conclusion RTs during Training show the characteristic reduction that occurs with practice. All of the groups appeared to improve at similar rates during Training but exhibited varying amounts of transfer. For example, the Blocked-to-Random subjects were amongst the fastest in premise RTs at the end of Training but were the slowest during the Transfer phase. These results will be analysed in greater detail below.

3.2.3.1 Training

Three ANOVAs were conducted on the Training data. A 2 (Training condition) x 6 (Training Block) x 2 (syllogism type) ANOVA was used to analyse the Premise RTs, whereas two 2(Training condition) x 6 (Training Block) x 2 (syllogism type) x 2 (conclusion type) ANOVAs were used to analyse Conclusion RTs and Accuracy.

The subjects as a whole improved significantly with each block of practice: Premise RTs went from 3619 ms at the beginning of Training to 2948 ms at
Figure 3.3: Mean Premise and Conclusion RTs in Experiment 1.
the end \(F(5,630) = 204.04, p<0.05\); Conclusion RTs went from 1900 ms to 825 ms \(F(5,630) = 164.08, p<0.05\); and Accuracy went from 78.30% to 96.91% \(F(5,630) = 90.53, p<0.05\).

The Premise and Conclusion RTs for the Random and Blocked Training conditions are presented in Figure 3.4. The figure has log-log axes and the lines drawn through the data points represent best-fit power functions. Consistent with the power law of learning (e.g., Newell and Rosenbloom, 1981) that underlies skill acquisition in ACT* (Anderson, 1982), these power functions provide a good fit to the data. It is apparent from Figure 3.4 that during Blocked Training subjects consistently studied the premises for a shorter time than those trained with randomly presented items (2948 ms vs. 3619 ms) \(F(1,126) = 9.78, p<0.05\). However there were no differences between the two conditions in Conclusion RTs \((F(1,126)<1)\) or Accuracy \((F(1,126)=3.05, p>0.05)\).

The fact that subjects trained with Blocked presentation studied premises for a shorter time than Random Trained subjects replicates half of the contextual interference effect. In addition this result supports the ACT* prediction that the different forms of training would lead to different processing strategies. This result suggests that Blocked subjects were able to process the premises with a smaller number of operations than Random subjects. Furthermore, the lack of any differences in Conclusion RTs suggests that the two groups processed the conclusions with a similar number of operations. This result was also predicted on the basis that the two forms of training would not induce different strategies for assessing conclusions.

During Random Training Premise RTs were fastest when the items were of the ABBC type (3487 ms vs. 3751 ms) \((F(1,63) = 31.94, p<0.05)\). There
Figure 3.4: Mean Premise and Conclusion RTs for combined Random and Blocked Training conditions in Experiment 1, plotted on log-log axes. Lines are best fit power functions with the following equations:

Random Premises: $RT = 5672.2 \text{ Block}^{(-0.44)}$, $r^2 = 0.996$;
Blocked Premises: $RT = 4962.6 \text{ Block}^{(-0.52)}$, $r^2 = 0.993$;
Random Conclusions: $RT = 1846.6 \text{ Block}^{(-0.48)}$, $r^2 = 0.992$;
Blocked Conclusions: $RT = 1798.1 \text{ Block}^{(-0.46)}$, $r^2 = 0.965$. 
was no such difference between ABBC and BCAB items during Blocked Training ($F(1,63) < 1$). Conclusion RTs showed no significant effect of syllogism type during Random Training ($F(1,63) = 2.05$, $p>0.05$), but showed an effect during Blocked Training. This effect was in the opposite direction with Conclusion RTs to ABBC syllogisms being slower than to BCAB items (1190 ms vs. 1079 ms) ($F(1,63) = 38.00$, $p<0.05$). The form of syllogisms affected Accuracy only during Random Training where subjects were more accurate with ABBC syllogisms (91.21% vs. 89.92%) ($F(1,63) = 6.62$, $p<0.05$). The only significant interaction of interest with regards to this effect was between syllogism type and block number during Random Training ($F(5,315) = 5.40$, $p<0.05$). Inspection of this interaction revealed that ABBC syllogisms were responded to faster than BCAB items, though this difference declined with practice, from 469 ms to 134 ms.

The finding that Premise RTs were faster with ABBC syllogisms than with BCAB problems during Random Training suggests that the processing strategy of these subjects created a bias in favour of ABBC problems. One strategy that could result in such a bias involves subjects scanning the elements in each premise in a conventional left-to-right, down-the-page manner in order to identify syllogism type. This strategy would enable subjects to locate the common elements of ABBC syllogisms after scanning only three of the four elements in each premise pair. All four elements would need to be processed in order to locate the common elements in BCAB syllogisms. If this were the case, then subjects could identify ABBC syllogisms in less time than BCAB syllogisms. An accuracy advantage for ABBC syllogisms compared to BCAB syllogisms was reported by Johnson-Laird (1983). Johnson-Laird proposed that this resulted from the manipulation of mental models of the meaning of each premise in order to derive a solution. BCAB syllogisms are suggested to require more
manipulation than ABBC syllogisms before a solution can be derived, thus increasing the likelihood of error. Unfortunately there is no way of verifying either of these strategies with the current methodology.

The fact that the syllogism type effect only occurred during Random Training provides support for the contention that the strategies developed by Blocked subjects would not feature steps that identify syllogism type. The same sort of reasoning predicted that identification of syllogism type would not be important at the conclusion processing stage and so would not affect Conclusion RTs. This prediction was supported by the lack of a difference between the two types of syllogism during Random Training. However during Blocked Training Conclusion RTs were slower with ABBC syllogisms. It is not clear what sort of strategy could have caused this difference nor why the two Training conditions should induce a strategy difference at this stage of processing. Certainly Conclusion RTs were similar throughout both forms of training which suggests that at least the strategies developed in these conditions were similar in terms of the number of processing operations.

The above results support the assumption of the ACT* account that the major difference between the strategies induced by Random and Blocked Training would concern whether subjects actively checked for the form of syllogisms before proceeding to derive a solution. It seems clear that Random Trained subjects did identify syllogism type as part of their processing strategy but there is no guarantee in the above results that the Blocked subjects did not pay attention to syllogism type. However, it would have been unnecessary for them to do so on each trial. In fact for these subjects only the first trial of each block required identification of syllogism type. After that, all items in the block were of the same type. It is possible to check whether Blocked subjects
were indeed following such a strategy. The serial-position curves in Figure 3.5 represent mean RTs on all trials for both Training conditions. The scalloping effect in the upper half of this figure shows clearly that Blocked Trained subjects increased the length of their premise study time considerably at the beginning of each new block, when syllogism type changed, but then required much shorter times during the rest of the block. Subjects trained with Random items on the other hand showed a uniform reduction in Premise RTs throughout practice. Therefore it appears that during Blocked Training subjects need only identify syllogism type once per block to enable efficient performance. On the other hand Random subjects need to identify syllogism type on every trial. Examination of serial-position curves for Conclusion RTs in the lower half of Figure 3.5 shows that there is a similar though much reduced effect of blocking on Conclusion RTs. Although it was assumed that syllogism type is not important at the conclusion processing stage, this result suggests that there is some carry-over of the effect of premise type into the time assumed to reflect the processing of conclusions only.

True conclusions were responded to faster than False ones in both Random Training (1076 ms vs. 1207 ms, F(1,63) = 36.96, p<0.05) and Blocked Training (1071 ms vs. 1198 ms, F(1,63) = 58.48, p<0.05). There was also a significant difference in Accuracy with subjects recognising True conclusions more accurately in both Random Training (92.12% vs. 89.01%, F(1,63) = 18.97, p<0.05) and Blocked Training (93.66% vs. 91.93%, F(1,63) = 7.63, p<0.05). There were significant interactions between the effect of True versus False conclusions and block number in Conclusion RTs during Blocked Training (F(5,315) = 4.02, p<0.05), and in Accuracy during Random Training (F(5,315) = 7.06, p<0.05) and Blocked Training (F(5,315) = 3.33, p<0.05). In all of these interactions the differences between True and False conclusions diminished with practice.
Figure 3.5: Serial position curves of mean Premise and Conclusion RTs during Training in Experiment 1.
The fact that True conclusions were responded to faster during both forms of Training supports a prediction made earlier on the basis of the ACT* account. This prediction was that Random and Blocked Training would lead to similar strategies for processing conclusions. Furthermore, this result suggests that with this task, subjects appear to have a bias for matching predicted events. That is, subjects derive hypothetical conclusions on the basis of premises, and if these match the presented conclusions, subjects will respond faster than if their expectations are not met. A similar result has been reported when subjects are explicitly requested to predict which stimulus will appear before each trial in a choice RT task (Simon & Craft, 1989). This phenomenon may be related to reports of a matching or confirmation bias in reasoning (e.g., Evans, 1989).

The major conclusions to be drawn from the Training results are that Random and Blocked Training lead to different strategies for processing premises but similar strategies for processing conclusions. The differences in processing premises relate to the identification of syllogism type. Random Training appears to induce strategies for this identification whereas Blocked Training precludes the need to develop such operations. Certainly such operations will have very little practice during Blocked Training. In contrast, the two Training conditions induce similar strategies for processing conclusions. These strategies involve a bias for True conclusions over False ones. In short, the Training results provide support for the ACT* account of performance with Random and Blocked Training; that Blocked Training will result in a processing strategy that relies on the execution of less productions than are necessary during Random Training. Furthermore, this difference between the processing strategies developed in the two Training conditions appears to be related to the identification of syllogism type. The intratrial
processing account is unable to account for the differences observed during Training.

3.2.3.2 Transfer

It is not immediately clear from Figure 3.3 whether there were any significant transfer effects. The first stage in examining this issue involved performing a number of analyses of variance: 2 (Training condition) x 2 (Transfer Block) x 2 (syllogism type) ANOVAs were used to analyse Premise RTs, and 2 (Training condition) x 2 (Transfer Block) x 2 (syllogism type) x 2 (conclusion type) ANOVAs were used to analyse Conclusion RTs and Accuracy. First of all the performance of subjects trained with either Random or Blocked presentation on Random Transfer items was assessed. These subjects did not differ in Premise RTs (Mean Premise RTs for Random Trained subjects = 2169 ms vs. Blocked Trained subjects = 2285 ms; F(1,62) < 1), Conclusion RTs (Random = 720 ms vs. Blocked = 824 ms; F(1,62) = 3.38, p>0.05) or Accuracy (Random = 97.65% vs. Blocked = 97.10%; F(1,62) < 1). In contrast, Blocked Trained subjects (1770 ms) studied the Premises of Blocked Transfer items for less time than Random Trained subjects (2240 ms; F(1,62) = 4.44, p<0.05). However there were no differences between these two groups of subjects in Conclusion RTs (Random = 812 ms vs. Blocked = 810 ms; F(1,62) = 0.00) or Accuracy (Random = 95.38% vs. Blocked = 96.88%; F(1,62) = 2.31, p>0.05).

The apparent lack of any difference between Random and Blocked Training in performance with Random Transfer items constitutes a failure to replicate the contextual interference effect. However, closer inspection of Figure 3.3 suggests that there may be some results that are consistent with this effect. The obvious advantage observed with Premise RTs during Blocked Training
compared to during Random Training appears to be reversed, or at least reduced, with the transition to Random Transfer items. Premise RTs in all conditions appear to improve with the transition from Training to Transfer except in the Blocked to Random condition where the reverse was observed. This interaction between condition (R-R, B-R, R-B, B-B) and the transition from Training (last block) to Transfer (first block) was significant (F(3,124) = 16.31, p<0.05). However the comparison of Random Trained subjects with Blocked Trained subjects on performance in the first block of Random Transfer failed to reach significance at the 0.05 level (F(1,62)<1). Therefore, despite the differences in Premise RTs observed during Training between Random and Blocked subjects, this difference was not reversed during Random Transfer, although the results were in this direction (in the first block, Random = 2142 ms vs. Blocked = 2282 ms). This failure to fully replicate the contextual interference effect will be examined in more detail in the final section of this chapter where results of Experiments 1, 2 and 3 will be combined in one analysis. However, some implications will be considered in this section following the presentation of the results of the second measure of transfer.

As described earlier, the second measure of transfer assumes that performance during Transfer can be predicted on the basis of the power function observed during Training if the same productions are executed during both phases. If more or less productions are applied during Transfer as were executed during Training then this will be evident by deviations from the power function. This assumption was used to investigate whether differences in production numbers could be used to account for Training and Transfer performance in the four conditions.
Power functions of the form $RT = N \cdot P^c$ (where $P$ = number of blocks of practice) were fitted to the Training data in all four conditions and then extrapolated into the Transfer phase. In order to decide whether Transfer performance constituted a significant deviation from the practice function observed during Training, confidence limits (alpha level = 0.05) were calculated for the Transfer data. If extrapolated performance falls within these confidence limits then Transfer performance can be said to be consistent with a continued improvement of the productions developed during Training. If extrapolated performance falls outside these limits then performance during Transfer is consistent with the application of a different number of productions than were executed during Training.

Figures 3.6, 3.7, 3.8 and 3.9 display Premise and Conclusion RTs of the four conditions during Training and Transfer. Power functions have been fitted to the Training data and extrapolated into the Transfer phase, where confidence limits for the Transfer data have also been included. It is clear from Figure 3.6 that the Premise and Conclusion RTs for the Random Transfer trials are consistent with what could have been expected on the basis of performance during Random Training. Only Conclusion RTs during the second block of Transfer appear to be slower than expected. A similar pattern of results was found with subjects trained with Blocked items and then presented with Blocked Transfer items (Figure 3.7). The results for these two conditions imply that the strategies developed during Training for processing premises and conclusions were sufficient for efficient performance during the Transfer phase. This is as would be expected because the Transfer conditions are not different to the Training conditions. Strategies that enabled efficient performance during Training should be just as efficient during Transfer.
Figure 3.6: Mean Premise and Conclusion RTs during Training (Random) and Transfer (Random) phases of Experiment 1. Error bars are confidence limits (alpha = 0.05). Lines are power functions fitted to Training data and extrapolated into the Transfer phase, with the following equations:

Premises: RT = 4947.2 Block (-0.43), \( r^2 = 0.956 \);
Conclusions: RT = 1774.1 Block (-0.50), \( r^2 = 0.971 \).
Figure 3.7: Mean Premise and Conclusion RTs during Training (Blocked) and Transfer (Blocked) phases of Experiment 1. Error bars are confidence limits (alpha = 0.05). Lines are power functions fitted to Training data and extrapolated into the Transfer phase, with the following equations:

Premises: \( RT = 4941.6 \text{ Block}^{-0.54}, r^2 = 0.993; \)
Conclusions: \( RT = 1771.9 \text{ Block}^{-0.46}, r^2 = 0.952 \)
Figure 3.8: Mean Premise and Conclusion RTs during Training (Random) and Transfer (Blocked) phases of Experiment 1. Error bars are confidence limits (alpha = 0.05). Lines are power functions fitted to Training data and extrapolated into the Transfer phase, with the following equations:

Premises: \( RT = 6274.2 \text{ Block}^{(-0.45)}, r^2 = 0.974; \)

Conclusions: \( RT = 1918.1 \text{ Block}^{(-0.46)}, r^2 = 0.999 \)
Figure 3.9: Mean Premise and Conclusion RTs during Training (Blocked) and Transfer (Random) phases of Experiment 1. Error bars are confidence limits (alpha = 0.05). Lines are power functions fitted to Training data and extrapolated into the Transfer phase, with the following equations:

- Premises: $RT = 4981.4 \text{ Block}^{-0.51}$, $r^2 = 0.989$;
- Conclusions: $RT = 1824.4 \text{ Block}^{-0.46}$, $r^2 = 0.974$
In the condition where subjects were trained with Random items and then presented with Blocked items, Transfer performance as measured by Premise RTs was better than expected. Figure 3.8 shows that extrapolated performance predicts a slower premise RT than was found in the first block of Transfer trials. The mean RT in the second block of Transfer trials however was no different than predicted. The Conclusion RTs in the Transfer phase of this condition were consistent with the extrapolated times.

Subjects who were trained with Random presentation studied premises during Blocked Transfer for less time than would be expected on the basis of their prior performance and further practice. The implication of this finding is that these subjects were able to process the premises during Blocked Transfer using fewer steps than were used during Random Training. Therefore it appears that where it was not necessary to identify syllogism type on each trial (i.e., Blocked Transfer) these subjects were able to circumvent those processing steps that previously performed this operation. A different result was observed with Conclusion RTs in this condition. Improvement from Training to Transfer trials appeared to be consistent with improvement predicted by the power function observed during Training. Therefore whatever strategies were developed during Random Training for processing conclusions were sufficient for efficient performance with Blocked Transfer items.

In the remaining condition subjects were trained with Blocked items and then presented with Random items during the Transfer phase. Figure 3.9 shows that during Transfer the premise RTs of these subjects were slower than was expected on the basis of their Training performance. This was the case in both blocks of Transfer trials. It is also clear from this figure that the average Conclusion RT during the first block of Transfer trials is not significantly
different to the extrapolated time but that the average RT for the second block is slower than expected.

As predicted by the ACT* account, Blocked Trained subjects were not well-equipped for Random Transfer items. The assumption underlying this prediction was that Blocked Training would reduce the necessity to practise the identification of syllogism type. However this skill is crucial to efficient performance during random presentation. Therefore, during Transfer, these subjects would be skilled at all aspects of the task except one and so would be much slower at that particular processing stage than at those that had been practised during Training. This accounts for the finding that Blocked Trained subjects were slower with Random Transfer items than was expected on the basis of their Training performance. The deficit resulted from having to execute productions that were not well-practised. In contrast, the strategies that were developed by these subjects to process conclusions appeared to have been sufficient for efficient processing during Random Transfer, as there was no significant slowing with Transfer.

In two conditions there appeared to be a significant slowing of Conclusion RTs in the second block of Transfer (see Figures 3.6 and 3.9) whereas in the first Transfer block RT was consistent with that expected on the basis of Training performance. It is unlikely that strategies that were appropriate for performance during the first block of Transfer suddenly lost their efficiency during the second Transfer block. Instead it is more likely that the slowing is simply a result of not including a non-zero asymptote in the power functions (i.e., functions of the form $RT = A + NP_0$, where $A>0$). It was difficult to fit power functions to the data with non-zero asymptotes. The simple power functions provided such a high degree of fit to the data that including such asymptotes did not improve the degree of fit appreciably. More particularly,
fitting curves to some of the learning functions benefited from including non-zero asymptotes, whereas no such benefit was obvious with other functions. For the sake of consistency only, non-zero asymptotes were used in all conditions. Thus these results are probably due to subjects approaching asymptotes sooner than the simple power functions describe. The key results for Premise RTs were not affected by the use of power functions without non-zero asymptotes.

The failure to fully replicate the contextual interference effect in this experiment creates difficulties in evaluating the ACT* and intratrial processing accounts of this effect. Both accounts predicted that, in the syllogism task, subjects trained with Random presentation would have shorter Premise RTs with Random Transfer items than subjects with Blocked Training. The results did not support this prediction, although they were consistent with the difference predicted. However, another feature of the contextual interference effect was observed in this experiment. Random Training was shown to be slower than Blocked Training. This result provides support for the ACT* account because it only occurred when the ACT* account predicted that the two forms of Training would induce different processing strategies. Thus Random Training was shown to induce a strategy that relied on the identification of syllogism type. Performance during Blocked Training did not require this operation. As a result the time needed to study premises was shorter during Blocked Training than during Random Training. In contrast, no differences were found between the Training conditions in terms of the strategies developed to process conclusions. The reaction time differences observed during Random and Blocked Training cannot be accounted for by notions of intratrial processing.
In addition to the support for the ACT* account of Training differences, the Transfer results were consistent with the suggestion that transfer is determined by common productions. Subjects who were trained with Random presentation had no trouble with Random Transfer items because they had been trained for that type of situation. Blocked Trained subjects were similarly well-equipped for Blocked Transfer items. In both of these conditions the ACT* account predicted that the productions developed during Training would be suitable for application during Transfer. This is because the stimulus conditions during the second phase were identical to those present during the first phase. In contrast Blocked Trained subjects had little practice identifying syllogism type and so lacked well-practised productions for performing this skill. As a result, slow productions were executed to process the premises of Random Transfer items. Hence these subjects performed slower during this phase than if they had continued to execute only the well-practised productions that were applied during Training. No clear prediction was made concerning transfer from Random Training to Blocked items. However, the fact that subjects in this condition processed premises faster than was expected on the basis of Training performance suggests that less productions were applied during Transfer than were executed during Training.

In conclusion, the results of Experiment 1 demonstrate that the ACT* theory of skill acquisition and transfer can be used to predict performance differences resulting from different training conditions, and also the extent of transfer associated with these different conditions. The transfer results that supported the ACT* predictions were all within-subjects analyses. The ACT* predictions that were not supported relied on between-subjects comparisons. This suggests that the failure to replicate the transfer feature of the contextual interference effect may result from a lack of sensitivity of the between-
subjects analyses with this particular task. This issue will be treated in more
detail in the final section of this chapter. In addition, it is for this reason that
the intratrial processing account cannot be disregarded at this stage as a
reasonable explanation of the contextual interference effect, despite the fact
that it is unable to account for the observed Training differences and the
within-subjects Transfer differences. Thus the ACT* and intratrial processing
accounts will be contrasted further in Experiments 2 and 3.

3.3 Experiment 2

3.3.1 Introduction

Experiment 2 was concerned with the effect of a Training condition similar to
Lee and Magill's (1983) serial condition. In this experiment the two types of
syllogism were presented to subjects in a fixed alternating order throughout
Training. During the Transfer phase presentation order was random. The
ACT* account and the intratrial processing account predict different effects of
Alternating and Random presentation. For example, predictions based on the
ACT* account centre on the fact that, with Alternating presentation, syllogism
type will be completely predictable from one trial to the next. This means that
subjects need not develop the skill of identifying syllogism type. These
subjects will be able to solve syllogisms without the need of goals 3 and 4 in
the strategy outlined in Figure 3.1. This suggests that during Training, these
subjects will perform faster than subjects who need these stages in their
solution method (e.g., Random Trained subjects). Only one condition was
used in this experiment, so this prediction is evaluated in the final section of
this chapter in a combined analysis of the results of Experiments 1, 2 and 3.
In addition to predicting that Alternating presentation will affect the type of strategy developed, the ACT* account predicts that this type of Training will not result in complete transfer to Random presentation. If the skill of identifying syllogism type is not developed during Training, subjects will not be well-equipped for Random presentation, where such a skill is necessary for efficient performance. As a result, transfer from Alternating to Random presentation will be positive, because some productions developed during Training will be applied during Transfer, but will only be partial because new productions will be necessary. Thus, the ACT* account predicts that the Random Transfer performance of subjects with Alternating Training will show a significant deviation from the power function observed during Training. This partial transfer should also be evident in subjects Trained with Alternating presentation performing slower during Random Transfer than subjects with Random Training (to be examined in the final section of this chapter).

The Alternating presentation during the Training phase of this experiment means that new solution methods will be loaded into working memory on each trial. According to the intratrial processing account this will result in subjects developing efficient procedures for loading and applying solution methods in working memory. Thus these subjects should have sufficient practice at this aspect of the task to cope with Random presentation during Transfer. Therefore, in contrast to the ACT* account, the intratrial processing account predicts that subjects with Alternating training will not be different from subjects with Random Training when Transfer involves randomly presented items (to be examined in the final section of this chapter). In addition, the intratrial processing account predicts that in terms of loading solution methods into working memory, Random Transfer will provide no more of a challenge than Alternating Training. Therefore the Transfer
performance of these subjects should not show any marked deviation from the learning function observed during Training.

The above predictions only apply to the processing of premises (i.e., Premise RTs). The occurrence of True and False conclusions in both Random and Alternating conditions is random. Therefore both accounts predict that complete transfer will be observed in Conclusion RTs from Alternating to Random presentation.

The Alternating Training phase of this experiment can be viewed as an attempt to contrast in one condition the features that, according to the two accounts, result in the contextual interference effect. According to the ACT* account Alternating Training and Blocked Training share the same degree of predictability with respect to the occurrence of the different types of syllogism. Therefore ACT* predicts that subjects in this experiment will behave like subjects in the Blocked Training - Random Transfer condition of Experiment 1. In contrast, Alternating Training has the same degree of contextual interference as Random Training. According to the intratrial processing account, this means that subjects with Alternating Training will have the same amount of practice at loading solution methods into working memory as subjects with Random Training. Thus the subjects in this experiment should exhibit results similar to those of the Random Training - Random Transfer condition of Experiment 1.
3.3.2 Method

3.3.2.1 Subjects

Thirty-three students from the University of Western Australia first year Psychology course volunteered to participate in this experiment as partial fulfilment of a course requirement. One student failed to reach the learning criterion of an error rate of not more than 25% in the last half of Training. Only the data from the remaining 32 subjects were analysed.

3.3.2.2 Materials

The items used in this experiment were identical to those used in Experiment 1. Only the order of presentation differed.

3.3.2.3 Design

One group of subjects was observed in this experiment. During Training, all subjects were presented with the same syllogisms in the same order. An ABBC syllogism was presented on the first trial, followed by a BCAB syllogism on the second trial, and then an ABBC syllogism on the third trial, and so on. This alternating order was fixed throughout Training. During Transfer, the presentation order of the two types of syllogism was random and so was similar to the Random Transfer conditions in Experiment 1.

3.3.2.4 Apparatus and Procedure

These were identical to those described in Experiment 1.
3.3.3 Results and Discussion

A number of ANOVAs were conducted to analyse the Training and Transfer data. With the Training data a 6 (Training block) x 2 (syllogism type) ANOVA was used to analysed Premise RTs, whereas two 6 (Training block) x 2 (syllogism type) x 2 (conclusion type) ANOVAs were used to analyse Conclusion RTs and Accuracy.

Mean Premise and Conclusion RTs observed during Training and Transfer are displayed in Figure 3.10. As is obvious from this figure, there was a significant improvement in Premise RTs during Training, from 6959 ms in the first block of trials to 2789 ms in the sixth block (F(5,155) = 65.12, p<0.05). Conclusion RTs were also reduced with practice, from 2219 ms in the first block of Training to 901 ms in the last block (F(5,155) = 69.62, p<0.05). Accuracy improved from 78.85% to 97.12% (F(5,155) = 29.23, p<0.05).

Syllogism type again had an effect on Premise RTs during Training. Subjects studied the premises of ABBC syllogisms for less time than the premises of BCAB problems (3942 ms vs. 4164 ms; F(1,31) = 13.10, p<0.05). During this phase, Conclusion RTs were faster with True conclusions than with False conclusions (1144 ms vs. 1322 ms; F(1,31) = 59.38, p<0.05).

With the Transfer data, a 2 (Transfer block) x 2 (syllogism type) ANOVA was used to analyse Premise RTs, and 2 (Transfer block) x 2 (syllogism type) x 2 (conclusion type) ANOVAs were used to analyse Conclusion RTs and Accuracy. The only interesting significant effect of these analyses was that of conclusion type: True conclusions were responded to faster than False conclusions (766 ms vs. 838 ms; F(1,31) = 6.12, p<0.05).
Figure 3.10: Mean Premise and Conclusion RTs during Training (Alternating) and Transfer (Random) phases of Experiment 2. Error bars are confidence limits (alpha = 0.05). Lines are power functions fitted to Training data and extrapolated into the Transfer phase, with the following equations:

Premises: \( RT = 6883.3 \text{ Block}^{-0.54}, r^2 = 0.988 \);
Conclusions: \( RT = 2035.2 \text{ Block}^{-0.51}, r^2 = 0.947 \)
With respect to the predictions based on the ACT* and intratrial processing accounts, the only ones that will be evaluated at this stage concern the comparison of performance during Transfer with that predicted on the basis of Training performance. It is clear from Figure 3.10 that in both Premise and Conclusion RTs, performance during Random Transfer did not deviate significantly from the performance predicted by extrapolating the learning functions observed during Training. Therefore the Random Transfer items apparently did not provide subjects with any more of a challenge than was posed by the Alternating Training phase. This result was predicted by the intratrial processing account and so supports the suggestion that practice at loading solution methods into working memory is an important consideration in transfer to a random presentation order.

The Transfer results of this experiment were not predicted by the ACT* account. However, what this says about the ACT* account is not entirely clear. On one hand it may indicate that transfer is not determined by the extent to which tasks share common identical productions. On the other hand, the results may simply indicate that Alternating Training did not affect processing strategy in the way predicted. For instance, consider the fact that subjects in this experiment processed the premises of ABBC syllogisms faster than the premises of BCAB syllogisms. This suggests that these subjects may not have made use of the predictable nature of the Alternating presentation to identify syllogism type. If they had made use of this feature, they would have known in advance of each trial which premise elements would comprise the true conclusion. If the speed advantage of ABBC syllogisms is a result of the procedure for scanning premise elements to locate common elements (see § 3.2.3.1), then being able to skip this process should eliminate the ABBC advantage. Certainly this appeared to be the case with Blocked Training in Experiment 1. Although the exact processing strategy used can only be
speculated upon at this stage, the observation of the ABBC advantage during Alternating Training does suggest that subjects were not taking advantage of the predictable nature of this presentation condition. As a result the Alternating Training condition was effectively the same as Random Training and so transfer to Random items should have been complete, as was observed. Therefore there is at least one reasonable interpretation of the results that is consistent with the ACT* account.

In conclusion, it is not clear if the results of this experiment are useful in distinguishing between the two accounts. However, these results will be compared with those of Experiments 1 and 3 in the final section of this chapter. This set of comparisons should indicate whether performance during Alternating Training was faster than observed during Random Training in Experiment 1, as predicted by the ACT* account. In addition, these comparisons should also indicate how the subjects in this experiment performed with Random Transfer items in relation to Random Trained subjects. Therefore these further comparisons will shed more light on the usefulness of the results of this experiment.

3.4 Experiment 3

3.4.1 Introduction

In Experiment 3 the two syllogism types were presented in a random order during both Training and Transfer phases. However, during Training, the identification of syllogism type was made easier by highlighting the feature that this identification was based on. The purpose of making this discrimination easier was to encourage subjects to rely on the highlighting
feature rather than learn to make the discrimination that is necessary when this cue is not present.

In Experiment 1 subjects trained with Blocked presentation only needed to identify syllogism type once per block. For all other trials in each block syllogism type was the same as in the first trial of the block. Thus, in a sense, except for the first trial of each block, identification of syllogism type was accomplished for these subjects. The strategy described in Figure 3.1 assumes that identification of syllogism type is associated with locating common elements in the premise pairs (goal 4 in Figure 3.1). This location of common elements relies on a search and match process that compares all four elements in premise pairs to determine which are the same. If this location of common elements could be accomplished for subjects then it is unlikely that they would develop the skill of performing such a location. For instance, if these elements were highlighted in comparison to other words in each premise pair then the task of locating them would be much easier. Subjects could then rely on the highlighting to locate these elements rather than developing the skill of determining which elements are common. As with Blocked Training then, highlighting would reduce the need to develop a particular skill and so would reduce the probability that the skill would be developed.

In this experiment one training condition was used. In this condition the categorical syllogisms used in Experiment 1 were presented in a random order. The common elements of each premise pair were capitalised in order to highlight them. Following Training, subjects were tested with random presentation of normal lower case items. These Transfer items were the same as those used in the Random Transfer phase of Experiments 1 and 2.
The assumption behind capitalising the common elements in the premise pairs is that it should make the task of identifying syllogism type easier by enabling subjects to omit locating common elements. The capitalising should provide a visual contrast between the common and uncommon elements, creating two diagonals; one of common elements, the other of uncommon elements. After some practice the contrast may be sufficient to allow subjects to refrain from checking that the common elements are indeed common. Subjects could simply identify syllogism form with a perceptual strategy. This strategy would rely on the direction of the diagonal of common elements. With practice subjects could associate a particular direction of the diagonal that contains the common elements with a particular syllogism type. For example, a syllogism of the ABBC type corresponds to a common elements diagonal that is inclined from the bottom left of the premises to the upper right. Thus perception of diagonal direction will amount to an identification of syllogism type. As a result, this identification (i.e., goal 3 in Figure 3.10) can be performed without the need to determine the location of common elements, as was necessary with Random presentation in Experiment 1. According to the ACT* account, this suggests that subjects will execute less productions to process premises with highlighting than to process premises without highlighting. Therefore ACT* predicts that this will result in faster Premise RTs during Training with highlighted common elements than is observed with items that lack this highlighting. This prediction will be evaluated in the final section of this chapter.

According to the ACT* account, the effect of capitalising premise elements on the development of strategies has implications for transfer. The prediction is that subjects will rely on superficial physical characteristics of premises to identify syllogism type rather than develop and practise productions for locating common elements. However, this is unlikely to benefit subjects
during the Transfer phase. As was found in Experiment 1, when Training conditions are such that certain processing operations can be omitted, transfer to conditions that require those operations is affected adversely. The Transfer items in this experiment were randomly presented without capitalised elements. Thus efficient performance with these items requires productions that locate common elements in order to identify syllogism type. This implies that Training in the present experiment will induce a set of productions that will not be well-equipped to process the normal Random Transfer items. As a result subjects should be slowed during the Transfer phase compared to their performance during Training. This result should only be apparent with Premise RTs. Conclusion RTs should not be affected by the change from capitalised elements to lower-case elements.

The intratrial processing account predicts a different set of results for this experiment. The fact that presentation during Training is random means that subjects will have practice at loading solution methods into working memory. This will then result in complete transfer to random presentation, as was observed with subjects trained with random presentation without the identification cue.

The inclusion of a highlighting cue in this experiment should be seen as a further attempt to simulate the features of Blocked Training that, according to the ACT* account, created the Training and Transfer effects observed in Experiment 1. In contrast, it is important to note that the random presentation order used in this experiment means that the degree of contextual interference is equivalent to that in the Random conditions of Experiment 1. As a result the experiment is also simulating the conditions that, according to the intratrial processing account, creates the Transfer advantage of Random Training. Therefore support for the ACT* predictions will suggest that the sharing of
identical productions underlies transfer and hence the contextual interference effect. On the other hand, support for the predictions based on the intratrial processing account will suggest that practice at loading solution methods into working memory underlies the contextual interference effect.

3.4.2 Method

3.4.2.1 Subjects

Forty volunteers from the University of Western Australia Psychology Department participated in this experiment for course credit or $4. Eight subjects failed to reach the learning criterion of an error rate not in excess of 25% in the last half of the Training phase. The data from these subjects were excluded from further analysis, leaving 32 subjects.

3.4.2.2 Materials

The items used in this experiment were identical to those presented in the Random Training-Random Transfer condition in Experiment 1 except for one feature. In the Training phase subjects were presented with premise pairs that had the common elements capitalised (see Table 3.3). In the Transfer phase the items were presented in the same fashion as in Experiment 1.

3.4.2.3 Design

There was one Training condition in this experiment. Subjects were presented with premise pairs with the capitalised common elements. The order of presentation of ABBC and BCAB problems was random, with the same constraints as described in Experiment 1 for Random Training. During the
Table 3.3: Sample syllogism used in Experiment 3 illustrating the premise pairs used (only ABBC shown), the two forms of conclusions (TRUE and FALSE) and the appropriate responses to each part of the syllogism.
Transfer phase of the experiment syllogisms were presented with all elements in lower case. The two types of syllogism were presented in a random order in this phase also.

3.4.2.4 Apparatus and Procedure

These were identical to those in Experiment 1.

3.4.3 Results and Discussion

A number of ANOVAs were conducted to analyse the Training and Transfer data. With the Training data a 6 (Training block) x 2 (syllogism type) ANOVA was used to analysed Premise RTs, whereas two 6 (Training block) x 2 (syllogism type) x 2 (conclusion type) ANOVAs were used to analyse Conclusion RTs and Accuracy.

Mean Premise and Conclusion RTs observed during Training and Transfer trials are presented in Figure 3.11. As is indicated clearly by this figure, the subjects improved with each block of Training. Premise RTs went from 6205 ms in the first block of Training to 1751 ms in the last block (F(5,155) = 99.21, p<0.05). Conclusion RTs went from 2005 ms to 791 ms (F(5,155) = 53.71, p<0.05). Accuracy also improved with practice, from 71.16% to 96.22% (F(5,155) = 30.26, p<0.05).

During the Training phase, the premises of ABBC syllogisms were studied for shorter durations than BCAB syllogisms (3057 ms vs. 3287 ms, F(1,31) = 23.12, p<0.05). Syllogism type had no effect on Conclusion RTs (F(1,31)<1). However, Accuracy was affected by syllogism type (ABBC=90.45% vs. BCAB=88.50%, F(1,31) = 7.51, p<0.05). True
Figure 3.11: Mean Premise and Conclusion RTs during Training (Highlight) and Transfer (Random) phases of Experiment 3. Error bars are confidence limits (alpha = 0.05). Lines are power functions fitted to Training data and extrapolated into the Transfer phase, with the following equations:

Premises: \( RT = 6105.1 \text{ Block} ^{-0.70}, r^2 = 0.999; \)

Conclusions: \( RT = 1831.8 \text{ Block} ^{-0.49}, r^2 = 0.941 \)
conclusions were responded to faster than False conclusions (1072 ms vs. 1178 ms, $F(1,31) = 16.10, p<0.05$). True conclusions were also responded to more accurately (91.23% vs. 87.72%, $F(1,31) = 8.58, p<0.05$).

These results replicate the observations in Experiment 1 that indicate biases in the processing strategies of subjects. Again ABBC syllogisms were found to provide an advantage over BCAB syllogisms in terms of processing premise pairs, and True conclusions appeared to have a processing advantage over False conclusions.

The Transfer results are presented in Figure 3.11. It is clear from this figure that Premise RTs were affected to a large extent by the change from capitalised elements to normal items. Conclusion RTs on the other hand appear not to have been affected by this change. In this figure the lines again represent power functions that provide the best fit to the Training data and have been extrapolated into the Transfer phase. It is clear that Premise RTs during the Transfer phase are significantly slower than would be expected on the basis of Training performance. Conclusion RTs in the second block of trials during Transfer also appear to be slower than the expected times though this again is a result of not using power functions with non-zero asymptotes. It is obvious though that Conclusion RTs were not affected by the change from Training to Transfer to the same extent as Premise RTs.

These Transfer results confirm the predictions of the ACT* account. Subjects were adversely affected by the change from capitalised elements to normal items during the Transfer phase. This effect was such that subjects studied premises in the Transfer items for longer durations than was observed at the end of Training. This result supports the assumption that subjects trained with capitalised elements would not be equipped to perform efficiently with items
that did not contain capitalised elements. Obviously these subjects relied on the capitalised elements feature of the premises to solve the syllogisms during Training. Relying on this feature meant that productions for locating common elements would not have been developed. These productions are essential for identifying syllogism type in normal items. Therefore, as with the Blocked Trained subjects in Experiment 1, these subjects did not have appropriate productions to process premises efficiently in the normal items during the Transfer phase of this experiment. However, the productions that were developed during Training to process conclusions appear to have been adequate for processing the conclusions of the Transfer items.

The results indicate that during Highlight Training subjects developed strategies which did not enable efficient performance during Random Transfer. This inefficiency in processing strategy was predicted to result from a lack of practice at locating common elements in order to identify syllogism type. However, the results do indicate that the performance of subjects during Highlight Training was sensitive to syllogism type. Despite the fact that common elements were highlighted during this phase, ABBC syllogisms were still processed faster than BCAB syllogisms. Thus, relying on perception of a diagonal of highlighted common elements to identify syllogisms type may also involve a speed advantage for ABBC syllogisms.

The Transfer results of this experiment do not support the predictions of the intratrial processing account. The presentation order of syllogism type (ABBC vs. BCAB) and conclusion type (True vs. False) was random throughout the experiment. However, the processing of premises, where syllogism type is important, revealed a different pattern of results to the processing of conclusions, where conclusion type is important. Furthermore, the significant slowing of Premise RTs with the transition from Highlight
Training to Random Transfer was similar to that observed in Experiment 1 from Blocked Training to Random Transfer. Therefore a random presentation order is not a sufficient condition for complete transfer to another random presentation condition. This suggests that practice at loading solution methods into working memory may not be the only factor underlying the contextual interference effect. In fact, given that the ACT* theory is able to account for all of the results in this experiment, intratrial processing may not play a role at all. This issue will be addressed further in the following section, the last of this chapter, where results from Experiments 1, 2 and 3 will be combined in a number of omnibus analyses.

3.5 Combined Analyses of Experiments 1, 2 and 3

3.5.1 Introduction

The aim of the combined analysis of selected results from Experiments 1, 2 and 3 was mainly to provide a further evaluation of some of the ACT* predictions. The first set of predictions concerned the effect of Training condition on the number of productions underlying performance. The ACT* account predicted that fewer productions would be executed in Blocked, Highlight and Alternating Training than in Random Training. This difference in production numbers should be evident in faster Premise RTs for these three conditions compared to performance during Random Training. No such differences were predicted by the ACT* account for Conclusion RTs. In contrast, the intratrial processing account made no predictions about relative performance during the four Training conditions.

The second set of ACT* predictions that were examined by the combined analysis concern transfer and are related to the predictions about the execution
of differing numbers of productions during Training. If transfer between two tasks is determined by the extent to which the tasks share identical productions, then performance differences between the Training conditions should predict Transfer differences. For instance, as described above, the Blocked, Highlight and Alternating Training conditions are predicted to result in strategies that involve fewer productions than are applied during Random Training. If it is the case that the productions developed with Random Training are necessary for efficient performance with random presentation, then the productions developed with the other three Training conditions will not be sufficient for such presentation. Subjects in these conditions will find it necessary to develop new productions to perform efficiently during Random Transfer. As a result, subjects in these three Training conditions will perform slower during Random Transfer than subjects with Random Training. Again these predictions only apply to Premise RTs. No differences are predicted between the four Training conditions in Conclusion RTs and so no differences in Conclusion RTs during Random Transfer should be observed. Thus Transfer predictions based on the ACT* account are tied to differences observed during Training. In contrast, Transfer predictions based on the intratrial processing account are unrelated to differences observed during Training because this account makes no predictions of differences during Training.

3.5.2 Subjects

In order to be able to make the most accurate assessment of Training differences, data from all of the 128 subjects of Experiment 1 were used in the combined analysis. This meant that Training data from 64 Random Trained subjects and 64 Blocked Trained subjects were examined in this analysis. Training data from the 32 Alternating Trained subjects from
Experiment 2 and from the 32 Highlight Trained subjects of Experiment 3 were also used. In order to have equal numbers of subjects in each Training condition, data from another 32 subjects with Alternating Training were included in the analysis. This data came from another experiment which will be reported in Part 2 of the thesis (Experiment 6). The subject population and the Training conditions in this experiment were identical to those in Experiment 2. Similarly, Training data from another 32 subjects with Highlight Training were obtained from another experiment to be reported later in the thesis (Experiment 4).

The data used in the combined analysis of Transfer differences were from those subjects in Experiments 1, 2 and 3 who were presented with Random Transfer items. Thus data from 32 subjects in each of the four Training conditions were available.

3.5.3 Results and Discussion

Mean Premise and Conclusion RTs during Training and Transfer for all conditions examined in this analysis are displayed in Figure 3.12. The analysis of the Training data will be described first, followed by the analysis of the Transfer data. Only those analyses that are related to the predictions will be described.

3.5.3.1 Training

A number of ANOVAs were conducted to analyse the Training data. A 4 (Training condition) x 6 (Training block) x 2 (syllogism type) ANOVA was used to analysed Premise RTs, whereas two 4 (Training condition) x 6 (Training block) x 2 (syllogism type) x 2 (conclusion type) ANOVAs were
Figure 3.12: Mean Premise and Conclusion RTs during Training and Transfer (all Random) phases of Experiments 1, 2 and 3. Extra data in Highlight Training condition are from subjects in Experiment 4. Extra data in Alternating Training are from Experiment 6 (see text for details).
used to analyse Conclusion RTs and Accuracy. These analyses indicated that there was an overall effect of Training condition in Premise RTs ($F(3,252) = 6.56$, $p<0.05$), but no such effect in Conclusion RTs ($F(3,252) = 1.24$, $p>0.05$) or Accuracy ($F(3,252) = 1.34$, $p>0.05$). In all measures there was a significant effect (all $p$'s$<0.05$) of syllogism type and an interaction of this variable with Training condition. Inspection of the data revealed that in all measures, ABBC syllogisms enjoyed a processing advantage over BCAB syllogisms in all conditions except Blocked Training, where the reverse was the case.

*Post-hoc* comparisons were made between the Premise RTs of the four conditions using Tukey's WSD procedure (Maxwell & Delaney, 1990). The mean Premise RT for Random Training (3619 ms) was compared with the mean Premise RTs of each of the other Training conditions. The comparison with Blocked Training (2948 ms) was significant ($F(1,126) = 9.78$, $p<0.05$). However, the comparisons with Highlight Training (3010 ms) and Alternating Training (3826 ms) were not significant (Highlight: $F(1,126) = 5.86$, $p>0.05$; Alternating: $F(1,126) = 0.70$). The initial perception then is that Premise RTs during Blocked Training were faster than during Random Training, and that performance during the other two Training conditions did not differ from Random Training. However, inspection of Figure 3.12 reveals a slightly different picture. Performance with Alternating presentation clearly did not differ from that with Random presentation during the whole of the Training phase. On the other hand, Premise RTs during Blocked Training are clearly faster than during Random Training. However, in the early blocks of Training, the Premise RTs of Highlighting Training do not differ substantially from those of Random Training. But in the later blocks of Training the Highlighting Training performance appears to be not only faster than the Random Training Premise performance but also the Blocked Training
performance. For this reason two further series of post-hoc comparisons were made, one series based on the data of the last three blocks and, the other based on the last block of Training only. With respect to the ACT* predictions that Transfer differences should be associated with Training differences, the number of productions executed in the final stages of Training is the important factor in the degree of transfer that will be observed. Performance in the early stages of Training is likely to be confounded with comprehending the task and developing appropriate sets of productions.

A 4 (Training condition) x 3 (Training block) x 2 (syllogism type) ANOVA was used to analyse Premise RTs in the last three blocks of Training. This analysis revealed that there was an overall effect of Training condition on Premise RTs (F(3,252) = 11.57, p<0.05). Post-hoc comparisons using Tukey's WSD procedure indicated that performance during both Blocked (2148 ms) and Highlight (1798 ms) Training was faster than during Random Training (2794 ms) (Random vs. Blocked: F(1,126) = 12.62, p<0.05; Random vs. Highlight: F(1,126) = 21.82, p<0.05). No difference was observed between the Random and Alternating (2865 ms) Training conditions (F(1,126) < 1). The same pattern of results was observed in comparisons of mean Premise RTs during the final block of Training.

Performance during Blocked Training appeared to be consistently faster than performance during Random training. In contrast, performance during Highlight Training was only faster than with Random Training in the latter half of this phase (see Figure 3.12). Although these results support the ACT* prediction that performance in both Blocked and Highlight Training would rely on the execution of fewer productions than with Random Training (at least in the latter half of Training), the different pattern of results observed in the two conditions suggests that subtle differences exist in the way Blocked
and Highlight Training encourage the development of particular processing strategies. This can be observed by comparing the learning curves of these two conditions during Training (see Figure 3.12). The learning curve of the Highlight Trained group has a slower intercept value than that of the Blocked Trained group. However, the curve of the Highlight Trained group improved at a faster rate than that of the Blocked Trained group, eventually overtaking that group’s performance, so that in the latter half of the Training phase, the Highlight Trained subjects were faster than the Blocked Trained subjects. Therefore performance in these two conditions is not simply a matter of bypassing some of the processing steps that performance during Random Training requires. The different ways in which these two training conditions enable performance without these processing steps may result in the different course of improvement.

One way to account for the different patterns of improvement observed during Blocked and Highlight Training is to consider the different ways in which these two training conditions enable performance without the need to execute all of the processing steps that are necessary during Random Training. Consider the Highlight condition. During the first block of trials, subjects may not realise the significance of the highlighting of common elements until they realise the significance of the common elements in solving the syllogisms. Initially then, these subjects have no advantage over the Random Trained subjects who also are faced with the task of understanding the role of the common elements. This suggestion is supported by results displayed in Figure 3.12, where, in the first block of trials, the performance of the Highlight Trained group is similar to that of the Random Trained group, if not slower. Eventually though, once the Highlight Trained group comprehend the association between the common elements and the solution to the syllogisms, the usefulness of the highlighting feature should be clear. Gradually, as
confidence in the relationship between the highlighting and the solution grows, reliance on this feature will become more frequent. This would then explain why the Highlight Trained subjects gradually become faster than the Random Trained subjects (see Figure 3.12). In contrast, consider the Blocked Trained group. The repetition of syllogism type, and therefore solution method, should have been obvious after only a few trials. Therefore the advantage over performance with Random Training should have been available almost immediately. This is certainly suggested by the results displayed in Figure 3.12. Furthermore, this advantage in processing time appeared to be constant throughout the Training phase. Thus, both the Blocked and Highlight Training conditions enabled solution of the syllogisms in less time than was observed with Random Training, but in different ways, and these different methods seem to have resulted in different patterns of improvement.

In summary, two of the ACT* predictions concerning effects on Premise RTs were supported by the Training results. Performance during both Blocked and Highlight Training was faster than observed during Random Training. These results support the suggestion that these two Training conditions would encourage the development of processing strategies that involved the execution of fewer productions per problem than would be the case with Random Training, although the results suggested that the execution of fewer productions was achieved in different ways in these two conditions. A third ACT* Training prediction, however, was not supported by the results. At no stage during this phase did performance during Alternating Training appear to be any faster than during Random Training.

The observation that performance with Alternating Training was not significantly different than with Random Training turns out to be open to an
interpretation that is not inconsistent with the ACT* account. Originally this account predicted that the Alternating condition would result in a processing strategy that relied on the predictable order in which ABBC and BCAB syllogisms would appear. If this were the case then fewer productions would underly performance in this condition resulting in faster Premise RTs. Apparently though, the subjects in this study did not take advantage of the alternating presentation order. The ACT* account does not actually predict that subjects will take advantage of such short-cut methods. Rather this account only predicts what will happen if subjects follow this course. It was assumed that subjects would take advantage of such an obvious time and effort saving solution method if they perceived it as such. The results suggest though, that the subjects either did not notice the alternating presentation order or did not realise its usefulness. Alternatively, maintaining a record of which type of syllogism was presented on the previous trial in order to anticipate the type of syllogism to be presented on the forthcoming trial may be an expensive or aversive procedure. Subjects may decide that such an effort is not worthwhile, and so will develop a strategy similar to that developed by subjects with Random Training. In other words, they will identify syllogism type by locating common elements. This type of decision about which strategy to adopt is considered by Anderson in his rational analysis theory (see Singley & Anderson, 1989, p. 275).

With respect to Conclusion RTs, the results were consistent with those predicted by the ACT* account. There did not appear to be any differences between Conclusion RTs in the four Training conditions.

Given the results of the combined analyses of the Training data, the ACT* Transfer predictions for Premise RTs need to be slightly revised. As originally predicted, the Blocked and Highlight Training conditions appear to
result in the execution of fewer productions than is the case with Random Training. However, subjects with Blocked and Highlight Training are assumed to lack well-practised productions for identifying syllogism type and so should be ill-equipped for random presentation. This was supported by the Transfer results reported in Experiments 1 and 3. Thus these subjects should perform slower during Random Transfer than subjects with Random Training. In contrast, Alternating Training now appears to be no different than Random Training in terms of the number of productions executed. In fact, Alternating Training may result in a similar set of productions to those developed with Random Training. As a result subjects with Alternating Training should be no different to subjects with Random Training during Random Transfer, in contrast to the original prediction. The Transfer predictions for Conclusion RTs remain the same because all of the Training predictions were supported. That is, there should be no differences in Conclusion RTs between the Training conditions during Random Transfer.

3.5.3.2 Transfer

A number of ANOVAs were conducted to analyse the Transfer data. A 4 (Training condition) x 2 (Transfer block) x 2 (syllogism type) ANOVA was used to analyse Premise RTs, whereas two 4 (Training condition) x 2 (Transfer block) x 2 (syllogism type) x 2 (conclusion type) ANOVAs were used to analyse Conclusion RTs and Accuracy. These analyses indicated that there were no overall effects of Training condition in Premise RTs (F(3,124) < 1), Conclusion RTs (F(3,124) < 1) or Accuracy (F(3,124) = 1.05, p>0.05). A series of one-tailed comparisons were made between the Transfer performance of Random Trained subjects and the performance of Blocked and Highlight Trained subjects. The first set of comparisons examined data from all of the Transfer trials and revealed that there were no differences
between these conditions (Random = 2169 ms vs. Blocked = 2285 ms, \( F(1,62) < 1 \); Random vs. Highlight = 2558 ms, \( F(1,62) = 1.88, p>0.05 \)). However, considering only the first block of Transfer trials, the difference between Random Trained subjects (2142 ms) and Highlight Trained subjects (2638 ms) was significant (\( F(1,62) = 3.17, p<0.05, \) one-tailed), but there was still no difference between Random and Blocked (2282 msecs) Trained subjects (\( F(1,62) < 1 \)). There were no significant differences between Random Trained subjects and Alternating Trained subjects in either the mean Premise RTs of the first block of Transfer (\( F(1,62) = 1.84, p>0.05 \)) or the overall means of the Transfer phase (\( F(1,62) = 1.41, p>0.05 \)).

The lack of any Transfer effects with Conclusion RTs supports the predictions of the ACT* account. So too does the observation that the Premise RTs of Alternating Trained subjects during Random Transfer were no different from those of Random Trained subjects. However, the remaining comparisons provided, at best, only weak support for the ACT* predictions. The Transfer Premise RTs of Blocked and Highlight Trained subjects were not markedly slower than those of Random Trained subjects, although the observed differences were consistent with the predicted effects.

It is interesting that the Transfer differences that were predicted by the ACT* account were not observed whereas the predicted Training differences were. One possible reason for this is that the Transfer comparisons had less statistical power than the Training comparisons. The comparisons made between Training means were based on 64 subjects per condition, whereas the Transfer comparisons were based on only 32 subjects per condition. This explanation is supported by the observation that the differences observed were at least in the predicted directions, and that the within-subject measure of Transfer reported in Experiments 1 and 3 did show the predicted Transfer
effects. It is possible then that thirty-two subjects may not provide sufficient statistical power with the task used in this study to detect between-subjects differences associated with the contextual interference effect. However, the spirit of the contextual interference effect is still apparent in the results of Experiment 1: Blocked Trained subjects performed faster during Training than Random Trained subjects but were slowed considerably when presentation was changed to a random order.

There is an alternate explanation for the failure to replicate in this study the Transfer differences associated with the contextual interference effect. This account suggests that there are both costs and benefits associated with Blocked Training in terms of later performance with Random Transfer items. The ACT* account describes the costs of this transition. These concern the lack of well-practised productions for identifying task type which are essential for efficient performance with random presentation. The effect of this cost was evident in the within-subject measure of transfer, that is, when the Transfer performance of Blocked Trained subjects was significantly slower than was predicted by extrapolating their Training performance. However, the results also indicated that, although these subjects were slowed significantly in comparison to their previous performance, their performance did not differ during Random Transfer from that of subjects with Random Training. So Blocked Training may enable the development of a skill which, when applied during Random presentation, overcomes the lack of other essential skills. The nature of the skill which provides such a benefit is open to speculation.

The most important finding of the combined analyses is that Transfer differences were associated with Training differences, at least with respect to the Transfer effects indicated by the within-subjects measure of Transfer reported in the earlier sections of this chapter. Thus Blocked and Highlight
Trained subjects were shown to be faster than Random Trained subjects during Training, but were also the only subjects to be slowed markedly by subsequent random presentation during Transfer. Alternating Trained subjects on the other hand were not affected by this change in stimulus condition. The implications of this finding for the ACT* and intratrial processing accounts of the contextual interference effect will be considered in the next section.

3.6 Conclusions

One of the major aims of Experiment 1 was to replicate the contextual interference effect with the syllogistic reasoning task. Although the previously reported form of this effect was not replicated, results consistent with the effect were observed. Certainly performance during Blocked Training was faster than during Random Training.

The second major aim of Experiment 1 was to contrast two accounts of the contextual interference effect: one based on the ACT* theory of skill acquisition and transfer (Anderson, 1987, 1989b), the other termed the intratrial processing account (Carlson & Yaure, 1990). The ACT* account proposes that faster performance during Blocked Training results from the execution of fewer productions than during Random Training. Blocked presentation enables subjects to perform without the need to develop and apply productions that are necessary for performance with random presentation. However, not having developed these crucial productions, Blocked Trained subjects will be forced to develop them when faced with random presentation, resulting in slower performance. In contrast, the intratrial processing account proposes that Random Training provides subjects with practice at loading solution methods into working memory, whereas Blocked Training provides very little of this sort of practice. As a
result Blocked Trained subjects will not be as well-practised at this component of the task as Random Trained subjects and so will exhibit slower performance with random presentation. Experiments 2 and 3 were designed to contrast further these two accounts, as was the combined analysis of results from all three experiments.

Most of the predictions based on the ACT* account were supported by the data. The ACT* account predicted that there would be Premise RT differences between the Training conditions on the basis that these Training conditions would result in different processing strategies. The fact that no such differences were observed with Conclusion RTs during Training was also predicted by the ACT* account.

According to the ACT* account, the Training differences were caused by the execution of differing numbers of productions in the different Training conditions. The ACT* theory of skill acquisition and transfer proposes that transfer between tasks is determined by the extent to which productions developed with one task can be applied with another task. On the basis of this proposal, the ACT* account predicted that the degree of transfer from the various Training conditions to Random Transfer items would be a function of the number of productions each Training condition had in common with the Random Training condition. The number of common productions was indicated by the Training differences - Blocked and Highlight Trained subjects were faster than Random Trained subjects and so could have been expected to have fewer common productions with the Random Trained subjects than Alternating Trained subjects, who were not different in Premise RTs from Random Trained subjects. Consistent with this, transfer from Alternating Training to Random Transfer was more complete than from either Blocked or Highlight Training. Furthermore, there were no differences in
Conclusion RTs between the Training conditions, suggesting that each condition resulted in the same set of productions, and predicting the complete transfer of all conditions to Random Transfer. Thus the ACT* account was able to predict both the Training and Transfer effects, and also to associate these with one theoretical mechanism - the transfer of identical productions.

The intratrial processing account was not able to account for as many findings as the ACT* account. The Training differences observed with Premise RTs could not be explained by this account, nor could the fact that no such differences were observed with Conclusion RTs. The observation that Highlight Trained subjects did not exhibit complete transfer to Random Transfer items was not predicted by this account. Furthermore, the intratrial processing account was unable to predict that in some conditions there was complete transfer with Conclusion RTs but not with Premise RTs, or that these conditions were those that exhibited Training differences. However, this account did predict the complete transfer from Alternating Training to Random Transfer.

The ACT* theory appears to account for more of the data in Experiments 1, 2 and 3 than the intratrial processing account. Certainly there are many results that cannot be explained by the intratrial processing account but which can be explained by the ACT* account. Furthermore, the reverse is not the case. Therefore the ACT* theory provides a sufficient explanation for most of the results, making the intratrial processing account redundant.

With respect to the contextual interference effect, the present results suggest that the ACT* account provides a much more parsimonious explanation than one based on intratrial processing. Both Training and Transfer performance rely on the execution of production rules. Therefore the development,
execution and transfer of productions can account for both Training and Transfer differences. In contrast, practice at loading solution methods into working memory can only account for Transfer differences. However, at this stage the intratrial processing account cannot be abandoned totally. The intervening-task condition examined by Carlson and Yaure (1990) had subjects perform unrelated tasks in between each trial of a blocked presentation condition. Although this condition did not reduce the Training advantage of these subjects compared to Random Trained subjects, it did reduce the disadvantage during Random Transfer so that these subjects performed as well as Random Trained subjects. Although the Transfer task in this experiment was a slightly different type of task to the Training task, it is still difficult to account for this result in terms of the transfer of identical productions. Therefore intratrial processing appears to play some role in contextual interference effects. Further experimentation will be required to determine how this role compares to that of production transfer. Considering the results of the experiments reported in this chapter, it is likely to be a much smaller role.
PART 2  The Shape of Learning Functions following Transfer

Chapter 4  Development of a model describing changes in the shape of learning functions following partial transfer

<table>
<thead>
<tr>
<th>Chapter 4</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Introduction</td>
<td>131</td>
</tr>
<tr>
<td>4.2 Anderson (1982) and Learning Rate</td>
<td>133</td>
</tr>
<tr>
<td>4.3 Old and New Task Components</td>
<td>135</td>
</tr>
<tr>
<td>4.4 Evidence in Support of Equation 6</td>
<td>146</td>
</tr>
<tr>
<td>4.4.1 Singley and Anderson (1985)</td>
<td>146</td>
</tr>
<tr>
<td>4.4.2 Duncan (1977)</td>
<td>150</td>
</tr>
<tr>
<td>4.4.3 Fitts (1964)</td>
<td>156</td>
</tr>
<tr>
<td>4.4.4 Woltz (1988)</td>
<td>159</td>
</tr>
<tr>
<td>4.4.5 MacKay (1982)</td>
<td>163</td>
</tr>
<tr>
<td>4.4.6 Snyder and Pronko (1952)</td>
<td>164</td>
</tr>
<tr>
<td>4.5 Summary and Conclusions</td>
<td>167</td>
</tr>
</tbody>
</table>
Chapter 4

4.1 Introduction

The first half of this thesis was concerned with evaluating the ACT* theory in terms of how well it could account for the contextual interference effect. The main hypotheses of ACT* enabled predictions of the extent of transfer in certain situations. These predictions were based on assumptions of the effect of presentation conditions on performance strategies. The various performance strategies were viewed in terms of the productions that underlie them. Transfer between tasks was then predicted on the basis of the extent to which tasks shared common productions. The results of Experiments 1-3 were shown to be consistent with these predictions. This suggested that ACT* provided an accurate account of the contextual interference effect and a useful account of the processes underlying skill acquisition. The second half of the thesis is concerned with more dynamic aspects of the ACT* theory - the shape of learning functions.

As described in Chapter 1, the reduction in performance time that is associated with learning can be described by a power function of the amount of practice. This is the power law of learning. Transfer of training was evaluated with respect to such functions in Experiments 1-3. The most general form of such functions is given by the following equation:

\[ T = X + N P^c \]

where \( T \) represents the time to perform the task, \( P \) represents the amount of practice on the task, \( X \) represents performance time at asymptote, \( N \) represents a constant relating to the task (\( N \) is proportional to the number of processing steps involved in the task), \( X + N \) represents the performance time on Trial 1, and \( c \) represents the rate of learning. The value of \( c \) is a constant for any particular situation and is negative, most commonly in the range -
$1 < c < 0$ (c can be positive when the dependent variable increases in value with learning, e.g., accuracy).

According to Anderson (1982) the value of $X$ is usually very small relative to $X + N$. In addition power functions approach asymptote slowly. Anderson suggests that it is for these reasons that power functions with zero asymptote (i.e., $X = 0$) can provide very good fits to practice data. The practice data in Experiments 1-3 were accounted for well by power functions with zero asymptotes. Furthermore, the degree of fit was not improved significantly by incorporating non-zero asymptotes. However, Anderson makes the point that when practice involves a sufficient number of trials, careful analysis of data will reveal evidence of non-zero asymptotes. This issue will be discussed further below.

Although the value of $c$ is usually found to be between -1 and 0 there appears to be substantial variation within this range between tasks and subjects (e.g., Newell & Rosenbloom, 1981). However a systematic relationship between type of task and subjects' learning rate has yet to be established. The aim of this chapter is to develop a model that provides an account of this relationship. At present, Anderson (1982) has provided the only account of the determinants of learning rate. This account basically suggests that, all things being equal, learning rate is a constant for each person. Given the fact that subjects exhibit varying learning rates dependent on task situation, this account is obviously not sufficient. However Anderson's account will be used as a starting point for the development of a more precise model. This model assumes that learning rate is affected by the relative amounts of practice of particular task components and also the relative number of processing steps involved in these components. Thus this model is in fact designed to account for the shape of learning functions following transfer, but will account for
changes in learning rate as a matter of course. In the remainder of this chapter, Anderson's account of learning rate will be described, followed by a description of the new model. Predictions that follow from this model will be compared with previously published data and then formally evaluated in experiments to be described in the following chapters.

4.2 Anderson (1982) and Learning Rate

As described in Chapter 1, ACT* predicts that performance will become faster with practice as a result of both algorithmic improvement and strengthening. ACT* states that the process of compilation improves the procedure for performing a task by reducing the number of steps involved in the procedure. It is assumed that the number of steps is reduced by a constant fraction with each improvement. Anderson (1982) suggests that this algorithmic improvement follows a standard power function for improvement to an asymptote:

\[ N = N^* + N_0 P^{-f} \]  

(1)

In this equation \( N \) represents the number of processing steps performed on Trial \( P \) (where \( P \) represents the amount of practice), \( N^* + N_0 \) represents the number of steps performed on Trial 1, and \( N^* \), the asymptote of this function, represents the minimum number of steps that constitute the optimal procedure for performing the task. The exponent \( f \) represents the constant fraction by which the number of steps is reduced with each improvement.

The other contributing factor to performance time reduction in the ACT* theory is that of strengthening. In ACT* it is assumed that the representation of task-specific information is strengthened in memory with practice. The strength of a memory element then determines the speed at which it can be accessed and applied. Anderson (1982) shows that this reduction in
application time with increased strength can also be described by a power function:

\[ T = C + A P^{-g} \]  

(2)

In this equation \( T \) represents the time to execute a series of productions, \( P \) represents the amount of practice, \( A \) is a constant that represents the time it takes to execute a certain number of productions and is therefore proportional to the number of productions involved in performing a task, the asymptote \( C \) represents the minimum time for execution of a certain number of productions, \( C + A \) represents the maximum time for execution, and the exponent \( g \) is a constant that represents the rate at which the strength of a memory element decays, and is a value greater than zero.

In order to derive a function that describes total time to perform a task Anderson (1982) combined Equation 1, which represented the number of productions, and Equation 2, which represents the time per production. This combination resulted in Equation 3:

\[ T_T = (N^* + N_0 P^{-f})(C + A P^{-g}) \]  

(3)

This equation simplifies to Equation 4:

\[ T_T = N_0 A P^{-(f+g)} \]  

(4)

if \( N^* \) and \( C \) are assumed to be zero. Anderson makes this assumption because Equation 3 is not a perfect power function, although it is a good approximation to one. Eliminating \( N^* \) and \( C \) results in a pure power function which can be further simplified to the equation introduced above as the general form of power function that describes learning:

\[ T = N P^c \]  

(5)

As described above, \( N \) is a constant related to the initial number of productions involved in performing the task and \( P \) is the amount of practice. This equation has a zero asymptote.
The step from Equation 4 to Equation 5 suggests that the learning rate for any task (c) is a negative constant that is determined by two other constants: the fraction by which the number of productions is reduced with compilation (f), and the decay rate of production strength (g). The latter two constants, according to Anderson (1982), are parameters of the cognitive system. Therefore learning rate is itself a parameter of this system. This implies that learning rate is determined by particular characteristics of a subject's mental functioning, that it is a constant which is 'built in' to each subject. Thus, in Anderson's account, there is no apparent facility for type of task to affect learning rate.

If learning rate is some form of constant for each subject, then presumably the range of learning rates that are observed between experiments is a result of different combinations of results from subjects with different learning rates. However this suggestion is not supported by experiments where learning rate was observed to vary from task to task within individuals and groups (e.g., Fitts, 1964; Grose & Damos, 1988; MacKay, 1982; Newell & Rosenbloom, 1981; Smith & Sussman, 1969; Snyder & Pronko, 1952). Therefore it does not appear that variation in observed learning rates results only from a sampling distribution of combinations of various constant learning rates. Variation in learning rate is also likely to result from an interaction between subject characteristics (i.e., their intrinsic learning rate) and task features. The form of this interaction will be considered below.

4.3 Old and New Task Components

The form of the interaction between a constant learning rate and task features could involve the fact that, except for infants, most tasks have components that involve previously learned skills as well as components that are peculiar
to the new task (Fitts & Posner, 1967, p. 19). The latter components will include both components for performing the new part of the task and components that integrate the functioning of the old and new components. This conception suggests that for any task there will be some components that have had more practice than others. The obvious consequences of such a suggestion are that (1) the older components will be faster than the new components (if the number of steps involved in the two sets of components is equivalent), and (2) the older components will have less room for improvement than the new components. These predictions in turn suggest further consequences. However, they also rely on a number of assumptions which require evaluation.

The first assumption underlying the above predictions is that for any one person the rate at which new skills are learned is a constant. This is as suggested by Anderson's (1982) conception of the power law of learning. A 'new' skill here is conceived of as a skill that involves no components which have had previous practice. This includes those skills that are necessary for integrating the functioning of old and new skills into the same goal structure. The second assumption is that all components underlying the performance of a task will improve according to the power law of learning and, with practice, will continue to do so at the same learning rate.

In summary then, the above conception suggests that, in most cases, learning a new task involves continued practice on old skills. These are skills that have been practised in the context of some other task. Learning a new task will also involve the learning of new skills. These are skills that are required to fill the gap between the repertoire of old skills possessed by the trainee and the skills necessary to perform the new task. These new skills will involve both task-related skills and skills for integrating the functioning of old and new task-
related skills. In order to evaluate the predictions based upon this concept of task learning it will be informative to contrast this concept with an idealised situation where all components of a task are learned from scratch.

The idealised situation is a simple one - improvement is a function of practice and follows the power law of learning. The initial time to perform the task is a function of the number of components or steps involved in the task. Thus the whole situation can be described by one power function. In contrast, the more realistic situation is unlikely to be accurately described by one power function. The simple reason is that in this situation each component will not have had equal amounts of practice. As a result a power function with one term that describes amount of practice is not sufficient. This then raises the question of how varying amounts of practice can be incorporated into a function that describes improvement on a task with old and new components.

One possibility is suggested by Anderson, Conrad and Corbett (1989) who propose that "acquisition of...skill can be predicted by composing simple learning functions for (the) units (p. 503)" underlying tasks. One interpretation of this proposal is that it is suggesting that components of a task have their own learning functions and that the learning function for the task as a whole is a combination of these separate functions. With respect to the current discussion, this would suggest that old and new components of a task improve with practice according to their own learning functions. These separate functions would then include the fact that the components have had unequal amounts of practice. The learning function for the task would then be a combination of these "old" and "new" learning functions.

The form of this combination needs to be considered before the implications of this suggestion can be examined. Underlying a great deal of the research
into skill acquisition is the assumption that the more steps involved in a task the longer it takes to perform (e.g., Anderson, 1982; Carlson et al., 1989; Staszewski, 1988). This assumption also was the basis of many of the predictions examined in Experiments 1-3. The assumption implies a serial process where each step contributes a particular amount of time to the total task time. Following this logic with the combination of old and new skills requires the combination to be a serial one. That is, the processing of one set of components should not impinge upon the processing of the other set except to provide input information. If this is the case then the learning function for the whole task should be a simple linear combination of power functions describing improvement in each of the underlying components. If the components can be separated into old and new then this function will have the following form:

\[ T_{\text{task}} = T_{\text{old}} + T_{\text{new}} = N_0 P_0^c + N_n P_n^c. \] (6)

This equation represents the linear combination of two power functions of the form described in Equation 5. Terms with the subscript "o" represent parameters of the old components of the task, and terms with the subscript "n" represent parameters of the new components.

There are a number of implications of Equation 6 that should be made explicit. The first is that the contribution of each set of components to the total task time is weighted by the number of steps involved in each set. That is, the greater the number of steps in a set of components, the greater will be the contribution of this set. The second implication is that this weighted combination will be qualified by the amount of practice that the sets of components have had prior to the combination. This qualification has two related forms: (1) the more practice a set of components have had, the faster they will be, and so practice serves to reduce the contribution of a set of
components to the overall performance time of a task; (2) as the amount of practice of a set of components increases the room for improvement decreases.

The most important implication of Equation 6 concerns the rate at which improvement will occur in the total task. In this equation the learning rate of the two separate power functions (c) is the same in each function. This represents the assumption described above that the learning of all components of a task for any one person is a constant. Incorporating this assumption into the equation results in a power function describing improvement in the overall task that has a different learning rate to that of each of the components. This difference is always in the direction of a reduction: the learning rate of the total task will be slower than the learning rate of its underlying components. The amount by which the learning rate will be reduced is a function of the relative number of steps in the old and new components, and of the relative amount of practice each set of components had prior to combination.

For example, consider the case of a subject who has practised a task for 6 sessions. Let the task have 100 steps (N = 100) and the learning rate be -0.8 (c = -0.8). The improvement in the time to perform the task can now be described by the equation T=100 P^{-0.8} (this is only a loose description as N in Equation 5 is only proportional to the number of processing steps/productions involved in a task, not equal to this number). Now suppose the subject is given a new task to practise that includes all of the steps in the old task plus a new set of steps that number 20. The subject will be able to perform the old steps quickly but will be starting from scratch with the new steps. The time to perform such a task that includes old and new components can be described by the combination of the power functions that would describe improvement on the separate components. Thus,
\[ T = T_{\text{old}} + T_{\text{new}} \]

\[ = 100 \, P_{\text{old}}^{-0.8} + 20 \, P_{\text{new}}^{-0.8} \]

where \( P_{\text{old}} = P_{\text{new}} + 6 \)

This function now has an overall learning rate of \( c = -0.44 \) (i.e., plot values for \( T \) against \( P_{\text{new}} \) on log-log axes and the gradient of the resulting straight line is \(-0.44\)). Therefore, learning rate has been attenuated as a result of combining two skills that differ in the amount of practice they have had and the number of steps involved with their execution. The rate of improvement in the overall task is slower than in the components underlying performance in the task. However the attenuation will not always be as dramatic as in this example. As mentioned above, the amount of attenuation is moderated by two factors.

The first moderating factor on the amount of attenuation in learning rate is the relative number of steps in the old and new components of a task. The effect of this factor on the overall learning rate is depicted in Figure 4.1. The data points were generated from the above example, where \( P_{\text{old}} = P_{\text{new}} + 6 \) and \( c = -0.8 \). The number of steps in the new component was varied from 0 to 500 and the number of steps involved in the old component was kept constant at 100. It is clear from this figure that an increase in the ratio of old to new steps increases the attenuation of the learning rate. The function depicted in Figure 4.1 has minimum and maximum boundaries. The minimum boundary corresponds to the situation where there are no old steps involved in the task. The learning rate in this situation corresponds to the intrinsic learning rate of the system, which is \(-0.80\). The maximum boundary represents the maximum attenuation effect. This corresponds to the situation where the old steps
Figure 4.1: Learning rate as a function of the ratio of the number of steps in old skills vs. the number of steps in new skills. The data points were generated from a learning example described in the text. The line represents an interpolation of the data points.
outnumber the new steps to the extent that the new steps have no effect on the overall learning rate.

The point of maximum attenuation is equivalent to measuring the learning rate of a task as if it were a new task and ignoring the fact that it has been practised for 6 sessions. The subsequent sessions of practice would elicit performance times that improved at what appeared to be a slower rate than the earlier sessions. However this attenuation is simply a result of using an inappropriate point to represent 'session one'. Figure 4.2 illustrates that in this situation a practice function with a 'slow' learning rate is in fact the tail end of a practice function with a faster learning rate. This phenomenon can result in inaccurate measures of learning rate when prior experience with a task is not taken into account. Newell and Rosenbloom (1981) suggest that using a more general form of the power function that incorporates the amount of prior practice can improve the accuracy of power function descriptions of practice data. Such a function would have the following form:

\[ T = N (P + E)^c \]  

(7)

This equation is the same as Equation 5 except that the term which represents the amount of practice on a task is now divided between practice that is observed (P) and practice prior to observation (E). This form of the power function has been shown to provide a better fit to some practice data but is also no better than simpler functions (Equation 5) with other data (Singley & Anderson, 1985). It appears that the reason why Equation 7 is no better than Equation 5 for some practice data is that the assumption of prior experience with a task is too general an assumption. As demonstrated above prior experience may only apply to some components of a task. Other components will not have had any practice. Hence Equation 6 may be a more accurate depiction of some situations.
Figure 4.2: Demonstration that a practice function with a 'slow' learning rate may be the tail-end of a function with a faster rate. The last six data points of the fast curve (learning rate = -0.80) have been displaced 6 practice units. The new curve now has the slower learning rate of -0.239.
The second moderating factor on the attenuation of learning rate is the relative amount of practice that the old and new skills underlying a task had prior to their combination. Figure 4.3 illustrates the effect on learning rate of increasing the amount of prior practice of old skills in the example introduced above. All of the data points in this figure were derived from the example situation with the number of old steps constant at 100 and the number of new steps constant at 20. The learning rates were calculated for the learning functions that resulted from varying the amount of practice of old skills in Equation 6. Figure 4.3 shows the result of varying $P_{old}$ from 0 to 100 sessions. When the old skills have had no practice prior to combination the learning rate is simply the intrinsic learning rate of the system (-0.80). This corresponds to the situation where all components of the task are new. As the amount of practice of old skills increases the attenuating effect on learning rate increases until the combination of old and new skills has its maximum effect. In this situation, when old skills have been practised for 8 sessions prior to the combination of old and new skills, the resulting learning rate is at its slowest at -0.439. Beyond this point, increasing the amount of practice old skills have prior to combination has a diminishing effect on learning rate. This diminishing effect continues until the old skills have had so much practice that any further practice results in only negligible improvement. At this point the combination of old and new skills has no effect on the overall learning rate. The learning rate for the overall task will now be completely determined by the rate at which performance on the new skills improves, and this will be at the intrinsic learning rate of the system (-0.80 in this situation).

In summary, Equation 6 leads to the prediction that when a task involves old and new components, this task will be learned at a slower rate than that at which each of the two sets of components improves (i.e., the constant intrinsic rate of each person). The amount by which this learning rate will be
Figure 4.3: Learning rate as a function of the amount of extra practice of old skills in comparison to new skills. The maximum attenuation of learning rate occurs when $P_{\text{old}} = 8$ and results in a learning rate $= -0.439$. The data points were generated from a learning example described in the text. The line represents an interpolation of the data points.
attenuated will be moderated by the relative number of steps between old and new components of the task, and by the amount of practice that the old skills had prior to learning the new task.

Some evidence exists in the research literature to support these predictions. However, the evidence should only be considered indirect as it relies on a fairly liberal view as to which components of a task are old and which are new. This evidence is presented below.

4.4 Evidence in Support of Equation 6

4.4.1 Singley and Anderson (1985)

Singley and Anderson (1985) examined the extent of transfer between different computer text-editing programs. Two basic types of editors were used: (1) line editors, where only one line of text in a file can be viewed at a time and editing is on a line-by-line basis; and (2) screen editors, where the screen is filled with the contents of a file and users are able to designate the location to be edited by moving around the screen with a cursor. Subjects in this study were trained to operate two line editors (ED and EDT) and one screen editor (EMACS). Singley and Anderson were interested in the extent to which training on one versus two line editors would transfer to performance on a screen editor. They found that there was positive transfer both between the line editors and from the line editors to the screen editor. These results were interpreted as suggesting that all of the editors shared a certain number of productions necessary for their performance. Thus training with one editor provided the subjects with a set of productions, of which some were useful for operating another editor. In more detail, there was almost complete transfer between the line editors which suggests that these
editors share a large number of productions. In contrast there was only partial transfer between the line editors and the screen editor, suggesting that the productions developed with the line editors were not totally sufficient for operating the screen editor. The development of more productions would have been necessary for efficient performance with this editor.

The most interesting result of this experiment for the current discussion concerns the differences in the shape of the practice functions during training and transfer. Simple power functions were fit to the data by Singley and Anderson and the equations of these functions are presented in Table 4.1. The design of the experiment was such that the subjects who performed in the Transfer phase with ED were trained with EDT and vice-versa. The subjects who performed in the Transfer phase with EMACS were subjects who had either been trained with one of the line editors or both, the total amount of training being equal for both groups of subjects. The equations in the Training column of Table 4.1 were derived from data from control subjects who practised with only one editor.

It is clear from the equations in Table 4.1 that in all cases the learning rate during Transfer was slower than the rate during Training. Furthermore, the amount of attenuation was greater when transfer was between line editors (-0.53 to -0.20 and -0.79 to -0.34) than when transfer was from line editors to the screen editor (-0.55 to -0.41). These results are as would be predicted on the basis of Equation 6. First consider the line editors. The fact that there was almost total transfer between these editors suggests that they share a large number of productions. Therefore when subjects switch to one line editor after operating with the other line editor, very few new productions need to be developed. In other words, when these subjects operate the new line editor, underlying their performance will be a large number of old skills and a small

<table>
<thead>
<tr>
<th>Editor</th>
<th>Training</th>
<th>Transfer</th>
</tr>
</thead>
<tbody>
<tr>
<td>ED</td>
<td>$T = 4.3 \times 10^{-0.53}$</td>
<td>$T = 3.7 \times 10^{-0.20}$</td>
</tr>
<tr>
<td>EDT</td>
<td>$T = 4.8 \times 10^{-0.79}$</td>
<td>$T = 3.9 \times 10^{-0.34}$</td>
</tr>
<tr>
<td>EMACS</td>
<td>$T = 3.9 \times 10^{-0.55}$</td>
<td>$T = 3.4 \times 10^{-0.41}$</td>
</tr>
</tbody>
</table>
number of new skills. As was shown in Figure 4.1, this is the type of situation where attenuation of learning rate is very large. In contrast, transfer from line editors to screen editors involves fewer shared productions. Thus subjects who trained with line editors and then switched to the screen editor would be able to use some of their previously developed productions but would also need to develop a relatively large set of new productions. As a result the ratio of old to new skills in this situation would be smaller than in the situation where transfer was between line editors. As shown in Figure 4.1, this predicts a smaller attenuation of learning rate in the former than in the latter situation.

A similar interpretation of the Transfer results was suggested by Singley and Anderson. Initially they attempted to account for the different practice functions that were observed in Training and Transfer with Newell & Rosenbloom’s (1981) general power function (i.e., Equation 7 above). However they concluded that the simple notion of including prior experience into the power function was not sufficient to account for the Transfer data. Singley and Anderson then suggested that in such transfer situations it was necessary to identify components of tasks that were general and specific. General components are those that are shared between tasks. Specific components are those that are peculiar to the particular task. Singley and Anderson then proposed that a more appropriate account of their transfer data would involve a power function that included two separate power functions, one for the general components and one for the specific components:

$$T = X + N_g P_g^c + N_s P_s^c$$  (8)

This equation is equivalent in form to Equation 6 with an asymptote. Furthermore equating general components with old components and specific components with new components results in the two equations being equivalent in function as well. Unfortunately Singley and Anderson did not
evaluate whether Equation 8 provided a better account of their transfer data than Equation 7. Furthermore, there was no explicit discussion of the implications of this conceptualisation of transfer on learning rates.

The changes in learning rate that were observed in this study are consistent with the predictions made on the basis of Equation 6. However the fact that the comparisons of learning rates were between subjects weakens the degree of support to some extent. Despite this however, the fact that the amount of attenuation in different conditions was in the predicted direction is encouraging.

4.4.2 Duncan (1977)

Duncan (1977) investigated how various S-R rules affected performance time in spatial choice reaction tasks. Duncan was interested in determining whether the 'obviousness' of S-R mappings could account for the speed at which responses were elicited. The basic task used in this experiment involved training subjects with various S-R mappings. On each trial one of four horizontally arranged lights would light up and the subject was required to press one of four buttons located directly below the lights. The particular button that a subject was to press was determined by the particular S-R mapping taught to the subject. There were four different S-R mappings used in this experiment (see Figure 4.4) representing different degrees of 'obviousness'. The Pure-Corresponding (P-C) condition was assumed to involve the most obvious mapping - when each light was lit the button directly below that light was to be pressed. The next most obvious condition was assumed to be the Pure-Opposite (P-O) condition. This mapping involved pressing the button that was "opposite" the stimulus that was lit. Implicit in this mapping is the rule "Give the response that is the mirror-
Figure 4.4: S-R mappings in the four conditions used in the experiment reported in Duncan, J. (1977). Response selection rules in spatial choice reaction tasks. In S. Domic (Ed.). Attention and Performance VI. Hillsdale NJ: Erlbaum.
opposite to the stimulus". The two Mixed conditions (M-1 and M-2), which contained two Corresponding mappings and two Opposite mappings, were assumed to be less obvious than the Pure conditions.

The main conclusions that Duncan reached on the basis of the experiment's results were that subjects did not appear to be performing on the basis of individual S-R associations. Instead it appeared that subjects were following a rule or system of rules. In addition, it seemed that performance was faster with the more obvious rules. For instance, Corresponding S-R mappings were faster than Opposite S-R mappings. This suggests that in the Pure conditions subjects could respond on the basis of one rule in each condition.

In the Pure-Corresponding condition subjects could follow the simple rule: "Press the button below the light that is on". In the Pure-Opposite condition the subjects could follow the slightly more complex rule: "Press the button that is the mirror-opposite of the light that is on". This of course would require further processing to determine which button is the mirror-opposite and so explains the extra time in this condition.

In the Mixed conditions subjects would apparently be required to develop a strategy that is a hybrid of the two rules described above. Although Duncan did not discuss the form of such a strategy, one plausible strategy has as the first step in either condition a determination of which stimuli has been presented. This would then determine which of the above rules would apply. Thus, in the Mixed-1 condition, this strategy would have the following form:

IF stimulus = 1 or 4
THEN Apply Opposite Rule

IF stimulus = 2 or 3
THEN Apply Corresponding Rule
Thus subjects in this condition could apply the same rules as subjects in the Pure conditions but would also need to apply rules that determined which of the Corresponding or the Opposite rule was appropriate on each trial.

The relevance of these conclusions to the present discussion concerns the rate at which subjects improved on the four different S-R mappings. The common ground shared by Duncan's experiment and the discussion of old versus new skills concerns the definition of "obviousness". Duncan defined "obvious" S-R mappings as those that arise from long hours of practice. Thus the Corresponding and the Opposite rules are obvious because subjects presumably have had prior experience applying them in other contexts. In contrast, the rules that subjects in the Mixed conditions needed to apply were less obvious because it is unlikely that subjects would have had much experience applying them in any context. Therefore the conditions of this experiment can be re-conceptualised as comprising various combinations of old and new skills. The Pure conditions can be considered to be comprised of old skills (either the Corresponding rule or the Opposite rule) only, whereas underlying the Mixed conditions are old skills (both the Corresponding and the Opposite rules) and new skills (rules for determining which of the Corresponding or Opposite rules to apply on each trial). Applying Equation 6 to this situation leads to two predictions: (1) The Pure conditions involve old skills only and so the ratio of old to new skills is very high. As illustrated in Figure 4.1 this results in maximal attenuation of learning rate. Thus learning rates in these conditions will be slower than the intrinsic learning rate of the subjects. (2) The Mixed conditions involve both old and new skills and so the ratio of old to new skills is lower than in the Pure conditions. Using Figure 4.1 as a guide, this predicts that learning rate will be attenuated to a lesser extent than in the Pure conditions. Combining these two predictions leads to
the further prediction that learning in the Mixed conditions will be faster than in the Pure conditions.

Figure 4.5 shows a re-plotting of the practice data reported by Duncan (1977) on log-log axes. The lines represent the best-fit power functions of the form described by Equation 5. The equations for these lines are presented in Table 4.2. The data points labelled as Mixed-Corresponding and Mixed-Opposite refer to those trials in the two Mixed conditions which involved a Corresponding and an Opposite S-R mapping respectively. Thus the data in this figure labelled Corresponding refers to those conditions where the Corresponding rule was applied, and the Pure versus Mixed labels refer to the contexts in which these trials occurred. The same convention applies to the Opposite data points.

The important feature to note about Figure 4.5 is that the rate of improvement in the Mixed conditions (-0.20 and -0.21) is faster than in the Pure conditions (-0.08 and -0.11). Therefore the predictions based on Equation 6 are consistent with the results of this experiment. However, this support for the equation must again be qualified to some extent by the fact that different subjects made up the four conditions in this experiment. It is interesting to note though that the learning rates of subjects in the Pure conditions were almost equal, as were the learning rates of subjects in the Mixed conditions. Therefore there was not a significant degree of variation between groups of subjects performing in similar conditions and the only obvious variation in learning rate was as predicted between the Pure and Mixed conditions. Despite this qualification then, it is again encouraging that the equation can predict differences in learning rate of certain tasks on the basis of particular characteristics of the tasks.
Figure 4.5: Mean RTs reported in Duncan, J. (1977). Response selection rules in spatial choice reaction tasks. In S. Domic (Ed.). Attention and Performance VI. Hillsdale, NJ: Erlbaum. Data points have been plotted on log-log axes. The lines represent the best-fit power functions. The equations for these lines are presented in Table 4.2.


<table>
<thead>
<tr>
<th>Condition</th>
<th>Equation</th>
<th>$r^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pure-Corresponding</td>
<td>$T = 481.34 , P^{-0.08}$, $r^2 = 0.974$</td>
<td></td>
</tr>
<tr>
<td>Pure-Opposite</td>
<td>$T = 561.04 , P^{-0.11}$, $r^2 = 0.924$</td>
<td></td>
</tr>
<tr>
<td>Mixed-Corresponding</td>
<td>$T = 685.45 , P^{-0.20}$, $r^2 = 0.980$</td>
<td></td>
</tr>
<tr>
<td>Mixed-Opposite</td>
<td>$T = 783.35 , P^{-0.21}$, $r^2 = 0.959$</td>
<td></td>
</tr>
</tbody>
</table>
4.4.3 Fitts (1964)

An unpublished experiment by Fitts and Switzer and reported by Fitts (1964, pp. 262-265) used a similar S-R paradigm to that used by Duncan (1977). Fitts and Switzer manipulated the obviousness or compatibility of stimuli and their associated responses. In this experiment pictures of common objects were presented to subjects and they were to make a vocal response to these pictures. Two variables were manipulated: (1) the number of alternative stimuli (12 vs. 3), and (2) the directness of the verbal mediation between stimulus and response. Subjects were required to respond with a letter of the alphabet to each of the stimuli. In the direct mediation case (Familiar condition) the response was the first letter in the familiar name of the object. For example, if a picture of a bird was presented, then "B" was the appropriate response. In the indirect mediation case (Unfamiliar condition) the response was one of the other 11 letters in the response set (determined randomly). Thus in the case of the bird picture being presented, and a picture of a chair being part of the same stimulus set, the appropriate response could have been "C", if this was the S-R pair that was taught to the subjects. Subjects practised the appropriate pairings to a criterion of two correct trials and then were tested in five sessions in which reaction time measures were taken.

Fitts (1964) makes the point that in both of the Familiar and Unfamiliar conditions the appropriate responses are not responses that subjects would usually make to the stimuli. However, there is no doubt that the two conditions differ in the obviousness of the appropriate responses. Fitts suggests that the responses are "hooked-up" to existing S-R pairs through associative chains. He proposes that in the Familiar condition this type of chain would involve the following steps: Stimulus -> Familiar object name ->
Vocalisation of first letter of name. In contrast the Unfamiliar condition is assumed to involve an extra step: Stimulus -> Familiar object name -> Some other name -> Vocalisation of the first letter of the other name. Fitts reports that following the testing sessions subjects reported that their performance relied on such associative chains.

The associative chains that Fitts describes suggest that in both the Familiar and Unfamiliar conditions performance of the task relies on both old and new skills. The old skills involved in both conditions relate to the processing of the stimulus object and accessing the familiar name for that object. The new skills involved in the Familiar condition relate to the processing of the familiar name of the object in order to determine the first letter. In the Unfamiliar condition the new skills are similar to the new skills in the Familiar condition in that a name is processed to determine the first letter. An additional set of new skills is involved in this condition which relate to the accessing of some other name which provides the bridge between the familiar object name and the appropriate response. Therefore both conditions will involve old and new skills but in the Unfamiliar condition the ratio of old to new skills will be less than in the Familiar condition. This situation should result in a difference in learning rates between the two conditions. An inspection of Figure 4.1 reveals that the degree of attenuation resulting from the combination of old and new skills will be less in the Unfamiliar condition. Therefore subjects in this condition should learn at a faster rate.

The test sessions data reported by Fitts (1964) have been re-plotted in Figure 4.6. The lines represent the best-fit power functions of the form described by Equation 5. The equations for these lines are presented in Table 4.3. The important result to note from this figure is that learning rate was faster in the Unfamiliar conditions (-0.13 and -0.13) than in the Familiar conditions (-0.08
Figure 4.6: Data taken from an unpublished study by Fitts and Switzer, reported in Fitts, P.M. (1964). Perceptual-motor learning. In A.W. Melton (Ed.). Categories of human learning. New York: Academic Press. Data points have been plotted on linear axes. The lines represent the best-fit power functions. The equations of these lines are presented in Table 4.3.


<table>
<thead>
<tr>
<th>Type</th>
<th>Equation</th>
<th>r²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large-Unfamiliar</td>
<td>( T = 780.01 P^{-0.13} )</td>
<td>0.978</td>
</tr>
<tr>
<td>Large-Familiar</td>
<td>( T = 684.23 P^{-0.08} )</td>
<td>0.994</td>
</tr>
<tr>
<td>Small-Unfamiliar</td>
<td>( T = 684.88 P^{-0.13} )</td>
<td>0.989</td>
</tr>
<tr>
<td>Small-Familiar</td>
<td>( T = 577.10 P^{-0.05} )</td>
<td>0.938</td>
</tr>
</tbody>
</table>
and -0.05). The direction of these differences is consistent with the predictions made on the basis of Equation 6. These differences are not large and this is probably related to the small differences in the ratio of old to new skills in the two conditions. Incorporating the additional step in the associative chain involved in the Unfamiliar condition does not constitute a large increase in the amount of processing necessary before a response can be made. The general rule of thumb used in skills analysis is that one production takes 100 msecs to apply (Carlson & Schneider, 1989). In the initial stages of the testing sessions the extra step involved in the Unfamiliar condition increased reaction times by approximately 100 msecs. This suggests that the extra step involved very few new productions, probably only one or two. It is surprising then that differences in learning rate were detected at all.

4.4.4 Woltz (1988)

As part of a larger study Woltz (1988) trained subjects on a task that involved the application of sequential rules. These rules involved evaluating stimulus conditions in order to provide an appropriate response. Woltz assumed that in the early stages of practice subjects would retrieve complete rule sequences and so maintain a considerable working memory load. Woltz manipulated the attention demands of this task by using two different orders in which the rules were to be applied. In the high attention demand condition those rules that were complex and so presented a greater load on working memory were to be applied early in the task. As a result subjects were likely to suffer a considerable load on working memory. This is because the temporary storage of the complete sequence of rules and the processing of the complex rules would be competing for working memory space. In contrast, there would not be such competition for working memory resources in the low attention demand condition. In this condition the complex rules were applied late in the
task. As a result rules in the early stages of the task need not be maintained in working memory at this point and this would release more of working memory for the application of the complex rules.

Woltz compared the subjects' efficiency on the individual components of the task with performance on the task as a whole. He found that in the low attention demand condition component efficiency skill was a good predictor of performance in the overall task. That is, subjects who were in the upper quartile of efficiency were faster than subjects in the lower quartile. This was also the case in the high attention demand condition. However, the performance advantage of efficient component execution was less in this condition. Woltz interpreted this result as suggesting that higher level cognitive processes, such as attention-related skills, may be more important than efficient componential skills when attention demands are increased.

Woltz's view suggests that in the low attention demand condition performance reflects the operation of task-related skills only, whereas in the high attention demand condition performance reflects both the operation of task-related skills and skills related to the directing of attention. A plausible assumption that can be made is that the attention-related skills are task-independent skills and therefore are likely to have had extensive practice compared to the more task-specific skills. In other words the attention-related skills could be considered "old" and the task-related skills "new". Hence this is another situation where learning rate predictions can be made on the basis of Equation 6.

If the high attention demand condition involves old and new skills and the low attention demand condition involves only new skills then learning rate will be faster in the latter condition. However it is unlikely that the difference
in learning rate will be substantial. The only old skills involved in the high attention demand condition are likely to be well-practised. Subjects will have had extensive experience allocating attention and working memory resources, certainly much more experience than with the task-related skills (e.g., evaluating whether a number is odd or even and then determining whether it is at the top or bottom of the screen). As was illustrated in Figure 4.3, the more practice that old skills have had prior to the combination with new skills, the less will be the amount of attenuation of learning rate. Figure 4.7 shows a re-plotting of the data reported by Woltz (1988, Figure 4). The lines represent the best-fit power functions of the form described by Equation 5 and appear to be good approximations to those included in Woltz's figure. Woltz concluded from his analyses that there was a significant difference in the learning rates of subjects in the two conditions. The direction of this difference was as predicted above: subjects in the low attention demand condition learned at a faster rate (-0.29) than subjects in the high attention demand condition (-0.26), although this difference was indeed small.

Woltz also predicted that there would be a learning rate difference between the two conditions. He proposed that increasing the attention demands of the task would limit the ability of subjects to proceduralise task instructions and thus inhibit the rate of skill acquisition. Although this conception led to the correct prediction in this case, it would lead to the same prediction regardless of the extent of subjects' experience in allocating working memory resources. In contrast, Equation 6 allows more specific predictions of the amount of learning rate attenuation, given that certain parameters (e.g., relative amounts of practice and number of steps involved in old and new components) are known.
Figure 4.7: Data reported in Woltz, D.J. (1988). An investigation of the role of working memory in procedural skill acquisition. *Journal of Experimental Psychology: General, 117*, 319-331. Data points have been plotted on log-log axes. The lines represent the best-fit power functions with the following equations:

**High Demand:** $T = 2.96p^{-0.26}, r^2 = 0.967$

**Low Demand:** $T = 2.80p^{-0.29}, r^2 = 0.989$. 
4.4.5 MacKay (1982)

MacKay (1982) reported data combined from two earlier studies (MacKay & Bowman, 1969; MacKay, 1981) in which the speed of reading sentences aloud was measured. Three types of sentences were examined: normal, scrambled and nonsense. The scrambled sentences were derived from the normal sentences by re-arranging the order of the words. The nonsense sentences were derived in turn from the scrambled sentences by substituting or rearranging letters in words to form pronounceable non-words.

MacKay reported two main results. The first was that normal sentences were read faster than scrambled sentences, which were read faster than nonsense strings. The second result was that the rate at which subjects improved their reading speed with practice was a function of the type of sentence being read. Subjects improved at the fastest rate with nonsense strings, at a slower rate with scrambled sentences, and at the slowest rate with normal sentences.

The difference in learning rates can be accounted for by considering, at a fairly superficial level, the skills that underlie reading aloud the three types of sentences. First consider the normal sentences. Competent readers rely on skills that (1) convert familiar words into sounds, and (2) use syntactically based meaning to increase fluency (e.g., noun phrases such as "the dog" are read faster than "the" and "dog" read separately) (e.g., Just & Carpenter, 1980). With adult readers these skills are unlikely to improve much with practice. Hence any improvement observed with this type of sentence will be at a very slow rate.

Now consider the scrambled sentences. Subjects who read such sentences aloud now perform without the benefit of the second type of skill involved in
normal reading. That is, because the scrambled sentences did not contain familiar noun or verb phrases, words can only be processed as individual meaning units rather than as part of higher level concepts. Therefore with this type of sentence, subjects would be required to develop some other strategy to increase fluency beyond the level at which they can read a list of unrelated words. This suggests that reading these sentences results in the combination of old and new skills. The ratio of old to new skills with these sentences will be smaller than with the normal sentences and so the attenuation of learning rate will be less than with the normal sentences. Thus learning will be faster with the scrambled sentences than with the normal sentences.

With the nonsense sentences subjects are at even less of an advantage in terms of being able to use already established skills. Because these sentences were derived from the scrambled sentences there is no syntactically based meaning to increase fluency. In addition the sentences contain no familiar words and so pronunciations are not as easily accessible as with real words. Instead subjects must rely on pronunciations derived from the pronunciations of real words that look similar. This derivation process is unlikely to be a well-practised skill. Without many other old skills to influence performance with these sentences, it seems plausible that the ratio of old to new skills in this condition is lower than both of the previous conditions. Therefore the combination of old and new skills in reading nonsense sentences will have little attenuating effect on learning rate and so subjects will improve with these sentences faster than with the other types.

4.4.6 Snyder and Pronko (1952)

It appears that Equation 6 can also provide a good account of the acquisition of perceptual-motor skills. Snyder and Pronko (1952) investigated the effect
of reverse lenses on performance in a motor task. For 15 days subjects were trained with normal vision on the Purdue pegboard task - a task that requires precise visual control of motor responses. Following this training phase, the subjects were then given 27 days practice on the same task whilst wearing reversing lenses. This latter condition had the obvious effect of slowing performance compared to the normal vision condition.

A more interesting result concerns the rate at which performance improved in the two conditions. The practice data from this experiment has been re-plotted in Figure 4.8. The lines represent the best-fit power functions of the form described by Equation 5. It is clear from this figure that, although performance with reversed vision never attains the speed of that with normal vision, the performance in the former condition improves at a faster rate (-0.20) than in the latter condition (-0.10).

The differences in learning rate can be accounted for by assuming that performance in the reversed vision condition involves the same skills as in the normal vision condition plus a new set of skills that cope with the reversed visual information. Using the same logic as used to account for the Fitts (1964) results, performance in the normal vision condition can be considered to involve both old and new skills. The old skills are concerned with coordinating visual information and motor responses. The new skills are more task related and concern rules specifying peg placement. In the same vein, the reversed vision condition involves the same old and new skills plus another set of skills concerned with translating the reversed visual information into normal orientation information. This means that subjects can adapt existing skills to operate in the new environment rather than having to develop a whole set of new skills. It also implies though that more new skills underly performance in the reversed vision condition than in the normal vision.
Figure 4.8: Data reported in Snyder, F.W. & Pronko, H.H. (1952). *Vision with spatial inversion*, Wichita, Kansas: University of Wichita Press. Data points have been plotted on linear axes. The lines represent the best-fit power functions with the following equations:

**Normal Vision**: \( T = 72.418 \times 0.10 \), \( r^2 = 0.913 \);

**Reverse Vision**: \( T = 111.98 \times 0.20 \), \( r^2 = 0.935 \).
condition. The fact that the performance on the first day of reversed vision was almost 60% slower than on the first day of normal vision suggests that the skills to be learned in order to cope with the reversed vision substantially outnumber the skills needed to place the pegs. Thus the ratio of old to new skills in the reversed vision condition is smaller than in the normal vision condition. The prediction can now be made on the basis of Equation 6 that the combination of old and new skills in this task will result in a greater attenuation of learning rate in the normal vision condition (see Figure 4.1). This then accounts for Snyder and Pronko's finding that performance with reversed vision improved at a faster rate than with normal vision.

4.5 Summary and Conclusions

The studies discussed above, when considered as a whole, provide consistent support for the notion that the combination of old and new skills can affect the rate at which a task is learned. The way in which learning rate is affected depends on the extent to which old components have been practised prior to combination with the new components, and the relative number of steps involved in the old and new components. One of the studies discussed above (Singley & Anderson, 1985) showed a slowing of learning rate from Training to Transfer. Another study (Snyder & Pronko, 1952) showed an increase in the learning rate from Training to Transfer. The remaining four studies did not involve transfer designs but did show differences in learning rates between conditions which were accounted for by considering the relative amounts of practice and number of steps involved in the underlying components.

The evidence is, by no means, conclusive. In only a limited number of cases were within-subject comparisons of learning rates available (Snyder &
Furthermore, the identification of old and new skills was based on subjective criteria. Similarly, assumptions concerning the relative number of steps involved in the underlying components and the amount of practice each set of components had prior to combination were also decided on a qualitative basis. However, the degree of support in the analysis of these studies is encouraging and suggests that further investigations are warranted. In the following chapters more direct experimental evidence is described. The next chapter describes Experiments 4, 5 and 6 which were designed to evaluate the effect on learning rate of combining old and new skills. Similar stimulus materials and methodology to Experiments 1, 2 and 3 of this thesis were used in these experiments.

Slightly different versions of the equations discussed above were evaluated in the experiments described in the next chapter. Strictly speaking, Equation 6 should include a term(s) describing asymptotic performance. If functions describing old and new components are combined then this suggests that two asymptotes will also be combined. That is,

\[ T_{\text{task}} = T_{\text{old}} + T_{\text{new}} = X_0 + N_0 P_0^c + X_n + N_n P_n^c \]

However, Anderson (1989b) suggests that the asymptote of a learning function represents "a constant associated with perceptual and motor processes" (p. 529). Presumably this refers to the minimum time required to process stimulus information and execute a response. If all tasks are subject to this same performance limit then this suggests that \( X_0 \) and \( X_n \) should refer to the same constant. This assumes that ultimately all productions that underlie performance on the task can be composed into one production. If this assumption is valid then the equation for \( T_{\text{task}} \) should only include one
constant representing asymptotic performance. Therefore, given this assumption, a more appropriate version of Equation 9 is:

\[ T_{\text{task}} = X + N_0 P_0^c + N_n P_n^c \]  

(10)

where \( X \) represents the asymptotic performance level that the whole task shares with the various components underlying this task. A similar argument can be made for the inclusion of an asymptote in Equation 7. Hence,

\[ T = X + N (E + P)^c \]  

(11)

In the experiments to be described in the following chapters, the ability of Equations 10 and 11 to describe transfer performance will be compared. Equation 10 will be referred to as the "Old/New Equation" because it is designed to describe performance following the combination of old and new skills. Equation 11 will be referred to as the "Old Equation" because it is designed to describe performance that results from continued improvement of old skills only. Thus performance predicted by this equation will be equivalent to extrapolating performance observed during Training. In other words, the Old/New Equation is designed to describe partial transfer, whereas the Old Equation is designed to describe complete transfer.
Chapter 5  Experiments 4, 5 and 6

5.1 Introduction

5.2 Experiment 4
   5.2.1 Introduction
   5.2.2 Method
      5.2.2.1 Subjects
      5.2.2.2 Materials
      5.2.2.3 Design
      5.2.2.4 Apparatus and Procedure
   5.2.3 Results and Discussion
      5.2.3.1 General Results
         5.2.3.1.1 Training
         5.2.3.1.2 Transfer
         5.2.3.1.3 Summary
      5.2.3.2 Learning Functions
      5.2.3.3 Summary and Conclusions

5.3 Experiment 5
   5.3.1 Introduction
   5.3.2 Method
      5.3.2.1 Subjects
      5.3.2.2 Materials and Design
      5.3.2.3 Apparatus and Procedure
   5.3.3 Results and Discussion
   5.3.4 Summary and Conclusions

5.4 Experiment 6
   5.4.1 Introduction
   5.4.2 Method
      5.4.2.1 Subjects
      5.4.2.2 Materials
      5.4.2.3 Design
      5.4.2.4 Apparatus and Procedure
   5.4.3 Results and Discussion
      5.4.3.1 General Trends
         5.4.3.1.1 Training
         5.4.3.1.2 Transfer
      5.4.3.2 Learning Functions

5.5 An Additional Parameter in the Old/New Equation
5.1 Introduction

One of the main conclusions that was drawn from the results of the first three experiments was that certain features of a training environment can induce particular types of processing strategies. The results of Experiment 3 demonstrated that highlighting selected features of syllogisms (i.e., the common elements of premises) would encourage subjects to develop strategies which involved fewer processing steps than if these features had not been highlighted. Furthermore these strategies were shown to be inefficient when the highlighting was removed.

Experiments 1, 2 and 3 were only concerned with the extent of transfer in particular situations. They were not concerned with how subjects would cope with environments where their strategies were deficient. It is clear though, that in such environments, subjects find it necessary to develop new skills. For instance, in order to solve the syllogisms in the Transfer phase of Experiment 3, the subjects found it necessary to learn to locate common elements without the benefit of highlighted elements.

The ability of subjects to cope in a new environment is determined by the number of old skills they have that are appropriate for the new environment and their ability to develop additional skills required by the new situation. For example, subjects who are trained to solve syllogisms with capitalised common elements develop a set of skills that rely on this feature. If these subjects are then given syllogisms to solve that do not have capitalised common elements the subjects will have to develop additional skills in order to function efficiently. The skills that were developed during training that are appropriate for the new task can be considered 'old'. The additional skills that are required for efficient performance with the new task are referred to as
'new'. Thus performance of the new task will involve the combination of old and new skills.

The fact that transfer in the above situation involves the combination of old and new skills suggests that an application of the Old/New Equation may be appropriate. However, as mentioned in the previous chapter, the Old/New Equation only applies when the combination of old and new skills is a serial one. In order to evaluate whether this is the case with the syllogism example it is necessary to consider the particular functions of the old and new skills alluded to above.

In the introduction to Experiment 3 (§ 3.4.1) a perceptual strategy was suggested as the most likely form of strategy that would be developed to solve syllogisms with capitalised common elements. This strategy would be centred on identifying syllogism form and would involve subjects perceiving the direction of the diagonal of common elements in the premise pairs highlighted by capitalisation. In contrast, when highlighting is removed, subjects would need to identify syllogism form by performing a search and match process that compares the four elements in each premise pair to determine which are common. Once this has been achieved, the syllogism could be solved as it was in the presence of highlighting. Therefore the new skill of locating common elements is actually providing information to the old productions that was originally provided by the highlighting feature. Thus the combination of old and new skills underlying performance of the new task is a serial one where the output of the new skills is used as input for the old skills. Thus this situation is appropriate for analysis using the Old/New Equation.
Predictions of learning rate attenuation based on the Old/New Equation involve considerations of the ratio of old to new skills in a new task, and the relative amounts of practice each set of skills have had prior to their combination. The exact values of these quantities for the above situation are not known. However, plausible estimates can be made which will be sufficient to derive a testable prediction.

The new productions that are necessary to solve syllogisms without capitalised common elements will not outnumber the old productions that are useful in this new situation. Figure 3.1 illustrates that locating the common elements will be a relatively small part of the processing of premise pairs. The results of Experiment 3 support this suggestion. Figure 3.11 shows that when Highlight Trained subjects processed premises without highlighting, their performance time increased by approximately 1000 ms. Using the rule of thumb described in the previous chapter, this suggests that subjects needed to develop approximately ten new productions to overcome the removal of the highlighting feature. This is small compared to the 60-70 productions that subjects developed in the initial stages of Training with the capitalised common elements (i.e., mean Premise RT in the first block of Highlight Training was approximately 6100 ms). Therefore the combination of old and new skills in the solution of syllogisms without highlighting is one where the old skills greatly outnumber the new skills.

The amount of practice that subjects were given in the Training phase of Experiment 3 is small in skill acquisition terms. Thousands of trials of practice are necessary to develop skills to the point where further practice provides little benefit (Anderson, 1989b; Schneider, 1985). Therefore the performance of subjects in Experiment 3 would still have been improving at
the end of the Training trials. As a result improvement in the Transfer phase would be due to the combined improvements of old and new skills.

In summary, the above situation is one where old skills have been practised to the point where there is still room for substantial improvement, and the ratio of old to new skills is high. Figures 4.1 and 4.3 show that this situation is one where attenuation of learning rate is considerable. Therefore, on the basis of the Old/New Equation, the prediction can be made that in a situation similar to that of Experiment 3, the rate at which subjects will improve during the Transfer phase will be slower than during the Training phase.

It is not possible to evaluate this prediction with the results of Experiment 3. As there were only two blocks of practice during the Transfer phase of this experiment, an accurate measure of the learning rate could not be made. The experiments described in this chapter were designed to overcome this difficulty. Experiment 4 involved a similar design to Experiment 3, but with Training and Transfer phases of equal length. Given the results of Experiment 3, the transition from Highlight Training to Random Transfer in Experiment 4 was considered to involve a combination of old and new productions. Therefore this experiment was designed as an explicit test of the predictions based on the Old/New Equation described in Chapter 4. Experiment 5 was designed to extend the results of Experiment 4 by observing the performance of two subjects with more extensive practice at the syllogism task. Experiment 6 involved a similar design to Experiment 2 where Alternating Training was examined. The Training and Transfer phases of this experiment were also of equal length. Given the results of Experiment 2, the transition from Alternating Training to Random Transfer in Experiment 6 was not considered to involve a combination of old and new productions. Thus
Experiment 6 was designed as a contrast to Experiment 4, with Transfer performance expected to be described by the Old Equation.

5.2 Experiment 4

5.2.1 Introduction

Experiment 4 was designed as an explicit test of the main prediction that can be derived from the Old/New Equation, that the combination of old and new productions will result in an attenuation of learning rate. The design of Experiment 4 was based on that of Experiment 3 and extended in two ways. First, the Transfer phase was extended to include as many trials as in the Training phase. This would provide a more reliable measure of learning rate in the Transfer phase than was possible in Experiment 3. Second, a control condition was used. Both the experimental group and the control group were presented with syllogisms during Training with the common elements of the premises highlighted. During Transfer, the control group again solved syllogisms with the highlighting feature. In contrast, the experimental group was presented with items without this feature.

A number of predictions can be made concerning performance in both the Training and Transfer phases of this experiment. The first prediction concerns performance during Training. There should not be any substantial differences between the groups, in either Premise or Conclusion RTs because both groups will be training with the same stimuli. The remaining predictions concern performance during the Transfer phase of the experiment.

The purpose of the highlighting manipulation in this experiment was the same as in Experiment 3. The experimental group were assumed to rely on the
highlighting feature during Training and so be ill-equipped for the Transfer situation where this cue was no longer present. As a result this group should develop new productions during Transfer to complement their old productions. Therefore the experimental group should be substantially slowed in Premise RTs because of the new productions they now have to execute (as was found in Experiment 3). In addition, according to the Old/New Equation, the learning function of this group during Transfer should improve at a slower rate than was observed during Training. However, this effect should only be evident with Premise RTs. The productions developed to process conclusions during Training are assumed to be adequate for this task during Transfer. Therefore Conclusion RTs should continue to improve according to the learning function observed during Training. In other words, the Conclusion RTs of the experimental group during Transfer should be described by the Old Equation. For similar reasons, the Premise and Conclusion RTs of the control group during Transfer should be described by extrapolating the learning function observed with these measures during Training. Therefore the Transfer performance of the control group should be accounted for by the Old Equation.

An important methodological point should be made at this stage. This point concerns determining whether Transfer performance reveals a learning rate attenuation as described by the Old/New Equation, or the continuation of a previously observed learning equation as described by the Old Equation. The latter situation will be observed as a learning function, during Transfer, that also has a slower learning rate than that observed during Training. However, as stated in Chapter 4, this attenuation is simply a result of using an inappropriate point to represent 'session one' in plotting the learning function and then determining learning rate. As was demonstrated in the previous chapter, though, the learning rate attenuation that results from the combination
of old and new productions is in fact less than that which is associated with further continuation of old productions. Despite this though, because both of these situations result in an attenuation of learning rate, a reduction in learning rate was not considered sufficient support for the model of Transfer described by the Old/New Equation. Therefore the degree to which Transfer performance can be accounted for by the various equations will also be considered.

5.2.2 Method

5.2.2.1 Subjects

Forty volunteers from the University of Western Australia first year Psychology course participated in this experiment for course credit or $5. Eight subjects failed to reach the learning criterion of an error rate not in excess of 25% in the last half of the Training phase. The data from these subjects were excluded from further analysis, leaving 32 subjects, 16 per condition.

5.2.2.2 Materials

The items used during the Training phase of this experiment were identical to those used in Experiment 3. That is, the common elements in the premise pairs of these items were presented in upper case to highlight them in comparison to the lower case uncommon elements.

In the Transfer phase of this experiment the first 96 items were identical to those used during Transfer in Experiment 3. An additional 192 items of this form were constructed by recombining the elements in the first 96 Training
items. The new combinations of elements followed the same constraints described in Experiment 1 concerning the plausibility and implications of the resulting syllogisms. In the Experimental condition of this experiment all of the elements in the premise pairs of these Transfer items were presented in lower case. In the Control condition the common elements of these items were presented in upper case.

5.2.2.3 Design

Two groups of subjects were used in this experiment. In the experimental condition subjects were presented with Training items that contained capitalised common elements and then Transfer items with all elements in lower case. In the control condition subjects were presented with items that contained capitalised common elements during both the Training and Transfer phases. For both conditions, the order of presentation of ABBC and BCAB problems during the Training and Transfer phases was random, with the same constraints as described in Experiment 1 for Random Training.

5.2.2.4 Apparatus and Procedure

These were identical to those in Experiment 1 except in two respects. Firstly, subjects in this experiment were presented with 288 Transfer items instead of 96 and so experimental sessions often lasted for longer than 60 minutes. Secondly, as a result of the greater number of trials, more rest periods were considered necessary. In this experiment the computer automatically signalled a one-minute rest period every 48 trials, and when the minute had elapsed, it automatically presented the next trial. The message that was presented by the computer to signal the rest period was the same in each period. Thus the
transition from the Training phase to the Transfer phase of the experiment was not announced, in contrast to the previous experiments.

5.2.3 Results and Discussion

The results of this experiment will be presented in two sections. The first section will be largely descriptive, presenting the gross patterns in the results. In contrast, the second section will evaluate the predictions made earlier regarding learning functions.

5.2.3.1 General Results

5.2.3.1.1 Training

A number of analyses of variance were performed on the Training data. A 2 (Training condition) x 6 (Training block) x 2 (syllogism type) ANOVA was used to analyse Premise RTs. Two 2 (Training condition) x 6 (Training block) x 2 (syllogism type) x 2 (conclusion type) ANOVAs were used to analyse Conclusion RTs and Accuracy.

The mean Premise and Conclusion RTs in the Training and Transfer phases of the experiment are presented in Figure 5.1. It is clear from this figure that the experimental and control groups did not differ substantially during Training in either Premise or Conclusion RTs. The analyses of variance support this observation. The difference between the mean Premise RTs of the two groups (experimental group = 2966 ms vs. control group = 2732 ms) was not significant (F(1,30) < 1), nor was the difference between the mean Conclusion RTs (experimental group = 1136 ms vs. control group = 933 ms, F(1,30) = 3.02, p>0.05). The difference in mean Accuracy was also not
Figure 5.1: Mean Premise and Conclusion RTs for both Experimental and Control groups, during Training and Transfer phases of Experiment 4.
significant (experimental group = 91.36% vs. control group = 91.69%, F(1,30) < 1).

The fact that the two groups did not differ substantially during Training means that unqualified comparisons can be made between the groups with respect to performance changes as a result of the different Transfer conditions.

Overall, subjects showed significant improvement in all performance variables during Training. Premise RTs improved from 5974 ms in Block 1 to 1365 ms in Block 6 (F(5,150) = 85.75, p<0.05). Conclusion RTs were reduced from 1858 ms in Block 1 to 729 ms in Block 6 (F(5,150) = 58.33, p<0.05), and Accuracy was increased from 75.20% in Block 1 to 98.11% in Block 6 (F(5,150) = 29.46, p<0.05).

Premises were studied for less time with ABBC syllogisms (2741 ms) than with BCAB syllogisms (2957 ms) (F(1,30) = 33.25, p<0.05). Subjects were also more accurate with ABBC problems (ABBC = 92.45% vs. BCAB = 90.60%, F(1,30) = 4.33, p<0.05). There was no effect of syllogism type on Conclusion RTs (F(1,30) < 1). The syllogism-type effect on Premise RTs and Accuracy appeared to diminish with practice. In Block 1 ABBC syllogisms were responded to 399 ms faster and 7.42% more accurately than BCAB syllogisms. This advantage was reduced to 172 ms and -0.13% in Block 6. Both of these interactions were significant (Premise RTs: F(5,150) = 2.28, p<0.05; Accuracy: F(5,150) = 3.27, p<0.05). These results replicate those of Experiment 3 and suggest that subjects in both conditions were processing premises with a bias in favour of ABBC syllogisms.
There was a significant effect of conclusion type on reaction time. True conclusions were responded to faster than False conclusions (962 ms vs. 1107 ms, F(1,30) = 47.11, p<0.05). They were also responded to more accurately (True = 92.58% vs. False = 90.47%; F(1,30) = 4.33, p<0.05). These results replicate those found in Experiments 1, 2 and 3 and again suggest that subjects developed a processing bias in favour of True conclusions.

5.2.3.1.2 Transfer

Two analyses were conducted to examine the effects of removing the highlighting feature in premise pairs. The first analysis compared performance in the last block of Training with the first block of Transfer. The second analysis was restricted to performance throughout the Transfer phase.

In the first analysis a number of analyses of variance were performed. A 2 (Training condition) x 2 (session) x 2 (syllogism type) ANOVA was used to analyse Premise RTs. Two 2 (Training condition) x 2 (session) x 2 (syllogism type) x 2 (conclusion type) ANOVAs were used to analyse Conclusion RTs and Accuracy. The most interesting result of this first analysis concerns the interaction between condition (Experimental vs. Control) and block (last block of Training vs. first block of Transfer). In all performance variables this interaction was significant (Premise RTs: F(1,30) = 38.25, p<0.05; Conclusion RTs: F(1,30) = 9.40, p<0.05; Accuracy: F(1,30) = 4.85, p<0.05). In all cases this interaction was such that the performance of the experimental group declined from Training to Transfer (i.e., Premise and Conclusion RTs became slower and Accuracy was reduced), whereas the performance of the control group improved (see Figure 5.1). However, a post-hoc test using Tukey's WSD procedure revealed that
the only significant change in performance was that the experimental group increased the time spent studying premise pairs from 1499 ms in the last block of Training to 2520 ms in the first block of Transfer ($F(1,30) = 31.27, p<0.05$).

The second analysis extends the results of the first analysis into the whole of the Transfer phase. A number of analyses of variance were performed on the Transfer data. A $2 \times 6 \times 2$ ANOVA was used to analyse Premise RTs. Two $2 \times 6 \times 2 \times 2$ ANOVAs were used to analyse Conclusion RTs and Accuracy.

Figure 5.1 shows that when premise pairs no longer contained capitalised common elements, the experimental group increased the time spent studying premises by a considerable amount, both in comparison to their previous performance with capitalised common elements, and to the continued performance of the control group. On average, the experimental group spent 2290 ms studying premises during the Transfer phase, whereas the control group spent only 1024 ms ($F(1,30) = 20.60, p<0.05$). In contrast, the change from Training to Transfer had no apparent effect on the Conclusion RTs of the experimental group, as these did not differ significantly from those of the control group (experimental group = 771 ms vs. control group = 653 ms, $F(1,30) = 2.38, p>0.05$). The Accuracy of the experimental group was similarly unaffected by this change (experimental group = 97.59% vs. control group = 96.81%, $F(1,30) < 1$).

The fact that Premise RTs of the experimental group were affected by the removal of the highlighting of common elements replicates a similar result in Experiment 3. The additional fact that no other performance variable was
affected by this change provides further support for the assumptions concerning strategies that would be developed with this task and the transfer predictions these assumptions enabled. These included the suggestion that subjects would rely on the capitalised common elements in premise pairs to determine syllogism type. When this highlighting feature was removed, subjects would find it necessary to develop new productions to identify syllogism type. Therefore it seems reasonable to conclude from the results of this experiment that subjects in the experimental group had to develop new productions to cope with the new task in the Transfer phase. Furthermore, these new productions would be concerned with the processing of premises only. Hence the selective effect on Premise RTs.

The overall performance of the two groups improved with practice during the Transfer phase. Premise RTs were reduced from 1821 ms in Block 1 to 1506 ms in Block 6 (F(5,150) = 8.48, p<0.05). Conclusion RTs also became shorter from Block 1 (764 ms) to Block 6 (654 ms) (F(5,150) = 5.16, p<0.05). However, there was no significant improvement in Accuracy (F(5,150) = 1.70, p>0.05).

There was a significant interaction between block number and condition in Premise RTs (F(5,150) = 2.31, p<0.05), with the difference between the experimental and the control groups diminishing with practice (difference at Block 1 = 1396 ms vs. difference at Block 6 = 1205 ms). This result suggests that the two groups improved at different rates during this phase. With the experimental group being the slower during the Transfer phase, and the difference diminishing with practice, the implication of this result is that the experimental group improved at a faster rate than the control group during this phase.
The differences in rate of improvement between the two groups is consistent with the predicted changes in learning rate. The control group carried on as if nothing had changed during this phase (as indeed it had not), improving at the same rate as during Transfer. That is, they improved according to the same learning function observed during Training. However, because performance during the Transfer phase is at the tail-end of the learning function for this group, improvement with practice would be minimal compared to that during Training. In contrast, the experimental group had to develop new productions to cope with the changes in the task. The productions that this group developed during Training that were appropriate for the new task would improve at the same rate as during Training. As with the control group, the amount by which these productions would improve would be small compared to that during Training. However, there would be substantial improvement in the new productions as these would be starting from scratch. The combination of these old and new productions would result in a set of productions that, on the whole, would be slower than the productions executed by the experimental group at the end of Training, but would have greater room for improvement with practice. Thus at the beginning of the Transfer phase the experimental group were slower than the control group. However, the reason for the slowing of performance - the new productions - resulted in more room for improvement for the experimental group. Hence this group improved at a faster rate than the control group during this phase.

5.2.3.1.3 Summary

The results of this experiment replicate and extend those of Experiment 3. Subjects were shown to develop processing strategies that included biases that favoured ABBC syllogisms over BCAB syllogisms, and True conclusions over False conclusions. In addition, subjects were shown to rely
on the capitalised common elements to solve the syllogisms as removal of this highlighting feature adversely affected performance. This effect resulted in slower performance throughout Transfer. This suggests that the effect was not simply a temporary shock that subjects would recover from quickly. Instead subjects had to develop new procedures for processing premise pairs, and as a result, performance could only improve with practice. This suggestion was given further support by the observation of a difference in learning rate between the two groups in Premise RTs. Furthermore this observation provides qualitative support for the predictions based on the Old/New Equation. A more quantitative evaluation of these predictions will be described in the next section.

5.2.3.2 Learning Functions

A number of analyses will be described in order to evaluate the predictions concerning learning functions. Two basic results were expected. (1) Where Transfer performance was assumed to rely on the same set of productions that were developed during Training, performance in the second phase would not deviate substantially from that expected on the basis of Training performance. In other words, extrapolating Training performance into the Transfer phase would provide a good account of Transfer performance. Thus the Old Equation should provide a good account of Transfer performance. This prediction is applicable to the processing of conclusions in both the Experimental and Control groups, and to the processing of premises in the Control group. In contrast, (2) when new productions are necessary to supplement old productions, Transfer performance will be different from that expected on the basis of Training performance. That is, extrapolating Training performance into the Transfer phase should not provide a good account of Transfer performance. This prediction is applicable to the Experimental
group's processing of premises. Thus the Old Equation should not provide a good account of Transfer performance in this situation, whereas the Old/New Equation should provide a superior account.

Figure 5.2 shows Premise and Conclusion RTs of the control group during Training and Transfer. The lines passing through the Training data are the best fit power functions of the form \( RT = X + NPC \). Non-zero asymptotes were included in the power functions of this experiment because accurate measures of learning rate were required. The parameters of these functions are presented in Table 5.1. The Training functions have been extrapolated into the Transfer phase. The Transfer data has been plotted with confidence limits (alpha = 0.05). Figure 5.3 presents the same information as Figure 5.2 for the experimental group. The third line in this figure is derived from an application of the Old/New Equation. The derivation will be described below.

An important feature to note about the functions fit to the data in this experiment is that performance time was plotted against Cumulative Accuracy, instead of against Block Number, as in all of the previous experiments. The reason for this modification is related to the ACT* account of the power law of learning. This account proposes that most of the improvement in performance time that results from practice is related to the strengthening of productions (Anderson, 1982). In ACT* productions are only strengthened as a result of successful application. Therefore plotting performance time against amount of successful practice (represented by Cumulative Accuracy, i.e., adding proportion correct over Blocks of practice) was considered to provide a more appropriate test of the predictions based on the Old/New Equation, which is itself based on the ACT* account of skill acquisition. (The major conclusions of this analysis of transfer were unaltered
Figure 5.2: Mean Premise and Conclusion RTs for the Control group during Training (Highlight) and Transfer (Highlight) phases of Experiment 4. Lines drawn through Training data are best-fit power functions (see Table 5.1 for equations). These lines have been extrapolated into the Transfer phase and so represent applications of the Old Equation in this phase. Error bars are confidence limits (alpha = 0.05). The abcissa represents a measure of successful practice which is calculated by adding proportion correct over blocks of practice.
### Table 5.1: Parameters of power functions fitted to Premise and Conclusion RTs during Training and Transfer phases of Experiment 4. Functions labelled "Observed" were fitted directly to the data. Functions labelled "Old Equation" were fitted to values extrapolated from Training performance. Functions labelled "Old/New Equation" were fitted to values predicted by various versions of the Old/New Equation (see text for details). $r^2$ = proportion of variance in the observed reaction time values accounted for by the predicted values. rmsd = root mean squared deviation between predicted and observed reaction time values.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$X$</th>
<th>$N$</th>
<th>$c$</th>
<th>$r^2$</th>
<th>rmsd</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experimental Group</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Premises</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training</td>
<td>0</td>
<td>5058.30</td>
<td>-0.703</td>
<td>0.993</td>
<td>152.841</td>
</tr>
<tr>
<td>Transfer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>0</td>
<td>2557.80</td>
<td>-0.105</td>
<td>0.901</td>
<td>48.367</td>
</tr>
<tr>
<td>Old Equation</td>
<td>0</td>
<td>1395.40</td>
<td>-0.220</td>
<td>0.937</td>
<td>1178.106</td>
</tr>
<tr>
<td>Old/New Equation</td>
<td>0</td>
<td>2517.20</td>
<td>-0.391</td>
<td>0.881</td>
<td>654.642</td>
</tr>
<tr>
<td><strong>Conclusions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training</td>
<td>600</td>
<td>1459.20</td>
<td>-1.209</td>
<td>0.999</td>
<td>237.663</td>
</tr>
<tr>
<td>Transfer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>0</td>
<td>860.81</td>
<td>-0.104</td>
<td>0.885</td>
<td>17.731</td>
</tr>
<tr>
<td>Old Equation</td>
<td>0</td>
<td>755.28</td>
<td>-0.059</td>
<td>0.883</td>
<td>68.790</td>
</tr>
<tr>
<td><strong>Control Group</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Premises</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training</td>
<td>0</td>
<td>5009.50</td>
<td>-0.835</td>
<td>0.979</td>
<td>267.335</td>
</tr>
<tr>
<td>Transfer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>0</td>
<td>1128.20</td>
<td>-0.093</td>
<td>0.700</td>
<td>37.193</td>
</tr>
<tr>
<td>Old Equation</td>
<td>0</td>
<td>1083.10</td>
<td>-0.260</td>
<td>0.707</td>
<td>212.357</td>
</tr>
<tr>
<td><strong>Conclusions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training</td>
<td>535</td>
<td>790.14</td>
<td>-0.993</td>
<td>0.999</td>
<td>12.315</td>
</tr>
<tr>
<td>Transfer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>0</td>
<td>675.62</td>
<td>-0.033</td>
<td>0.286</td>
<td>20.005</td>
</tr>
<tr>
<td>Old Equation</td>
<td>0</td>
<td>660.36</td>
<td>-0.047</td>
<td>0.309</td>
<td>31.940</td>
</tr>
</tbody>
</table>
by using Cumulative Accuracy as the measure of practice as opposed to Block Number which was used in the previous experiments.)

Table 5.1 presents the parameters for power functions that provide the best-fit to the Training and Transfer Premise and Conclusion RTs of both groups. Also included in this table are the parameters of curves fitted to points extrapolated from Training performance. In other words, the best-fit power functions for the Training data were extrapolated a further 6 blocks. This data was then plotted from Block 1 and power functions were then fitted to the resulting curves. This resulted in functions with slopes that were smaller than the slopes of the original Training functions. As mentioned in the previous chapter, this is a natural result of plotting data from inappropriate starting points and is the rationale behind incorporating previous experience in power functions (i.e., the Old Equation). The Transfer data was also plotted from Block 1 before the fitting of power functions. This was to allow suitable comparisons of learning rate with not only the extrapolated Training curves, but also the original Training data.

The most appropriate evaluation of the predictions based on the Old/New Equation should involve a test of how well the equation can account for Transfer data in comparison to alternative models (e.g., the Old Equation). Therefore the ability of each function to account for the Transfer data was assessed using $r^2$ (proportion of variance in observed data accounted for by predicted values) and rmsd (root mean squared deviation between observed and predicted values) as measures of goodness-of-fit. The larger the $r^2$ values and the smaller the rmsd values, the better the fit between predicted and observed performance.
The control group experienced the same stimulus conditions in both phases of the experiment. Therefore it was assumed that during Transfer these subjects would apply the same productions executed during Training. This was predicted to result in improvement during Transfer that would be described by versions of the Old Equation derived from learning functions observed during Training. Figure 5.2 shows that this prediction was supported by the results of this experiment. Premise and Conclusion RTs during Transfer were consistent with performance predicted on the basis of Training performance. Table 5.1 shows that fitting power functions to the observed Transfer data and to performance predicted by the Old Equation results in similar functions. These functions are similar in terms of their equations and therefore the degree to which they can account for the data. Certainly there is very little separating the equations describing predicted and observed Conclusion RTs during Transfer. However, the equations describing predicted and observed Premise RTs in this condition do differ slightly. Although the latter equation provides a better fit to the data, Figure 5.2 indicates that the Old Equation predicts values which closely approximate the observed values. Note that in both Premise and Conclusion RTs, the learning rate observed during Transfer is slower than observed during Training, but is at least of similar magnitude to that predicted by the Old Equation. Therefore, as predicted, when Transfer performance is assumed to rely on continued improvement of old productions, this performance will be described by extrapolating the learning functions that describe the initial learning of these old productions (i.e., the Old Equation).

The experimental group was assumed to perform during Transfer with a combination of old and new productions. The old productions are those developed during Training and the new productions are those developed during Transfer to cope with the new stimulus conditions during this phase.
Therefore Transfer performance was predicted to be described by application of the Old/New Equation, rather than by the Old Equation. This prediction only applied to Premise RTs, however. Conclusion RTs were predicted to be described by the Old Equation. Figure 5.3 shows that Conclusion RTs during Transfer were accounted for well by the Old Equation. This is supported by the fact that the power functions fitted to the predicted and observed Conclusion RTs are very similar and that the $r^2$ and rmsd values are also of similar magnitude (see Table 5.1). In contrast, the Transfer Premise RTs were far removed from the reaction times predicted by the Old Equation, as is indicated by the large rmsd value (1178 ms).

A version of the Old/New Equation was derived for this data set and compared with the observed Transfer Premise RTs of the experimental group. The Old/New Equation is supposed to account for improvement with practice in a situation where new productions are developed to fill the gap between a repertoire of old skills and those skills necessary to perform a task. Underlying part of the improvement with the new task will be continued improvement with the old skills. This improvement is described by extrapolating the Training function. In addition to this improvement is improvement of the new skills. Description of this improvement will require a new power function. As discussed in the text, this new power function will have a learning rate that is the same as the learning rate of the Training function. This new function will also have an intercept that reflects the number of new productions involved in the new task. In practice this intercept will reflect the time that the extra productions add to the initial overall performance time. This suggests that the intercept of this new power function can be approximated by subtracting the performance time for the first block of Transfer that is expected on the basis of Training (i.e., execution of old productions) from the observed performance time in Block 1 of the Transfer
Figure 5.3: Mean Premise and Conclusion RTs for the Experimental group during Training (Highlight) and Transfer (Random) phases of Experiment 4. Lines drawn through Training data are best-fit power functions (see Table 5.1 for equations). These lines have been extrapolated into the Transfer phase and therefore represent applications of the Old Equation in this phase. The third line is a specific version of the Old/New Equation derived from the data of this experiment (see text for details). Error bars are confidence limits (alpha = 0.05). The abcissa represents a measure of successful practice which is calculated by adding proportion correct over blocks of practice.
phase. Unfortunately the practice of fitting curves to data plotted against Cumulative Accuracy complicates this estimation procedure. In this experiment, the expected Premise RT of the experimental group in the first block of Transfer was 1363.37 ms. This was calculated by extrapolating the Training function of this group ($T = 5058.3 P^{-0.70275}$) one block from the end of Training (i.e., $P = 6.460 = \text{Cumulative Accuracy up to Block 7}$). The observed RT for this group in the first block of Transfer was 2519.51 ms (when $P_n = 0.978 = \text{Cumulative Accuracy in Block 7}$). Therefore the intercept value for the new power function ($N_n$) can be estimated by the following application of the Old/New Equation:

$$T = X + N_0P_0^C + N_nP_n^C$$

where $P = \text{Cumulative Accuracy}$

$$=> 2519.51 = 0 + 5058.3 (6.460)^{-0.70275} + N_n (0.978)^{-0.70275}$$

$$=> 2519.51 - 1363.37 = N_n (1.014299)$$

$$=> 1139.84 = N_n$$

Therefore $T = 5058.3 P_0^{-0.70275} + 1139.84 P_n^{-0.70275}$

The version of Old/New Equation described above has been plotted in Figure 5.3. It is clear from this figure that although this function accounts for the initial slowing in the first block of Transfer, it accounts for little more of the data in the remaining blocks of Transfer. The learning rate predicted by this equation is much faster than was observed (-0.391 vs. -0.105). However, this equation did predict Premise RTs that were much closer to the observed values than those predicted by the Old Equation (rmsd of Old/New Equation = 654.64 ms vs. rmsd of Old Equation = 1178.11 ms). Thus the Old/New Equation provides a better account of Transfer performance in this situation than the Old Equation. However, this equation does not fully capture the
pattern of improvement that follows the initial slow performance in this condition. Learning rate appears to be attenuated even more than is predicted by the Old/New Equation.

5.2.3.3 Summary and Conclusions

In those conditions where productions developed during Training were appropriate for performance during Transfer, the rate of improvement during Transfer was as predicted on the basis of Training performance. When Transfer data was plotted from Block 1, as if it was from a new task, this resulted in an attenuation of learning rate in comparison to the rate observed during Training. However this was to be expected if Transfer performance was merely the continuation of improvement begun in Training. In this respect, the Old Equation, which includes previous experience with a task as part of the practice component, was shown to provide a good account of the data in these conditions.

A different pattern of results was found in the Premise RTs of the experimental group, which were expected to show the effects of the combination of old and new skills. The analyses of variance of these results discussed above, and the results of Experiment 3 supported the conclusion that Transfer performance in this condition reflects the need to develop new productions. This still appears to be a reasonable conclusion. However, what the results cast doubt upon is the prediction of the effect of this combination of old and new skills on the learning rate during Transfer. It was predicted on the basis of the Old/New Equation that this combination would result in a learning rate that was slower than the learning rate during Transfer, but that was faster than that expected if Transfer performance was merely a continuation of Training performance. This result was not apparent in the
data. Instead, the Transfer learning rate was found to be slower than both of these other learning functions. In other words, the degree of learning rate attenuation from Training to Transfer was found to be more than the expected maximum amount of attenuation based on extrapolated Training performance.

In summary, Experiment 4 was designed to examine improvement following transfer. When transfer was complete predictions were straightforward and supported by the data. However, when transfer was only partial, and it was necessary to develop new productions, predictions were more complicated. In some respects these predictions were supported by the data (e.g., the slowing effect throughout Transfer). However the predicted attenuation of learning rate was less than was observed in the data. But the fact that Transfer performance in this situation could not be accounted for simply by extrapolating Training performance suggests that something more complicated than continued improvement of old skills is occurring.

The main result that casts doubt upon the accuracy of the Old/New Equation is the observation that the Experimental group's rate of improvement in Premise RTs was slower than the expected minimum learning rate. There is no obvious alternative view of the processes underlying transfer in this situation that can account for this observation. It is possible, though, that this observation is a methodological artifact rather than a telling falsification of the theory underlying the Old/New Equation. In only 2 of the 8 sets of RT data that were fitted with power functions were functions with non-zero asymptotes found to provide better degrees of fit than functions with zero asymptotes. Rather than suggesting that in the remaining 6 conditions subjects would eventually be able to perform the tasks in zero time, this result implies that the amount of practice in each phase may not have been sufficient to allow accurate estimates of asymptotes. If this is the case then the estimates
of learning rate that were obtained would not have been accurate either. Furthermore, any factor that affects the accuracy of power functions describing Training performance will also affect the ability of the Old/New Equation to account for the Transfer data, because the derivation of this function is based on the Training functions. It seems prudent then that before the Old/New Equation is abandoned as a model of Transfer, more extensive amounts of practice are examined in order to increase the likelihood of obtaining accurate measures of learning rate.

5.3 Experiment 5

5.3.1 Introduction

Experiment 5 was conducted to examine further the learning rate predictions of the Old/New Equation. The amount of practice in the Training and Transfer phases of this experiment was substantially increased in an attempt to increase the probability of obtaining accurate estimates of asymptotic performance. This in turn would increase the accuracy of learning rate estimates.

The experiment was also designed to examine the Transfer effects reported in earlier experiments in relation to the performance of individual subjects. It was not feasible to examine this in the experiments that have already been described. Most of the effects in these experiments were considered with respect to power functions fitted to group data. Fitting power functions to individual data in these experiments would not result in accurate representations of performance because there was considerable within-subject variation. Over 6 blocks of practice, the change in performance of each subject was not always in one direction and so power functions could not always account for changes with practice. Experiment 5, though, enabled a
more extensive look at performance changes with practice and so individual learning functions could be generated on the basis of general within-subject trends.

The data from two subjects were studied in this experiment. The design was similar to the Experimental condition in Experiment 4 in that two phases were used. The Training phase involved practice with syllogisms that contained premise pairs with capitalised common elements. In the Transfer phase this highlighting feature was removed from the premise pairs.

The amount of practice involved in the experiment was equivalent to six times one Experiment 4 session. In the Training phase the subjects were given 3 sessions of practice on 3 separate days. This same amount of practice was then given in the Transfer phase, again on separate days.

The predicted results for this experiment were essentially the same as for Experiment 4. Basically these were as follows: (1) Because subjects need to develop new productions during Transfer, Premise RTs should be slowed in the initial blocks of Transfer compared to performance at the end of Training. (2) The rate at which Premise RTs improve during Transfer should be slower than the learning rate observed during Training. However, the Old/New Equation predicts that this Transfer learning rate should be faster than the rate predicted by extrapolating Training performance into this phase. (3) The removal of the highlighting feature from premises should not affect the processing of conclusions. Hence Conclusion RTs should be unaffected by the change from Training to Transfer. That is, Conclusion RTs during Transfer should not deviate significantly from extrapolated Training performance.
5.3.2 Method

5.3.2.1 Subjects

Two female subjects, H.S. and M.S., were used in this experiment. H.S. was 28 years old and M.S. was 23 years old. Both subjects were paid $5 per session.

5.3.2.2 Materials and Design

The 576 items that were used in Experiment 4 were presented in each session in this experiment. In the three sessions in the Training phase of this experiment all items were presented with the common elements in upper case. In the three Transfer sessions these same items were presented with all of the elements in lower case.

The first 288 items that were presented to H.S. were the Training items presented in Experiment 4. The next 288 items that were presented in this first session were the Transfer items of Experiment 4. In H.S.'s second session the order of presentation of these two sets of items was reversed. This order alternated with each session for the remaining 4 sessions. For M.S. the initial order was the reverse of H.S.'s first order and this alternated with each successive session.

5.3.2.3 Apparatus and Procedure

These were identical to those described in Experiment 4 except in this experiment six experimental sessions were used. The first three sessions were Training sessions where the premise pairs of each syllogism were presented
with common elements in upper case. The remaining three sessions were Transfer sessions, with all elements in lower case. The Training sessions were conducted on the Monday, Wednesday and Friday of one week and the Transfer sessions were conducted on the same days of the following week. The first session for both subjects lasted approximately 80 minutes, but with practice, the subjects were able to reduce subsequent session times to under 60 minutes, although session times increased in the Transfer phase. The subjects were given one minute rest periods every 48 trials. Practice trials were only presented at the beginning of session one.

5.3.3 Results and Discussion

The data was divided into blocks of 48 trials. As with the previous experiments, only correct trials were analysed. The mean Premise and Conclusion RTs for each subject during the Training and Transfer phases are displayed in Figures 5.4 and 5.5. Reaction times have again been plotted against Cumulative Accuracy. Power functions based on Training performance have been included in these figures and these have been extrapolated into the Transfer phase (see Tables 5.2 and 5.3 respectively for equations of these power functions). Curves describing applications of the Old/New Equation to the Transfer data have also been included in these figures and will be discussed below.

Both subjects improved their performance times substantially during the Training phase. H.S. reduced her Premise RTs by approximately 3000 ms and her Conclusion RTs by approximately 700 ms. M.S. improved her Premise RTs by almost 6000 ms and her Conclusion RTs by approximately 2500 ms. Although there was a fair degree of variation around this improvement with practice, the power functions that provide the best fit to the
Training RTs account for, on average, 94% of the variance in the data (see Tables 5.2 and 5.3). This is a remarkably good fit for data from individual subjects. Despite this though, and the greater amount of practice observed during Training in this experiment compared to previous experiments, only the power functions fit to H.S.'s Training data revealed non-zero asymptotes. The Training data of M.S. were best described by power functions without asymptotes.

It is clear from the figures that both subjects were affected by the removal of the capitalised common elements from premise pairs. The Premise RTs of both subjects during the initial Transfer blocks were substantially slower than during the final Training blocks, by approximately 1200 ms. Surprisingly, the Conclusion RTs of both subjects were similarly affected. H.S. spent approximately 400 ms longer processing conclusions in the first Transfer block than she took in the last Training block. M.S. was slower by approximately 800 ms.

The best-fit power functions for the Transfer data (see Tables 5.2 and 5.3) show that both subjects improved during this phase at a slower rate than during Training. Whether this is a result of continued improvement upon Training performance or the result of a combination of old and new skills will be evaluated with respect to the size of learning rate differences and the ability of the Old and Old/New Equations to account for the Transfer data. Before this issue can be evaluated though, the derivation of the various versions of the Old/New Equation will be described.

Following the same method described in the previous chapter, each version of the Old/New Equation was the sum of two power functions. The first function provided the best fit to the Training data and the second function was
Figure 5.4: Mean Premise and Conclusion RTs for H.S. during Training and Transfer phases of Experiment 5. Lines drawn through Training data are best-fit power functions (see Table 5.2 for equations). These lines have been extrapolated into the Transfer phase. The other lines are specific versions of the Old/New Equation derived from the data of this experiment (see text for details).
Table 5.2: Parameters of power functions fitted to Premise, Conclusion and Total RTs of H.S. during Training and Transfer phases of Experiment 5. Functions labelled "Observed" were fitted directly to the data. Functions labelled "Old Equation" were fitted to values extrapolated from Training performance. Functions labelled "Old/New Equation" were fitted to values predicted by various versions of the Old/New Equation (see text for details). $r^2 =$ proportion of variance in the observed reaction time values accounted for by the predicted values. rmsd = root mean squared deviation between predicted and observed reaction time values.

<table>
<thead>
<tr>
<th></th>
<th>Parameters</th>
<th>Goodness of Fit</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>X</td>
<td>N</td>
<td>c</td>
<td>$r^2$</td>
</tr>
<tr>
<td><strong>Premises</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training</td>
<td>45</td>
<td>1871.60</td>
<td>-0.957</td>
<td>0.956</td>
</tr>
<tr>
<td>Transfer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>310</td>
<td>585.57</td>
<td>-0.534</td>
<td>0.840</td>
</tr>
<tr>
<td>Old Equation</td>
<td>0</td>
<td>114.92</td>
<td>-0.010</td>
<td>0.338</td>
</tr>
<tr>
<td>Old/New Equation</td>
<td>65</td>
<td>1272.60</td>
<td>-0.881</td>
<td>0.939</td>
</tr>
<tr>
<td>Conclusions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training</td>
<td>210</td>
<td>499.43</td>
<td>-0.628</td>
<td>0.930</td>
</tr>
<tr>
<td>Transfer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>50</td>
<td>539.14</td>
<td>-0.178</td>
<td>0.832</td>
</tr>
<tr>
<td>Old Equation</td>
<td>0</td>
<td>267.68</td>
<td>-0.022</td>
<td>0.569</td>
</tr>
<tr>
<td>Old/New Equation</td>
<td>225</td>
<td>404.17</td>
<td>-0.534</td>
<td>0.887</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training</td>
<td>325</td>
<td>2647.10</td>
<td>-1.056</td>
<td>0.970</td>
</tr>
<tr>
<td>Transfer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>625</td>
<td>927.88</td>
<td>-0.567</td>
<td>0.798</td>
</tr>
<tr>
<td>Old Equation</td>
<td>0</td>
<td>393.38</td>
<td>-0.026</td>
<td>0.413</td>
</tr>
<tr>
<td>Old/New Equation</td>
<td>340</td>
<td>1533.20</td>
<td>-0.951</td>
<td>0.960</td>
</tr>
</tbody>
</table>
Figure 5.5: Mean Premise and Conclusion RTs for M.S. during Training and Transfer phases of Experiment 5. Lines drawn through Training data are best-fit power functions (see Table 5.3 for equations). These lines have been extrapolated into the Transfer phase. The other lines are specific versions of the Old/New Equation derived from the data of this experiment (see text for details).
### Table 5.3: Parameters of power functions fitted to Premise, Conclusion and Total RTs of M.S. during Training and Transfer phases of Experiment 5.

<table>
<thead>
<tr>
<th>Premises</th>
<th>Parameters</th>
<th>Goodness of Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>X</td>
<td>N</td>
</tr>
<tr>
<td>Training</td>
<td>0</td>
<td>4963.40</td>
</tr>
<tr>
<td>Transfer</td>
<td>Observed</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>Old Equation</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Old/New Equation</td>
<td>55</td>
</tr>
<tr>
<td>Conclusions</td>
<td>Training</td>
<td>0</td>
</tr>
<tr>
<td>Transfer</td>
<td>Observed</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Old Equation</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Old/New Equation</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>Training</td>
<td>0</td>
</tr>
<tr>
<td>Transfer</td>
<td>Observed</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>Old Equation</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Old/New Equation</td>
<td>70</td>
</tr>
</tbody>
</table>

- **Premises**: Parameters of power functions fitted to Premise, Conclusion and Total RTs of M.S. during Training and Transfer phases of Experiment 5. Functions labelled "Observed" were fitted directly to the data. Functions labelled "Old Equation" were fitted to values extrapolated from Training performance. Functions labelled "Old/New Equation" were fitted to values predicted by various versions of the Old/New Equation (see text for details). \( r^2 \) = proportion of variance in the observed reaction time values accounted for by the predicted values. rmsd = root mean squared deviation between predicted and observed reaction time values.
a new power function. The intercept of this new function was calculated by subtracting the reaction time for the first Transfer block, predicted on the basis of Training performance, from the observed mean reaction time of the first Transfer block. Thus for H.S., her observed Premise RT in the first Transfer block was 1597.52 ms. On the basis of her Training performance (i.e., RT = 45 + 1871.6 \( P^{-0.957} \)) the RT for the first Transfer block (i.e., when \( P = 35.458 \) = Cumulative Accuracy up to Block 37) was predicted to be 106.54 ms. Therefore the intercept value for the new component of the version of the Old/New Equation to describe Premise Transfer can be estimated by the following application of the Old/New Equation:

\[
T = X + N_0 P_0^c + N_n P_n^c
\]

where \( P = \) Cumulative Accuracy

\[
=> 1597.52 = 45 + 1871.6 (35.458)^{-0.95697} + N_n (0.833)^{-0.95697}
\]

\[
=> 1251.7903 = N_n
\]

therefore \( T = 45 + 1871.6 P_0^{-0.95697} + 1251.79 P_n^{-0.95697} \)

For M.S. the derivation of a version of the Old/New Equation to describe Premise Transfer is the same. The observed first Transfer block RT was 1355.9 ms and the prediction based on Training was 289.20 ms. Thus the appropriate equation is derived by the following application of the Old/New Equation:

\[
T = X + N_0 P_0^c + N_n P_n^c
\]

where \( P = \) Cumulative Accuracy

\[
=> 1355.9 = 0 + 4963.4 (36.292)^{-0.79149} + N_n (1)^{-0.79149}
\]

\[
=> 1066.7 = N_n
\]

therefore \( T = 4963.4 P_0^{-0.79149} + 1066.7 P_n^{-0.79149} \)
The fact that the Conclusion RTs of both subjects were slower during the first Transfer blocks than in the final Training blocks suggests that, for some reason, the processing of conclusions was affected by the removal of capitalised common elements from premise pairs. This is a surprising result considering that in none of the previous experiments was a similar "sympathetic" slowing observed. The most obvious implication of this result is that these two subjects were carrying over the processing of the premises into the time that previously has been assumed to involve only the processing of conclusions. This suggestion is supported by the observation that, in the later stages of Training, both subjects spent less time studying premises than was taken to respond to conclusions (see Figures 5.4 and 5.5). Considering the information in premise pairs that must be processed before a conclusion can be evaluated (see Figure 3.1), it seems unlikely that the processing of premises could be faster than the processing of conclusions. Presumably then, the subjects in this experiment were pressing "READY" before completing the processing of the presented premises, but were able to continue this processing without requiring that the premises remain visible. If this is the case then it suggests that improvement in Conclusion RTs during the Transfer phase should also reflect improvement in both old and new productions. Therefore it would seem appropriate to derive versions of the Old/New Equation to test in relation to Conclusion RTs as well.

The mean Conclusion RT in the first Transfer block of H.S. was 678.58 ms and the Training prediction was 263.21 ms. Therefore the appropriate version of the Old/New Equation is derived as follows:

\[ T = X + N_0P_0^C + N_nP_n^C \]

where \( P = \text{Cumulative Accuracy} \)

\[
\begin{align*}
678.58 &= 210 + 499.43 \times (35.458)^{0.62751} + N_n \times (0.833)^{0.62751} \\
370.37 &= N_n
\end{align*}
\]
therefore $T = 210 + 499.43 P_o^{-0.62751} + 370.37 P_n^{-0.62751}$

For M.S., the mean Conclusion RT for the first Transfer block was 871.73 ms and the Training prediction was 420.63 ms. This gives rise to the following equation:

$$T = X + N_0 P_o^c + N_n P_n^c$$

where $P =$ Cumulative Accuracy

$$\Rightarrow 871.73 = 0 + 2208.8 (36.292)^{-0.46176} + N_n (1)^{-0.46176}$$

$$\Rightarrow 451.10 = N_n$$

therefore $T = 2208.8 P_o^{-0.46176} + 451.10 P_n^{-0.46176}$

For both subjects, learning rate during Transfer was slower than during Training, in both Premise and Conclusion RTs, and was also slower than that predicted by the Old/New Equation (see Tables 5.2 and 5.3). For H.S., the Transfer learning rate of Premise RTs was faster than that predicted by extrapolating Training performance (i.e., the Old Equation). There did not appear to be much difference between these observed and predicted rates in Conclusion RTs. In both the Premise and Conclusion RTs of M.S., there also did not appear to be much difference between learning rate during Transfer and the learning rate predictions based on the Old Equation.

If Premise and Conclusion RTs during Transfer were both reflecting the effect of removing the highlighting in premise pairs, then it seems anomalous that the differences in learning rate described above were not uniform across these performance variables. For this reason, and the fact that the two performance variables apparently do not provide a clear distinction in terms of what is being processed, Total RTs were calculated in order to reflect the
overall processing strategy of the two subjects. The predictions for the behaviour of this variable are identical to those for Premise RTs. In other words, this variable should reflect the effects of combining old and new skills.

Power functions were fit to the Total RTs of both subjects and the parameters of these functions are presented in Tables 5.2 and 5.3. The mean Total RTs for both subjects during Training and Transfer are displayed in Figure 5.6. This figure includes curves fit to the Training data and extrapolated into the Transfer phase and curves representing appropriate versions of the Old/New Equation. The version for H.S. was:

\[ T = 325 + 2647.1 P_o^{-1.0555} + 1558.37 P_n^{-1.0555} \]

and the version for M.S. was:

\[ T = 6840.6 P_o^{-0.63339} + 1524.35 P_n^{-0.63339} \]

Tables 5.2 and 5.3 show that with the Total RTs of both subjects, the learning rate during Transfer was slower than during Training and slower than the rate predicted by the Old/New Equation. For H.S. this learning rate was faster than that predicted on the basis of the Old Equation. For M.S. this learning rate did not differ substantially from the Old Equation prediction.

The analysis of learning rate differences does not provide a clear picture of the relative merits of the Old and Old/New Equations. In the results of H.S., the Old/New Equation was more accurate in predicting learning rate in Premise RTs, the Old Equation was more accurate with Conclusion RTs, and neither equation was close to the observed rate with Total RTs. In contrast, the Old
Figure 5.6: Mean Total RTs for H.S. and M.S. during Training and Transfer phases of Experiment 5. Lines drawn through Training data are best-fit power functions (see Tables 5.2 and 5.3 for equations). These lines have been extrapolated into the Transfer phase. The other lines are specific versions of the Old/New Equation derived from the data of this experiment (see text for details).
Equation appeared to provide the best learning rate predictions for all three performance variables in the results of M.S. However, the power functions that were fitted to the Transfer data of M.S. did not account for a large degree of variance in her performance. These functions accounted for an average of 39% of the variance in the Transfer data, whereas the functions fitted to M.S.’s Training data accounted for an average of 94% of the variance in performance. The main reason for this reduction in the ability of power functions to accurately describe M.S.’s learning performance was that M.S. suffered from influenza during Day 2 of the Transfer phase. The effect of this illness was to slow performance and increase the variability of performance. This effect is clearly visible in Figure 5.5 and is most obvious in Figure 5.6. By the third day of Transfer the effects of the illness were gone, but the damage had been done. The overall Transfer performance of M.S. could not be described well by a power function. Therefore this analysis is not informative as to the effect on learning rate of combining old and new skills.

A more informative analysis involves the degree to which the various equations can account for Transfer performance. The goodness-of-fit of the Old and Old/New Equations to the Transfer data was measured with $r^2$ and rmsd. These results are presented in Tables 5.2 and 5.3. The degree to which these equations account for the Transfer data is also represented graphically in Figures 5.4, 5.5 and 5.6. In these figures, the Old Equation is represented by the extrapolation of Training performance into the Transfer phase.

It is clear from Tables 5.2 and 5.3 that in all cases - both subjects and all three performance variables - the Old/New Equation provides a superior account of the Transfer data than the Old Equation, accounting for a greater proportion of the variance in the data and predicting reaction times that, on average, were closer to the observed values. This result is demonstrated clearly also in the
three figures. However, it is also clear that the Old/New Equation does not provide a perfect description of Transfer in this situation. In all cases the rate of improvement in Transfer that was predicted by the Old/New Equation was faster than was observed (see Tables 5.2 and 5.3). The Old/New Equation is able to account for the first four blocks of Transfer, which show most of the effect of removing the highlighting from premise pairs, but in subsequent blocks, performance improves at a slower rate than is predicted. This result will be discussed further below.

5.3.4 Summary and Conclusions

Experiment 5 had two major aims. The first was to replicate the effects observed in Experiment 4. The second was to observe these effects under conditions involving greater amounts of practice in order to evaluate further the ability of the Old/New Equation to describe partial transfer. With respect to the first aim, all of the major effects observed in Experiment 4 were also observed in this experiment. Both subjects in this experiment were clearly affected by the removal of highlighting in the premise pairs. Premise and Conclusion RTs were slower in the first blocks of the Transfer phase than in the final blocks of Training. This result replicates the results of Experiment 3 and 4 and suggests that the subjects found it necessary to develop new productions in order to solve the syllogisms without capitalised common elements.

The fact that the processing of conclusions was affected by the removal of highlighting suggests that it is not always the case that subjects have finished processing premise pairs when they press the "READY" button. The two subjects in this experiment appeared to continue this processing of the premises into the time previously assumed to reflect conclusion processing
only. The previous support for this assumption came from experiments where group data was analysed. It is possible that subjects differ in the exact strategy used at this point of processing, with the majority in the previous experiments using the simple strategy (i.e., the strategy described in Figure 3.1). The results of this majority would then dilute the effects of subjects who followed the strategy apparently used in this experiment. The fact that Premise and Conclusion RTs in the previous experiments exhibited results which were predicted on the basis of the simple strategy supports the assumption that most subjects follow this strategy. The results of Experiment 5 though, illustrate that this assumption is not always a safe one to make.

The Old/New Equation was shown to provide a better account of Transfer performance than the Old Equation. This implies that in this experiment, Transfer performance was not simply a continuation of Training performance (as described by the Old Equation). Therefore a good account of Transfer in this situation needs to involve the combination of improvement of old skills with improvement of new skills. The form of this combination appears to be similar to that described by the Old/New Equation.

Although the Old/New Equation provided the best account of the Transfer data in this experiment, it was not very accurate in predicting the pattern of improvement observed during Transfer. In fact the predicted learning rate was always faster than was observed. A similar observation was made of the learning rates in Experiment 4. This suggests that the Old/New Equation is not a totally adequate description of improvement with practice following the combination of old and new skills. Apparently some other factor or factors need to be considered.
One of the major aims of Experiment 5 was to explore one potential explanation for why the Old/New Equation is unable to account for Transfer performance when old and new skills are combined. The Old/New Equation is derived from the power function that provides the best fit to the Training data. Thus the accuracy of the Old/New Equation is dependent on the accuracy of the Training function. Hence the hypothesis tested by this experiment was that some factor was affecting the accuracy of functions describing Training performance. The specific hypothesis under test was that the amount of practice observed in Experiment 4 was not sufficient to derive an accurate power function description of Training performance. However, the results of Experiment 5 appear to falsify this hypothesis. Three times the amount of practice given in Experiment 4 was given in this experiment and yet the ability of the Old/New Equation to account for Transfer performance was not improved. Furthermore, the Old/New Equation again underestimated the degree to which learning rate is attenuated following partial transfer. Thus some other factor is operating during partial transfer that is not accounted for by the Old/New Equation.

Before another hypothesis is considered, it is necessary to establish that the effects reported in Experiments 4 and 5 concerning changes in learning functions are peculiar to situations involving partial transfer. Experiment 6 was designed to examine learning functions following complete transfer. If learning functions are affected in this experiment in a similar way to that observed in the previous experiments, then the Old/New Equation will be rendered redundant. The attenuation of learning rate following transfer will need to be attributed to factors other than the combination of old and new skills.
5.4 Experiment 6

5.4.1 Introduction

The results of Experiment 4 revealed changes in learning functions that were almost as predicted by the Old/New Equation. It was assumed that these changes resulted from the combination of old and new skills. There was also evidence in Experiment 4 that when Transfer performance relied on old productions only, this performance could be predicted by extrapolating the learning function that described Training performance. However, in those conditions where such complete transfer was observed, the transition from Training to Transfer was not marked by any change in stimulus conditions, in contrast with the obvious change associated with removing highlighting from premise pairs. Thus performance in these conditions would be expected to continue improving as if nothing had changed, as indeed it had not.

If the Old/New Equation is to provide a reasonable account of the combination of old and new skills then it is necessary to demonstrate that old skills will continue to improve according to the learning function observed during Training. Preliminary evidence in support of this was provided in the complete transfer conditions described above. However a more stringent requirement is that old skills continue to improve in this fashion when stimulus conditions have changed. In other words, it is necessary to demonstrate that old productions will function as before when executed in different contexts. Such a demonstration would support the assumption underlying the Old/New Equation that, when improvement in a new task is partly associated with continued improvement of old skills, this component can be described by the learning function describing initial improvement of these old skills.
Experiment 6 was designed to contrast learning functions following complete transfer in two conditions. The first condition was a control condition and involved the same stimulus conditions during Transfer as were present during Training. The experimental condition involved slightly different conditions during Transfer to those that were present during Training. However, despite the differences between these two conditions it was assumed that complete transfer would occur in both conditions.

The design of Experiment 6 was similar to that of Experiment 2, with the addition of a control group. Both groups were given Alternating Training. The experimental group was then given Random Transfer items to solve, whereas the control group was presented with Alternating Transfer items. Thus stimulus conditions for the control group were equivalent in both phases of the experiment but were different for the experimental group. Because there was no difference between Training and Transfer for the control group it was expected that complete transfer would result. That is, this group would execute the productions developed during Training in response to the Transfer items. As a result, the Transfer performance of this group was expected to be predicted by extrapolating their Training performance. In other words, the Old Equation should provide a good account of the control group's Transfer performance.

Considering the results of Experiment 2, it was assumed that, despite the change in stimulus conditions from Training to Transfer, the experimental group would be able to apply the productions developed during Training to the Transfer problems. Furthermore, there would not be any need to develop new productions. Thus complete transfer was assumed to occur in this condition also. In addition, the old productions were assumed to improve during Transfer according to the learning function describing Training
performance, despite there being a change in stimulus conditions. As a result, the Transfer performance of the experimental group should also be accounted for by the Old Equation.

Experiment 6 should be considered a control experiment for Experiment 4. Experiment 4 demonstrated changes in learning functions that resulted from changes in stimulus conditions. The changes in stimulus conditions in Experiment 4 were assumed to result in the development of new productions to complement the old productions. Experiment 6 was designed to demonstrate that changes in stimulus conditions that do not require the development of new productions will not affect learning functions.

5.4.2 Method

5.4.2.1 Subjects

Thirty-six volunteers from the University of Western Australia first year Psychology course participated for course credit. Four subjects did not reach the learning criterion of an error rate of not more than 25% in the last half of the Training phase. The data from these subjects were not analysed further, leaving data from 32 subjects, 16 per condition.

5.4.2.2 Materials

The syllogisms presented during the Training phase were the same items presented during Training in Experiment 2. The order of presentation in this experiment was also the same as in Experiment 2.
The syllogisms presented to the experimental group during Transfer were the same as those presented during Transfer in Experiment 4. The items presented to the control group during this phase were the same as those presented to the experimental group, except for the control group these items were presented in an alternating order. That is, on the first trial, the syllogism was of the ABBC type. On the second trial, a BCAB syllogism was presented, followed by another ABBC syllogism on the third trial, and so on.

5.4.2.3 Design

The design of this experiment was similar in some respects to that of Experiment 4. The experimental and control groups performed in equivalent circumstances during Training (i.e., Alternating Training). During Transfer the control group performed with more Alternating items, whereas the experimental group were presented with Random items.

5.4.2.4 Apparatus and Procedure

These were identical in all respects to those of Experiment 4.

5.4.3 Results and Discussion

The results will be presented in two sections. The first concerns the general trends in the data. The second concerns learning functions.
5.4.3.1 General Trends

5.4.3.1.1 Training

A number of analyses of variance were performed on the Training data. A 2 (Training condition) x 6 (Training block) x 2 (syllogism type) ANOVA was used to analyse Premise RTs. Two 2 (Training condition) x 6 (Training block) x 2 (syllogism type) x 2 (conclusion type) ANOVAs were used to analyse Conclusion RTs and Accuracy.

Mean Premise and Conclusion RTs of the experimental and control groups during Training and Transfer are presented in Figure 5.7. Despite what the figure suggests, there were no differences between the two groups during the Training phase in Premise RTs (experimental group = 3874 ms vs. control group = 3323 ms, F(1,30) = 1.16, p>0.05). Conclusion RTs (experimental group = 1307 ms vs. 1110 ms. F(1,30) < 1) or Accuracy (experimental group = 89.70% vs. control group = 92.66%, F(1,30) < 1).

Overall performance improved with practice during Training. Premise RTs were reduced from 5759 ms in the first block of Training to 2507 ms in the final block (F(1,150) = 70.56, p<0.05). Conclusion RTs improved from 1986 ms to 888 ms (F(1,150) = 67.15, p<0.05). Accuracy increased from 78.86% to 94.45% (F(1,150) = 17.55, p<0.05).

Syllogism type again affected performance, with the premises of ABBC syllogisms being studied for less time than the premises of BCAB syllogisms (ABBC = 3482 ms vs. BCAB = 3715 ms, F(1,30) = 11.02, p<0.05). ABBC syllogisms were also solved correctly more often (ABBC = 91.77% vs. BCAB = 90.59%, F(1,30) = 4.62, p<0.05). There were also significant
Figure 5.7: Mean Premise and Conclusion RTs for both Experimental and Control groups, during Training and Transfer phases of Experiment 6.
interactions of syllogism type with practice in Premise RTs and Accuracy
(Premise RTs: F(5,150) = 3.16, p<0.05; Accuracy: F(5,150) = 3.11, p<0.05). These interactions both took the form of the performance difference between the two types of syllogism being reduced with practice. Syllogism type did not affect Conclusion RTs (ABBC = 1211 ms vs. BCAB = 1206 ms, F(1,30) < 1).

True conclusions were again responded to faster than False conclusions (True = 1126 ms vs. False = 1290 ms, F(1,30) = 38.83, p<0.05). A similar, but non-significant advantage for True conclusions was found with Accuracy (True = 92.72% vs. False = 89.64%, F(1,30) = 4.06, p>0.05).

5.4.3.1.2 Transfer

A number of analyses of variance were performed on the Transfer data. A 2 (Training condition) x 6 (Transfer block) x 2 (syllogism type) ANOVA was used to analyse Premise RTs. Two 2 (Training condition) x 6 (Transfer block) x 2 (syllogism type) x 2 (conclusion type) ANOVAs were used to analyse Conclusion RTs and Accuracy.

The only significant effects observed during the Transfer phase were an improvement with practice with Premise RTs (Block 1 = 2365 ms to Block 6 = 1741 ms, F(5,150) = 8.79, p<0.05), an effect of syllogism type on Conclusion RTs (ABBC = 836 ms vs. BCAB = 798 ms, F(1,30) = 8.05, p<0.05), and the ubiquitous advantage of True conclusions over False conclusions (True = 780 ms vs. False = 853 ms, F(1,30) = 21.00, p<0.05). The two groups did not differ in any of the three performance measures (Premise RTs: experimental group = 2292 ms vs. control group = 1786 ms, F(1,30) = 2.14, p>0.05; Conclusion RTs: experimental group = 845 ms vs.
control group = 789 ms, \( F(1,30) < 1 \): Accuracy: experimental group = 92.25% vs. 94.27%, \( F(1,30) < 1 \).

5.4.3.2 Learning Functions

Power functions have been fitted to the Training data of both groups. Figure 5.8 displays these functions for the control group. Figure 5.9 displays the functions for the experimental group. In both of these figures the functions describing Training performance have been extrapolated into the Transfer phase and so represent applications of the Old Equation. The equations of these functions are presented in Table 5.4.

The learning function results are straightforward. In the Premise and Conclusion RTs of both groups, the Old Equation was able to provide accurate predictions of Transfer performance, predicting power functions which were similar to those that provided the best fit to the data and which had \( r^2 \) and rmsd values of comparable size to those of the functions derived from the data. This is obvious from Figures 5.8 and 5.9. In addition, the equations of the functions fitted to the observed and predicted reaction times are quite similar in all cases except one (see Table 5.4). The Conclusion RTs of the experimental group during Transfer were not accounted for well by a simple power function, which accounted for only 13.8% of the variance. Interestingly, the Old Equation was able to improve on this, accounting for 25.2% of the variance in the observed data. However, neither of these functions account for appreciable proportions of variance compared to all of the other functions. The main reason for this is that, overall, there was not a significant amount of improvement in Conclusion RTs during this phase. Even the functions fitted to the Conclusion RTs of the control group only accounted for about 67% of the variance in the data. All the functions fitted to
Figure 5.8: Mean Premise and Conclusion RTs for the Control group during Training (Alternating) and Transfer (Alternating) phases of Experiment 6. Lines drawn through Training data are best-fit power functions (see Table 5.4 for equations). These lines have been extrapolated into the Transfer phase and so represent applications of the Old Equation in this phase. Error bars are confidence limits (alpha = 0.05).
Figure 5.9: Mean Premise and Conclusion RTs for the Experimental group during Training (Alternating) and Transfer (Random) phases of Experiment 6. Lines drawn through Training data are best-fit power functions (see Table 5.4 for equations). These lines have been extrapolated into the Transfer phase and therefore represent applications of the Old Equation in this phase. Error bars are confidence limits (alpha = 0.05).
## Table 5.4: Parameters of power functions fitted to Premise and Conclusion RTs during Training and Transfer phases of Experiment 6. Functions labelled "Observed" were fitted directly to the data. Functions labelled "Old Equation" were fitted to values extrapolated from Training performance. $r^2 = \text{proportion of variance in the observed reaction time values accounted for by the predicted values}$. $\text{rmsd} = \text{root mean squared deviation between predicted and observed reaction time values.}$

<table>
<thead>
<tr>
<th></th>
<th>Parameters</th>
<th>Goodness of Fit</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$X$</td>
<td>$N$</td>
<td>$c$</td>
<td>$r^2$</td>
</tr>
<tr>
<td><strong>Experimental Group</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Premises</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Training</strong></td>
<td></td>
<td></td>
<td>300</td>
<td>4909.80</td>
<td>-0.385</td>
<td>0.993</td>
</tr>
<tr>
<td><strong>Transfer</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Observed</strong></td>
<td></td>
<td></td>
<td>55</td>
<td>2547.30</td>
<td>-0.137</td>
<td>0.928</td>
</tr>
<tr>
<td><strong>Old Equation</strong></td>
<td></td>
<td></td>
<td>40</td>
<td>2694.60</td>
<td>-0.106</td>
<td>0.981</td>
</tr>
<tr>
<td><strong>Conclusions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Training</strong></td>
<td></td>
<td></td>
<td>445</td>
<td>1349.10</td>
<td>-0.581</td>
<td>0.983</td>
</tr>
<tr>
<td><strong>Transfer</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Observed</strong></td>
<td></td>
<td></td>
<td>760</td>
<td>176.48</td>
<td>-1.170</td>
<td>0.138</td>
</tr>
<tr>
<td><strong>Old Equation</strong></td>
<td></td>
<td></td>
<td>40</td>
<td>881.81</td>
<td>-0.088</td>
<td>0.252</td>
</tr>
<tr>
<td><strong>Control Group</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Premises</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Training</strong></td>
<td></td>
<td></td>
<td>155</td>
<td>4917.70</td>
<td>-0.508</td>
<td>0.995</td>
</tr>
<tr>
<td><strong>Transfer</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Observed</strong></td>
<td></td>
<td></td>
<td>0</td>
<td>2204.00</td>
<td>-0.214</td>
<td>0.896</td>
</tr>
<tr>
<td><strong>Old Equation</strong></td>
<td></td>
<td></td>
<td>0</td>
<td>2077.60</td>
<td>-0.143</td>
<td>0.899</td>
</tr>
<tr>
<td><strong>Conclusions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Training</strong></td>
<td></td>
<td></td>
<td>655</td>
<td>972.74</td>
<td>-1.036</td>
<td>0.998</td>
</tr>
<tr>
<td><strong>Transfer</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Observed</strong></td>
<td></td>
<td></td>
<td>155</td>
<td>658.91</td>
<td>-0.037</td>
<td>0.660</td>
</tr>
<tr>
<td><strong>Old Equation</strong></td>
<td></td>
<td></td>
<td>0</td>
<td>795.97</td>
<td>-0.044</td>
<td>0.678</td>
</tr>
</tbody>
</table>
the data describe improvement over time, and therefore, if the data does not show such an improvement, the functions will obviously not describe performance well.

The fact that the learning functions of both groups during Transfer were predicted by the Old Equation supports the assumption that both of these groups would execute old productions developed during Training. The observation that the performance of the two groups did not differ substantially during Transfer supports the assumption that old productions would continue to improve according to their original learning function, despite a change in stimulus conditions. This assumption underlies the Old/New Equation. Thus it appears that changes in learning functions predicted by the Old/New Equation are in fact peculiar to situations involving partial transfer (i.e., when old and new skills are combined), and are not a feature of complete transfer.

An interesting feature of the data from Experiment 6 should be made explicit. The power functions that provided the best fit to the Training Premise and Conclusion RTs of both groups all included non-zero asymptotes (see Table 5.4). This constitutes further evidence that insufficient practice during Training is not a factor in the inability of the Old/New Equation to account for learning function changes following partial transfer. An alternate explanation for the apparent failure of the Old/New Equation to account for the pattern of improvement observed in such situations will be discussed in the following section.

5.5 An Additional Parameter in the Old/New Equation

The Old/New Equation is able to predict that learning rate will be attenuated when old and new skills are combined to perform a task. However, as has
been demonstrated in Experiments 4 and 5, the degree of attenuation predicted by the Old/New Equation is not as great as is observed. Furthermore, the Old/New Equation predicts that learning rate in this situation will not be attenuated to the same degree as when performance involves the continued improvement of old skills only. However, in Experiment 4 it was found that learning rate following partial transfer was even slower than if performance involved continued improvement of old skills only. Thus the Old/New Equation severely underestimates the degree to which learning rate will be attenuated following partial transfer.

Another feature of the learning functions following partial transfer that is not accounted for by the Old/New Equation is an apparent change in asymptote from Training to Transfer. The power functions that provided the best fit to the Premise and Total RTs of both subjects in Experiment 5 revealed an increase in asymptote from Training to Transfer. This increase is quite obvious in Figures 5.4, 5.5 and 5.6. A change in asymptote is not a feature of the Old/New Equation. In fact, in Chapter 4 (§4.5) it was assumed that the asymptote of a learning function would remain the same following transfer. This is because the asymptote was assumed to be a constant performance limit of the cognitive system, at least for a particular type of task. Thus the results of Experiment 5 appear to indicate that this assumption is unfounded. Another change in asymptote was observed in the results of Experiment 6. However, this change was in the opposite direction to that observed in Experiment 5. In the three out four situations where power functions provided a good account of the data (i.e., Premise RTs of the experimental group, and Premise and Conclusion RTs of the control group), asymptotes became smaller following the transition from Training to Transfer. The most obvious explanation for this particular change is that the asymptotes measured during Training are estimates of asymptotic performance after only six blocks of practice. The
estimates made during Transfer are made on the basis of performance that is obviously closer to the 'true' asymptotic level. Therefore the Transfer estimates are likely to be more accurate simply because they are based on data which provides a better indication of the asymptotic level of performance. Thus it appears that the increase in asymptote may only be associated with partial transfer.

If performance asymptotes are affected by partial transfer then this may explain the greater degree of learning rate attenuation than is predicted by the Old/New Equation. The learning rate and asymptote of a learning function are related, such that for a particular set of data, changing the value of one of these parameters will affect the value of the other. This suggests that if the Old/New Equation was to incorporate a change in asymptote following the combination of old and new skills, learning rate predictions based on this equation would be affected. The following will describe the effects of such an addition to this equation.

The derivation of the data-specific versions of the Old/New Equation involved estimating the intercept of a power function that described improvement of new productions. This value was assumed to be indicated by the difference between the reaction time in the first block of Transfer and the reaction time predicted by extrapolating Training performance. The latter reaction time was assumed to represent the performance time of old productions. However, if asymptotes are affected by partial transfer, then this will contaminate estimation of the intercept of the power function describing the new skills. The difference value that was used to estimate this intercept will be a result of both the execution of additional productions (i.e., the intercept) and an increase in asymptote.
One method for overcoming the contaminating effect of an asymptote increase is to incorporate the increase into the Old/New Equation. This, in fact, requires application of an earlier version of the Old/New Equation which was rejected by the assumption of a constant asymptote. This particular version was Equation 9:

\[ T_{\text{task}} = T_{\text{old}} + T_{\text{new}} \]
\[ = X_0 + N_0 P_0^c + X_n + N_n P_n^c \]  \hspace{1cm} (9)

If an estimate for \( X_n \) can be made, then this will result in an uncontaminated estimate of \( N_n \). For example, consider the Premise RTs of H.S. in Experiment 5 (see Table 5.2). The asymptote of the power function fitted to the Training data was 45 ms. The asymptote for the Transfer data was 310 ms. \( X_n \) should represent the increase in asymptote. In other words, \( X_n = X_t - X_0 \), where \( X_t \) = asymptote of overall task during Transfer. Therefore, for H.S., \( X_n = 310 - 45 = 265 \). The calculation of \( N_n \) can now proceed as before:

\[ T = X + N_0 P_0^c + X_n + N_n P_n^c \]  \hspace{1cm} \text{where } P = \text{Cumulative Accuracy}

\[ \Rightarrow 1597.52 = 45 + 1871.6 (35.458)^{-0.95697} + 265 + N_n (0.833)^{-0.95697} \]
\[ \Rightarrow 1597.52 - 106.54 - 265 = N_n \]
\[ 1.191078 \]
\[ \Rightarrow 1029.30 = N_n \]

therefore \( T = 45 + 1871.6 P_0^{-0.95697} + 265 + 1029.3 P_n^{-0.95697} \)

Incorporating the change in asymptote in the derivation of the Old/New Equation has the effect of reducing the weighting of the 'new' component of this equation compared to that of the 'old' component. In the original version
of the Old/New Equation derived for the Premise RTs of H.S., N_h was estimated at 1251.79 ms which is larger than the value estimated above (1029.30). Recall from the previous chapter (see Figure 4.1) that, according to the Old/New Equation, an increase in the ratio of old to new skills in performing a task will increase the attenuation of learning rate of that task. Thus this reduction in the weighting of the new component has the effect of attenuating the learning rate of the overall task even more than was originally predicted by the Old/New Equation. Therefore incorporating the increase in asymptote into the Old/New Equation has the effect of increasing the attenuation of learning rate. This effect is displayed in Figure 5.10, which reproduces the Premise RTs part of Figure 5.4 and includes the reaction times predicted by the new version of the Old/New Equation. It is obvious from this figure that the new version of the Old/New Equation provides a much more accurate account of the Transfer performance of H.S. than the original version. Considering this result, it appears anomalous that the learning rate predicted by this new version of the Old/New Equation is still substantially larger than the learning rate of the simple power function which provides the best fit to the Transfer data (see Table 5.5). However, it is also interesting to note that this new version of the Old/New Equation provides a better fit to the data than this simple power function, accounting for a greater proportion of the variance in the observed reaction times and predicting values which, on average, are closer to the observed values.

A similar analysis could not be performed on the Conclusion RTs of H.S. because there was a reduction in asymptote following transfer. However, there was an increase in asymptote in Total RTs. A new version of the Old/New Equation was calculated for this data:
Figure 5.10: Mean Premise RTs for H.S. during Training and Transfer phases of Experiment 5. The line drawn through the Training data is the best-fit power function (see Table 5.5 for equation). This line has been extrapolated into the Transfer phase. The other lines are specific versions of the Old/New Equation derived from the data of this experiment (see text for details). The line labelled "Old/New Equation*" represents the new version of the Old/New Equation which incorporates a change in asymptote.
### H.S.

<table>
<thead>
<tr>
<th>Premises</th>
<th>Training</th>
<th>Observed</th>
<th>Old Equation</th>
<th>Old/New Equation</th>
<th>Old/New Equation*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>310</td>
<td>65</td>
<td>330</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>45</td>
<td>1871.60</td>
<td>114.92</td>
<td>1272.60</td>
<td>1052.00</td>
</tr>
<tr>
<td>c</td>
<td>-0.957</td>
<td>-0.534</td>
<td>-0.010</td>
<td>-0.881</td>
<td>-0.868</td>
</tr>
<tr>
<td>$r^2$</td>
<td>0.956</td>
<td>0.840</td>
<td>0.338</td>
<td>0.939</td>
<td>0.938</td>
</tr>
<tr>
<td>msd</td>
<td>245.784</td>
<td>126.680</td>
<td>463.502</td>
<td>247.968</td>
<td>56.062</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Total</th>
<th>Training</th>
<th>Observed</th>
<th>Old Equation</th>
<th>Old/New Equation</th>
<th>Old/New Equation*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>325</td>
<td>2647.10</td>
<td>393.38</td>
<td>1533.20</td>
<td>1292.10</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td>-1.056</td>
<td>-0.026</td>
<td>-0.951</td>
<td>-0.939</td>
</tr>
<tr>
<td>c</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r^2$</td>
<td></td>
<td>0.970</td>
<td>0.413</td>
<td>0.960</td>
<td>0.959</td>
</tr>
<tr>
<td>msd</td>
<td></td>
<td>236.047</td>
<td>599.335</td>
<td>347.402</td>
<td>89.390</td>
</tr>
</tbody>
</table>

### M.S.

<table>
<thead>
<tr>
<th>Premises</th>
<th>Training</th>
<th>Observed</th>
<th>Old Equation</th>
<th>Old/New Equation</th>
<th>Old/New Equation*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>625</td>
<td>0</td>
<td>340</td>
<td>640</td>
</tr>
<tr>
<td>X</td>
<td>4963.40</td>
<td>927.88</td>
<td>393.38</td>
<td>1533.20</td>
<td>1292.10</td>
</tr>
<tr>
<td>c</td>
<td>-0.791</td>
<td>-0.567</td>
<td>-0.026</td>
<td>-0.951</td>
<td>-0.939</td>
</tr>
<tr>
<td>$r^2$</td>
<td>0.972</td>
<td>0.798</td>
<td>0.413</td>
<td>0.960</td>
<td>0.959</td>
</tr>
<tr>
<td>msd</td>
<td>234.703</td>
<td>167.483</td>
<td>599.335</td>
<td>347.402</td>
<td>89.390</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Total</th>
<th>Training</th>
<th>Observed</th>
<th>Old Equation</th>
<th>Old/New Equation</th>
<th>Old/New Equation*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>6840.60</td>
<td>340.68</td>
<td>1196.30</td>
<td>840.60</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td>-0.633</td>
<td>-0.172</td>
<td>-0.530</td>
<td></td>
</tr>
<tr>
<td>c</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r^2$</td>
<td></td>
<td>0.956</td>
<td>0.383</td>
<td>0.383</td>
<td></td>
</tr>
<tr>
<td>msd</td>
<td></td>
<td>311.890</td>
<td>379.030</td>
<td>379.030</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Transfer</th>
<th>Observed</th>
<th>Old Equation</th>
<th>Old/New Equation</th>
<th>Old/New Equation*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>45</td>
<td>1038.80</td>
<td>340.68</td>
<td>1196.30</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td>-0.144</td>
<td>-0.172</td>
<td>-0.530</td>
</tr>
<tr>
<td>c</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r^2$</td>
<td></td>
<td>0.342</td>
<td>0.299</td>
<td>0.383</td>
</tr>
<tr>
<td>msd</td>
<td></td>
<td>133.597</td>
<td>573.216</td>
<td>415.200</td>
</tr>
</tbody>
</table>

### Table 5.5: Parameters of power functions fitted to Premise and Total RTs of H.S. and M.S. during Training and Transfer phases of Experiment 5. Functions labelled "Observed" were fitted directly to the data. Functions labelled "Old Equation" were fitted to values extrapolated from Training performance. Functions labelled "Old/New Equation" were fitted to values predicted by various versions of the Old/New Equation (see text for details). Functions labelled "Old/New Equation*" were fitted to values predicted by the new version of the Old/New Equation which incorporates a change in asymptote (see text for details). $r^2 =$ proportion of variance in the observed reaction time values accounted for by the predicted values. msd = root mean squared deviation between predicted and observed reaction time values.
\[ X_n = X_t - X_0 \]
\[ = 625 - 325 \]
\[ = 300 \]

\[ T = X + N_0P_0^c + X_n + N_nP_n^c \quad \text{where} \quad P = \text{Cumulative Accuracy} \]

\[ \Rightarrow 2276.1 = 325 + 2647.1 (35.458)^{-1.0555} + 300 + N_n (0.833)^{-1.0555} \]

\[ \Rightarrow 2276.1 - 386.24 - 300 = N_n \]
\[ 1.212716 \]

\[ \Rightarrow 1310.99 = N_n \]

Therefore, \( T = 325 + 2647.1 P_0^{-1.0555} + 300 + 1310.99 P_n^{-1.0555} \)

Figure 5.11 displays the reaction times predicted by this new version of the Old/New Equation along with the Total RTs of H.S. Again this new equation has provided a superior account of the pattern of improvement during Transfer than was provided by the original version of the Old/New Equation. Furthermore, this new version again predicts a faster learning rate than that of the simple power function which provided the best fit to the data (see Table 5.5). However, as is also evident in Table 5.5, the new version provides a better account of the data (i.e., larger \( r^2 \) and smaller rmsd) than the simple power function.

New versions of the Old/New Equation incorporating the increase in asymptote were also derived for the Premise and Total RTs of M.S. from Experiment 5. For the Premise RTs, the new equation is:

\[ T = 0 + 4963.4 P_0^{-0.79149} + 45 + 1021.7 P_n^{-0.79149} \]
Figure 5.11: Mean Total RTs for H.S. during Training and Transfer phases of Experiment 5. The line drawn through the Training data is the best-fit power function (see Table 5.5 for equation). This line has been extrapolated into the Transfer phase. The other lines are specific versions of the Old/New Equation derived from the data of this experiment (see text for details). The line labelled "Old/New Equation*" represents the new version of the Old/New Equation which incorporates a change in asymptote.
For the Total RTs, the new equation is:

\[ T = 0 + 6840.6 P_o^{-0.63339} + 60 + 1464.35 P_n^{-0.63339} \]

Reaction times predicted by these equations have been plotted against M.S.'s data in Figures 5.12 and 5.13. Although not obvious in these two figures, the new versions of the Old/New Equation only marginally improve upon the account of the Transfer data provided by the original versions. This is evident in Table 5.5, where the new versions differ from the old versions only in rmsd values, and these differences are small. This result does not necessarily count against the new version of the Old/New Equation because of the effect of M.S.'s illness during Transfer noted before. Certainly the substantial improvement provided by this new version in accounting for H.S.'s performance is encouraging and suggests that further investigation of this new version is warranted.

Unfortunately an increase in asymptote following partial transfer was not detected in the results of Experiment 4. In fact no asymptotes were detected in Premise RTs at all in this experiment. Therefore it is difficult, and maybe inappropriate, to apply the new version of the Old/New Equation to this set of data. However, given that this new version can account for a large attenuation of learning rate from Training to Transfer, and that such a large attenuation was observed with the experimental group in Experiment 4 (i.e., -0.703 to -0.105), it would appear prudent to apply this new version for the sake of consistency.

Because no asymptotes were observed in the Premise RTs of the experimental group in Experiment 4, an arbitrary value for \( X_n \) will have to be assigned. Assuming that this value should be somewhere between zero and the last
Figure 5.12: Mean Premise RTs for M.S. during Training and Transfer phases of Experiment 5. The line drawn through the Training data is the best-fit power function (see Table 5.5 for equation). This line has been extrapolated into the Transfer phase. The other lines are specific versions of the Old/New Equation derived from the data of this experiment (see text for details). The line labelled "Old/New Equation*" represents the new version of the Old/New Equation which incorporates a change in asymptote.
Figure 5.13: Mean Total RTs for M.S. during Training and Transfer phases of Experiment 5. The line drawn through the Training data is the best-fit power function (see Table 5.5 for equation). This line has been extrapolated into the Transfer phase. The other lines are specific versions of the Old/New Equation derived from the data of this experiment (see text for details). The line labelled "Old/New Equation*" represents the new version of the Old/New Equation which incorporates a change in asymptote.
observed reaction time during Transfer (i.e., 2109.0 ms), a completely arbitrary value of 1000 ms was chosen. The derivation of the new version of the Old/New Equation can now proceed as before:

\[ T = X + N_0 P_0^C + X_n + N_n P_n^C \]

where \( P = \text{Cumulative Accuracy} \)

\[ => 2519.51 = 0 + 5058.3 (6.46)^{-0.70275} + 1000 + N_n (0.978)^{-0.70275} \]

\[ => 2519.51 - 1363.37 - 1000 = N_n \]

\[ 1.014299 \]

\[ => 153.94 = N_n \]

therefore \[ T = 0 + 5058.3 P_0^{-0.70275} + 1000 + 153.94 P_n^{-0.70275} \]

Assigning \( X_n \) the value of 1000 ms has the obvious effect of reducing the value of \( N_n \) compared to the original version of the Old/New Equation derived for this data (i.e., 1139.84 down to 153.94). This has the result of increasing the ratio of old to new productions and so increases the attenuation of learning rate. This effect is obvious in Figure 5.14, which displays the reaction times predicted by both versions of the Old/New Equation and the Premise RTs of the experimental group in Experiment 4. The most interesting feature of this figure is the vast improvement in the degree to which the Transfer performance can be accounted for by the new version of the Old/New Equation compared to the original version. This is supported by the larger \( r^2 \) value and smaller rmsd value of the new version compared to the original version (see Table 5.6). So it would appear that incorporating a change in asymptote in the Old/New Equation also goes some way towards accounting for the substantial attenuation of learning rate observed in Experiment 4. Although this particular example is a relatively contrived one, it does serve to illustrate how the derivation of the new version of the Old/New
Figure 5.14: Mean Premise RTs for the Experimental group during Training (Highlight) and Transfer (Random) phases of Experiment 4. The line drawn through the Training data is the best-fit power function (see Table 5.6 for equation). This line has been extrapolated into the Transfer phase and therefore represents an application of the Old Equation in this phase. The other lines are specific versions of the Old/New Equation derived from the data of this experiment (see text for details). The line labelled "Old/New Equation*" represents the new version of the Old/New Equation which incorporates an increase in asymptote. Error bars are confidence limits (alpha = 0.05).
Table 5.6: Parameters of power functions fitted to Premise RTs during Training and Transfer phases of Experiment 4. The function labelled “Observed” was fitted directly to the data. The function labelled “Old Equation” was fitted to values extrapolated from Training performance. Functions labelled "Old/New Equation" were fitted to values predicted by various versions of the Old/New Equation (see text for details). The function labelled "Old/New Equation*" represents the new version of the Old/New Equation which incorporates a change in asymptote (see text for details). $r^2 = \text{proportion of variance in the observed reaction time values accounted for by the predicted values.}$\msd = \text{root mean squared deviation between predicted and observed reaction time values.}
Equation results in a function which provides a better description of the data than the original version of this equation.

All of the above suggests that the combination of old and new skills may have some effect on performance asymptotes, and so have an even greater effect on observed learning rate than originally suspected. However, before this effect is considered further, it is necessary to eliminate other potential causes of the large learning rate attenuation. In particular, it is necessary to examine an assumption underlying the Old/New Equation that, if shown to be false, would substantially reduce the importance of the above discussion.

The Old/New Equation is based on the assumption that new skills will improve at the same rate as old skills (i.e., \( c \) is a constant). It is possible that this assumption is unfounded, despite the arguments presented in support of it in Chapter 4. If the assumption is incorrect, and new skills improve at a rate that is different to the improvement rate of old skills, then this may account for the observation of slower learning rates during Transfer than were predicted on the basis of the Old/New Equation. Experiment 7 was designed to examine this issue by measuring the learning rates of new and old components underlying performance of a task.
## Chapter 6  Experiment 7

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1 Introduction</td>
<td>243</td>
</tr>
<tr>
<td>6.2 Overview of Task</td>
<td>244</td>
</tr>
<tr>
<td>6.2.1 Training</td>
<td>247</td>
</tr>
<tr>
<td>6.2.2 Transfer</td>
<td>250</td>
</tr>
<tr>
<td>6.2.2.1 Flow-Route RTs</td>
<td>255</td>
</tr>
<tr>
<td>6.2.2.2 3-tank RTs</td>
<td>256</td>
</tr>
<tr>
<td>6.2.2.3 1-tank RTs</td>
<td>257</td>
</tr>
<tr>
<td>6.2.2.4 Conclusion RTs</td>
<td>258</td>
</tr>
<tr>
<td>6.2.2.5 Total RTs</td>
<td>259</td>
</tr>
<tr>
<td>6.2.3 Summary of Predictions</td>
<td>260</td>
</tr>
<tr>
<td>6.3 Method</td>
<td>260</td>
</tr>
<tr>
<td>6.3.1 Subjects</td>
<td>260</td>
</tr>
<tr>
<td>6.3.2 Stimuli</td>
<td>262</td>
</tr>
<tr>
<td>6.3.3 Apparatus and Procedure</td>
<td>267</td>
</tr>
<tr>
<td>6.3.4 Design</td>
<td>270</td>
</tr>
<tr>
<td>6.4 Results and Discussion</td>
<td>270</td>
</tr>
<tr>
<td>6.4.1 General Trends</td>
<td>270</td>
</tr>
<tr>
<td>6.4.1.1 Training</td>
<td>270</td>
</tr>
<tr>
<td>6.4.1.2 Transfer</td>
<td>271</td>
</tr>
<tr>
<td>6.4.2 Learning Rate</td>
<td>273</td>
</tr>
<tr>
<td>6.4.2.1 Flow-Route RTs</td>
<td>275</td>
</tr>
<tr>
<td>6.4.2.2 3-tank RTs</td>
<td>283</td>
</tr>
<tr>
<td>6.4.2.3 1-tank RTs</td>
<td>288</td>
</tr>
<tr>
<td>6.4.2.4 Conclusion RTs</td>
<td>293</td>
</tr>
<tr>
<td>6.4.2.5 Total RTs</td>
<td>296</td>
</tr>
<tr>
<td>6.4.3 Summary and Conclusions</td>
<td>300</td>
</tr>
</tbody>
</table>
6.1 Introduction

The purpose of Experiment 7 was to measure the learning rates of old and new skills underlying performance of a new task. The derivation of the Old/New Equation relied on the assumption that all skills are learned at the same rate. This implies that the learning rate of new skills is equal to the rate at which old skills were originally learned. Therefore if one task is practised, and then another task is performed that includes components of the previous task, the new components of the second task should be learned at the same rate as the first task was learned. In contrast, the old components will improve as part of the second task at a rate which appears to be slower than that of the new components. However, improvement of these old components will in fact be described by the tail-end of the learning function of the first task. Most of the improvement of these old components has already taken place, whereas improvement of the new components has just begun.

The crucial test of the assumption of constant learning rates is that once the learning rate of new components has been measured, the rate at which subsequent new components are learned should be equivalent to this previously measured rate. Unfortunately it is not possible to evaluate this assumption with respect to the results of the previous experiments as there is no way of separating reaction times for old and new skills. For the same reason the syllogism task could not be modified to enable this separate measurement of reaction times. Hence it was necessary to develop a new task in order to evaluate this assumption.

A new task was designed that was based loosely on the principles underlying the syllogism task. Subjects again performed the task in two phases: a Training phase where a simple version of the task was practised, and a
Transfer phase where a more complex version of the task was performed. In the simple version of the task, variable stimuli were presented to subjects to study. A number of rules were associated with particular features of the stimuli and these determined a 'conclusion' that followed. Subjects again were to indicate when they had finished processing each stimulus and then they were required to evaluate a conclusion stimulus as 'true' or 'false' on the basis of the 'premise' stimulus. Using the feedback provided at the end of each trial, subjects were to discover the rules implicit in the task and continue practising the task until the Transfer phase when the task was changed. The task was modified in the Transfer phase of the experiment to make processing more complex, whilst retaining the same set of rules. This more complex task was constructed so that it was possible to measure the time to perform skills that could be transferred from the original task (i.e., old skills) and also the time to perform new skills that were necessary in the new situation. Thus the experiment was not only designed to provide a means of examining the improvement of old and new components of a task, but also to provide a further test of the ability of the Old/New Equation to account for performance following the combination of old and new skills.

6.2 Overview of Task

In this experiment subjects were taught to perform a task which involved predicting the outcome of certain configurations of stimuli. The stimuli were figures on a computer screen which represented tanks in a fuel refinery. Figure 6.1 illustrates the basic stimulus configuration presented to subjects. Subjects were informed that liquid could flow from the small tanks labelled 1 to 4 into the large tanks, A and B. On each trial the four small tanks were shown containing a certain volume of liquid. In addition, the configuration of piping connecting tank 1 to tank 4, and tank 2 to tank 3 (labelled 'Flow-Route
Figure 6.1: Basic stimulus configuration presented in Experiment 7. Labelling in this figure was not part of the stimulus configuration.
Indicator" in Figure 6.1) was shown with liquid flowing in a particular direction. This flow route indicated whether the liquid in tanks 1 and 2 would flow into either tank 3 or 4, and then subsequently into large tank A or B. When liquid flowed from tanks 1 and 2 into tank 3, liquid did not flow from tank 2 into tank 4. Similarly, when the flow route was from tanks 1 and 2 into tank 4, liquid did not flow from tank 1 into tank 3. Thus the Flow-Route Indicator was crucial for deciding whether liquid from three tanks or one tank went into either of the large tanks.

The Flow-Route Indicator was not the only indicator of the final volumes in tanks A and B. The volume level marks in the large tanks did not correspond to the same volume levels marked in the small tanks. However there was a simple rule relating the two sets of volume levels. One volume unit in the large tanks was equal to two volume units in the small tanks.

The task involved assessing the volumes and flow route in the small tanks and deciding whether the configuration of volumes presented in the large tanks corresponded with that expected on the basis of the small tank configuration. Thus the structure of the task was similar to the syllogism task. The small tank configuration can be seen as analogous to a pair of premises, with the Flow-Route Indicator serving a similar function to the location of common elements. This function was to determine the structure of the 'conclusion' to the problem, which was the volumes in tanks A and B.

The mode of presentation of the tank problems was also similar to that of the syllogisms. The configuration displayed in Figure 6.1 was presented with the small tanks containing certain volumes of liquid and the Flow-Route Indicator revealing a particular direction of flow. Tanks A and B were presented as empty. Subjects indicated when they had finished studying the configuration
(the time to perform this step = Flow-Route RTs). The display disappeared and in its place appeared the large tanks again, this time containing liquid. The subjects then indicated whether they thought this large tank configuration was TRUE or FALSE of the small tank configuration they had just seen (the time to perform this step = Conclusion RTs).

6.2.1 Training

Subjects practised a simple version of the task during Training and a more complex version during Transfer. In the simple version each of the small tanks was always presented as containing two volume units of liquid. The only feature of the small tank configuration that varied from trial to trial was the Flow-Route Indicator. This would either indicate that liquid flowed from tank 1 to tank 4 or from tank 2 to tank 3. Therefore, following the rule described above relating small and large tank volumes, there were only two possible outcomes for tanks A and B. Either tank A would contain three volume units and tank B one unit, or vice-versa. These two possibilities were the only configurations presented during Training. Therefore subjects needed only to note whether A contained more liquid than B or vice-versa to decide the truth of the presented configuration.

There are a number of similarities between the tank task and the syllogism task in terms of the underlying logic and the mode of presentation. For instance, both tasks involve a 'premise' which includes a feature that determines the solution (i.e., common elements in the syllogism task, flow route in the tank task). Both tasks also are presented with the 'premise' first, followed by a 'conclusion'. This suggests that subjects could perform on the tank task with a similar strategy to the one described in relation to the syllogism task. Figure 6.2 illustrates the goal structure inherent in such a
Figure 6.2: Goal structure underlying solution of simple version of tank problem. Flow of control indicated by arrows.
strategy for the tank task. This goal structure is based on the one presented in Figure 3.1.

Although performance in this task has not been observed before, considering the obvious similarities with the syllogism task it seems reasonable to expect that subjects will adopt a perceptual strategy similar to that developed with syllogisms. Like the strategy described for syllogisms, the strategy described in Figure 6.2 is hierarchical in nature. Thus the goal of solving each tank problem is divided into sub-goals, and some goals need to be satisfied before others. For example, in order to derive a conclusion on the basis of the small tank configuration (goal 2 in Figure 6.2), it is necessary to determine the flow-route (goal 3). Once this has been determined, a preliminary conclusion can be constructed (goal 4). The following production rules describe how this might be achieved:

P1: IF Flow-Route = Tanks 1.2+3 flow into Tank A
    THEN set Preliminary A = sum of volumes in Tanks 1,2+3
        (= 6 units)
        and Preliminary B = volume in Tank 4
        (= 2 units)

P2: IF Flow-Route = Tanks 1.2+4 flow into Tank B
    THEN set Preliminary B = sum of volumes in Tanks 1,2+4
        (= 6 units)
        and Preliminary A = volume in Tank 3
        (= 2 units)

These conclusions are preliminary because another processing step is required (goal 5) that halves the volume levels in these conclusions in order to produce similar levels to those that appear in the presented conclusions. So, the strategy depicted in Figure 6.2 suggests that a number of processing stages must proceed before goal 2 can be achieved.
Subjects are unlikely to focus on goal 5 to a large extent because expected true conclusions need only be sufficiently accurate to predict whether A will contain more liquid than B, or vice-versa. In a similar vein, the decision underlying goal 8 need only rely on a perceptual strategy to compare the volume levels of A and B, such as the one presented in Figure 6.3.

6.2.2 Transfer

The complex version of the tank task practised during the Transfer phase of Experiment 7 had the same basic structure as the simpler version but the volume levels in all of the tanks varied to a greater extent. The volume levels in the small tanks could vary from one to four units. As a result the volume levels in the large tanks could also vary from one to four units. Thus subjects were now obliged to pay attention to the exact volume level in each tank. In addition, the false large tank configurations that were presented on half of the trials were not always simply the mirror reflection of the true configurations, as was the case in the simple version of the task. Instead, half of the false configurations represented the same gross relationship between the volume levels of A and B as the true configurations (e.g., \( A > B \)) but the exact value of the difference between the two volume levels was not the same. Thus subjects' expected true conclusions needed to be accurate in terms of absolute volume level, not just relative volume level. In addition, it was necessary for the large tank configurations that were presented to be assessed to the same degree of accuracy.

Another difference between the simple and complex versions of the task is that the small tank configurations were presented in a sequential form (see Figure 6.4). The first part of the configuration that was presented on each trial was the entire display with all tanks empty and the Flow-Route Indicator.
"Is the relationship between A & B in the expected direction?"
e.g., A > B, A < B

YES

"TRUE"

NO

"FALSE"

Figure 6.3: Decision tree underlying goal 8 in simple version of the tank task (see Figure 6.2).
Figure 6.4: Display sequence in the complex version of the tank task presented during the Transfer phase of Experiment 7. The labels of the various stages in the task represent the reaction time measures taken when subjects respond at each stage.
showing a particular flow-route. Subjects responded when this had been processed and then the three small tanks that all flowed into the same large tank were shown with particular volumes of liquid. Subjects responded when this information had been processed and then the remaining small tank was shown containing some liquid. Finally, when subjects responded that this information had been processed, the display disappeared, and in its place appeared the large tanks, each containing a certain volume of liquid. As before, subjects were required to evaluate the large tank configurations as TRUE or FALSE on the basis of the small tank configurations.

The goal structure of a strategy likely to be adopted by subjects for performing the complex version of the tank problem is presented in Figure 6.5. Most of the stages in this strategy are duplicated from the strategy in Figure 6.2 which suggests that there will be a substantial degree of transfer in this task from the simple version, at least on some parts of the task. The rationale behind presenting the small tank configurations in a sequential manner was to attempt to separate those parts of the task where transfer was complete from those parts where transfer was only partial or where there was no transfer at all. In other words, the complex version of the task was seen to be a combination of parts that relied solely on productions developed in the simple version of the task, parts that relied on the development of new productions, and parts that relied on a combination of old and new productions.

Five different reaction time measures were collected in this phase, compared to only two during the Training phase. These were: (1) the time to process the Flow-Route Indicator (Flow-Route RTs), (2) the time to process the volume levels of the three tanks (3-tank RTs), (3) the time to process the volume level in the remaining small tank (1-tank RTs), (4) the time to evaluate the large
Figure 6.5: Goal structure underlying solution of complex version of tank problem. Flow of control indicated by arrows.
tank configuration as TRUE or FALSE (Conclusion RTs), and (5) Total RTs (i.e., sum of the other four RTs). Each of these reaction time measures was expected to reflect the operation of particular stages in the strategy in Figure 6.5. For instance, it was assumed that Flow-Route RTs would reflect goals 1, 3 and 4, 3-tank RTs would reflect goals 6, 7 and 8, 1-tank RTs would reflect goals 9, 5, 10, 2 and 11. Conclusion RTs would reflect goals 12, 13 and 14, and Total RTs would reflect the achievement of all goals. The expectation therefore was that the various reaction time measures would reflect different combinations of old and new productions and this would allow comparisons of the improvement rate of old and new skills. A number of predictions based on the Old/New Equation and the strategy in Figure 6.5 will now be described. These predictions will be presented in separate sections for each of the reaction time measures.

6.2.2.1 Flow-Route RTs

The presentation of flow-route information in the Transfer phase was not substantially different from in the Training phase. The only difference was that the four small tanks were empty. In the Training phase they were all shown containing two volume units of liquid each. This difference was not expected to affect the processing of flow-route information. In other words, subjects should be able to process this information with the same productions that were developed for this purpose during Training.

Flow-Route RTs measured during Training should reflect the processing of more than just flow-route information. It can be seen in Figure 6.2 that Flow-Route RTs in the simple version of the task reflect the construction of expected true conclusions as well. This part of the task is likely to be postponed in the complex version and so will not be reflected in Flow-Route
RTs during Transfer. Therefore Flow-Route RTs during Transfer may reflect less processing than was the case during Training. That is, the processing associated with goals 2, 4 and 5 in Figure 6.2 is not necessary at this stage of the complex version of the tank task. This may then result in a speed-up in performance similar to that observed in Experiment 1 when Random Trained subjects performed with Blocked Transfer syllogisms. However, Flow-Route RTs during Transfer should still reflect the improvement of productions developed during Training which determine flow-route. If these productions continue to improve during Transfer as they did during Training, then extrapolating Training Flow-Route RTs should at least predict the learning rate of Flow-Route RTs during Transfer. The predicted Flow-Route RTs may, however, generally overestimate the Flow-Route RTs observed during Transfer, because the latter should reflect the execution of fewer productions.

6.2.2.2 3-tank RTs

As mentioned above, 3-tank RTs were expected to reflect the processing of the volume levels in the first three small tanks presented. This process is represented by goals 6, 7 and 8 in Figure 6.5. These goals involve the processing of the volume levels in each tank, adding the three separate volume levels, and signalling when this processing is complete. None of these skills were developed in the context of the simple version of the task. Therefore 3-tank RTs should reflect the development and execution of an entirely new set of productions. As discussed in the introduction to this chapter, the learning rate of new productions should be the same as the rate at which other new productions improve. This suggests that the rate at which 3-tank RTs improve should be equal to the learning rate observed during Training. Therefore measurement of the rate at which 3-tank RTs improve represents a crucial test of the assumption underlying the Old/New Equation
that the learning rate of new skills is equal to the rate at which old skills were originally learned.

An important issue underlying the prediction of equal learning rates concerns the most suitable measure of the learning rate of new productions. With respect to the present experiment, it is necessary to consider which of the two reaction time measures taken during Training is the most appropriate indicator of the rate at which new skills are learned. The assumption of equal learning rates only applies to entirely new productions. Therefore an appropriate initial measure of learning rate must reflect the same sort of processing as the Transfer task without the same productions. This will ensure that the initial estimate of learning rate will not be contaminated by the effects of more or less well-practised old skills than are involved in the Transfer task. The reaction time measure during Training that most closely reflects the same type of processing as 3-tank RTs is Flow-Route RTs. This measure reflects the development and practice of productions for processing tank volume levels, although not the same productions that are required to process the volume levels of the three tanks during Transfer. In contrast, Conclusion RTs during Training reflect pattern matching (i.e., matching presented large tank configurations with expected configurations) and so may not be the most suitable measure of the improvement of processes similar to those reflected by 3-tank RTs. Thus the prediction for 3-tank RTs is that learning rate will be equal to that observed with Flow-Route RTs during Training.

6.2.2.3 1-tank RTs

1-tank RTs were expected to reflect goals 9, 5, 10, 2 and 11 in Figure 6.5. Goal 9 is a new goal that involves processing the volume level of the small tank presented last. Goal 5 is a modified version of goal 4 in Figure 6.2. Goal
Chapter 6

4 involved the process of setting preliminary values of the expected volume levels of large tanks A and B. These preliminary values are those set prior to applying the halving algorithm (i.e., goal 5 in Figure 6.2). In the simple version of the task these preliminary values were always two volume units or six volume units. So, goal 4 would involve simply assigning these values to A and B as a function of flow-route. The new version of this goal (goal 5 in Figure 6.5) must take into account the fact that the preliminary values of the expected volume levels of A and B vary as a function of the volume levels in each of the small tanks. Therefore, in the more complex task, it is necessary on most occasions to assign values other than two or six units to these preliminary values. Hence this goal should include productions for calculating these particular values. Thus goal 5 in the complex strategy should be considered as mainly involving a new set of productions. Goals 10, 2 and 11 are the same as goals 5, 2 and 6 in the simple version of the task (see Figure 6.2). Therefore 1-tank RTs should reflect the combination of old skills (i.e., goals 10, 2 and 11) with new skills (i.e., goals 9 and 5). As a result, the rate of improvement in 1-tank RTs should be slower than that observed in Flow-Route RTs during Training. The amount of this learning rate attenuation should be predicted by application of the Old/New Equation.

6.2.2.4 Conclusion RTs

The strategy depicted in Figure 6.5 shows that the goals involved in processing the large tank configurations in the complex task are the same as those that underlie the performance of this operation in the simple version of the task. It was suggested earlier that in the simple version of the task a perceptual strategy may be used in determining the truth of presented large tank configurations (e.g., Figure 6.3). This perceptual strategy was said to involve a simple check for whether tank A contained more liquid than tank B,
or vice-versa. However, because it is necessary in the complex version of the task for subjects to pay attention to the exact volume levels in the large tanks, a more refined strategy than the earlier perceptual strategy is required at this stage of processing. Therefore goal 13 in Figure 6.5 will need to be a more refined version of goal 8 in Figure 6.2. Instead of simply checking for a perceptual difference between the volume levels of tanks A and B, the exact levels will need to be processed and compared with the expected levels. Therefore it will be necessary to develop new productions to achieve this. As a result, it is expected that Conclusion RTs during Transfer should reflect the combination of old and new skills and so be described by an application of the Old/New Equation. This would result in Conclusion RTs improving during the Transfer phase at a slower rate than during Training. However, the Transfer learning rate would be faster than extrapolated Training performance.

6.2.2.5 Total RTs

The strategy in Figure 6.5 describes performance in the complex version of the tank task. From the above discussion of the various features of this strategy it is obvious that it contains most of the skills developed in the context of the simple version of the task plus a new set of skills. Therefore Total RTs during Transfer were expected to reflect the combination of old and new skills and so be described by an application of the Old/New Equation. As a result Total RTs during Transfer should improve at a slower rate than during Training but at a faster rate than extrapolated Training performance.
6.2.3 Summary of Predictions

There are three basic outcomes predicted for Experiment 6 (see Table 6.1). (1) Where performance in the Transfer phase relies solely on skills developed in the Training phase (Flow-Route RTs), the learning function during Transfer will fall along a curve extrapolated from Training performance (or possibly below this curve but with the same learning rate). In other words, learning rate in the Transfer phase will appear to be slower than during Training and this will be a result of the improvement in these skills being at the tail-end of a function that began at the start of Training. (2) Where Transfer performance involves the combination of old and new skills (1-tank RTs, Conclusion RTs and Total RTs), learning rate during Transfer will be described by the Old/New Equation. (3) Where performance during Transfer is a result of the development of new skills only (3-tank RTs), learning rate will be equal to the rate at which productions were developed during Training.

6.3 Method

6.3.1 Subjects

Sixteen volunteers from the University of Western Australia Psychology Department participated in this experiment for course credit or $5. All subjects satisfied the learning criterion of an error rate not in excess of 25% in the latter half of the Training trials.
<table>
<thead>
<tr>
<th>Performance Measure</th>
<th>Productions</th>
<th>Equation Describing Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flow-Route RTs</td>
<td>old</td>
<td>Old Equation</td>
</tr>
<tr>
<td>3-tank RTs</td>
<td>new</td>
<td>equation with same learning rate as function describing Flow-Route RTs during Training</td>
</tr>
<tr>
<td>1-tank RTs</td>
<td>old &amp; new</td>
<td>Old/New Equation</td>
</tr>
<tr>
<td>Conclusion RTs</td>
<td>old &amp; new</td>
<td>Old/New Equation</td>
</tr>
<tr>
<td>Total RTs</td>
<td>old &amp; new</td>
<td>Old/New Equation</td>
</tr>
</tbody>
</table>

Table 6.1: Summary table of predictions for performance during the Transfer phase of Experiment 7. The 'Productions' column summarises the predictions concerning the type of productions that will be executed during the phases of the task reflected by the various performance measures (e.g., Flow-Route RTs are predicted to reflect the execution of old productions only). Further predictions as to which equations will provide the best description of performance during Transfer are summarised in the third column (e.g., Flow-Route RTs are predicted to be described best by a version of the Old Equation).
6.3.2 Stimuli

During Training subjects were presented with tank configurations like the one in Figure 6.6. The structure was designed to depict a series of tanks that contained certain volumes of liquid. The tanks were connected by a series of pipes through which the liquid flowed in certain prescribed ways. At the beginning of each trial tanks 1 to 4 were shown with certain volumes of liquid. During Training this volume was constant across all trials (i.e., 2 units). In addition to the volumes in the tanks certain pipes were shown to contain liquid also. This designated the route that the liquid could flow on that particular trial. Two different routes were used in this experiment. Figure 6.6 shows one of these routes - the liquid in tanks 1 and 2 flows into tank 4. The other route involved liquid flowing from tanks 1 and 2 into tank 3. The direction in which the liquid was shown to flow then determined the volumes of liquid that would be expected in the tanks in the lower half of the configuration. These tanks were larger than tanks 1 to 4. The rules underlying the configuration were such that one unit volume in these larger tanks was equal to twice the volume of one unit in the smaller tanks (ignoring the volume of liquid shown in the pipes). Thus in the configuration depicted in Figure 6.6 the liquid in tanks 1, 2 and 4 flows into the larger tank labelled B. There are six volume units of liquid in these three smaller tanks and so this leads to half that number of volume units in tank B, that is, three. Only tank 3 flows into tank A in this configuration. Tank 3 contains two volume units so this will lead to one volume unit in tank A.

Two hundred and forty trials were presented to subjects during the Training phase of this experiment. Half of these trials involved the flow route depicted in Figure 6.6 (i.e., tanks 1 and 2 flow into tank 4). The other half of these trials involved the other flow route (i.e., tanks 1 and 2 flow into tank 3). In
Figure 6.6: Example of stimulus configuration presented in the simple version of the tank task in Experiment 7. Labelling in this figure was not part of the stimulus configuration.
addition half of the trials were presented with "true" configurations of liquid volumes in tanks A and B. These were configurations that showed volumes consistent with what would be expected from use of the above rule (the configuration of tanks A and B shown in Figure 6.6 is "true" because it is consistent with the configuration of tanks 1 to 4). The other half of the trials were presented with "false" configurations of liquid volumes in tanks A and B. These were volumes that were opposite to what would be expected from the configuration of small tanks. These also were the "true" tank volumes for the opposite configuration of tanks 1 to 4. For example, the false configuration of tanks A and B for the small tank configuration in Figure 6.6 would show tank A with three units and tank B with one unit. This in turn would be the true configuration for the situation where tanks 1 and 2 flow into tank 3. In summary then, subjects were presented with small tank configurations where tanks 1 and 2 either flowed into tank 3 or tank 4, and large tank configurations where tank A contained three units and tank B contained one unit or vice-versa. Thus there were only four different types of stimulus and these were presented 60 times each.

During Transfer the situation was changed so that the volumes shown in tanks 1 to 4 were no longer always two units. In this phase the volumes of each of these tanks could range from one to four units. As a result the volumes of tanks A and B could also range from one to four. This situation leads to 256 possible configurations in the small tanks. The same rule relating volumes in tanks 1 to 4 to volumes in tanks A and B that was used during Training was also used in this phase of the experiment. As a result some of the total number of possible configurations would lead to volumes with half units. These configurations were excluded from this experiment, as were configurations that led to volumes in tanks A and B that were greater than four units. A total of 100 legal configurations were selected as base items for
the Transfer phase. An example of a legal configuration is shown in Figure 6.7.

Four hundred Transfer items were derived from the 100 base items. The 100 base items were presented twice each with true configurations of tanks A and B. The other 200 items involved false configurations of these larger tanks. These false configurations were of two types. One type was simply the opposite of the true configuration. For example, consider the configuration shown in Figure 6.7. The large tank configuration shown has tank A with three units and tank B with two units which is true of the small tank configuration. An example of the first type of false configuration with respect to this example would have tank A with two units and tank B with three units. Thus this first type of false configuration is similar to the false configurations in the Training phase (i.e., they are both the reverse of the true configurations) except now the volumes in tanks A and B are not simply either one or three but can range from one to four.

The second type of false configuration of the large tanks involved the same relative volumes as the true configurations but the difference between the volumes of tanks A and B was altered by one volume unit. Thus if a true configuration had one large tank with a greater volume than the other large tank, then this relationship was retained but the difference between the two volumes was altered by one unit. For example, a false configuration of this type corresponding to the small tank configuration in Figure 6.7 could have tank A with four units and tank B with two units. Thus tank A still has a greater volume than tank B though now the difference in volumes has increased by one unit. False configurations of this type were balanced as to whether they involved an increase or a decrease in the volume difference between the tanks.
Figure 6.7: Example of stimulus configuration presented in the complex version of the tank task in Experiment 7. Labelling in this figure was not part of the stimulus configuration.
In addition to the balance of true and false configurations of tanks A and B, each item was presented with one of two flow routes. Two hundred items involved tanks 1 and 2 flowing into tank 3 (e.g., Figure 6.7). The other 200 items had tanks 1 and 2 flowing into tank 4.

6.3.3 Apparatus and Procedure

The tank displays were presented on a CRT and responses were made on a keyboard. All aspects of the experiment were controlled by a PDP 11/73 computer.

Subjects were tested individually in one session that lasted between 75 and 90 minutes. Subjects were informed that the experiment would involve them learning about a particular aspect of a fuel refinery. The part of the refinery that the experiment would focus on would involve six tanks and a series of pipes that connected these tanks to each other. The subjects' task was to learn the rules that determined how fuel moved through this system. Subjects could then use these rules to judge whether configurations of one part of the system were consistent with a presented outcome in another part of the system.

During the Training phase each item was presented as follows: The tank display was presented on the screen of the CRT with each of the four small tanks (tanks 1 to 4) filled with two units of liquid. The direction in which the liquid flowed from tanks 1 and 2 to tanks 3 and 4 was indicated by the presence of liquid in particular pipes. In this first part of each trial the large tanks, A and B, were shown to be empty. Subjects were instructed to press the "READY" button (space bar of the keyboard) when they had processed the display. The display was visible for a maximum of ten seconds. When a subject pressed "READY", or ten seconds had elapsed without a response,
the whole display disappeared. The lower part of the display was then redrawn on the screen. Tanks A and B were shown with configurations of either one and three volume units respectively, or three and one. Subjects were then to press "TRUE" (the 'z' key of the keyboard) or "FALSE" (the 'j' key of the keyboard) dependent on how they felt the configuration of tanks A and B corresponded to the configuration of small tanks that preceded it. Again subjects were given a maximum of ten seconds to respond. When "TRUE" or "FALSE" had been pressed, or ten seconds elapsed without a response, the display disappeared. Following this subjects were provided with feedback on the screen for two seconds. This consisted of a sentence of the form: "Correct/Incorrect, volumes were True/False", dependent on the subjects' response. No feedback was provided if subjects did not press "TRUE" or "FALSE". After the feedback disappeared from the screen the next item was presented automatically.

Subjects were given two practice trials at the beginning of the experiment. They were presented in the same fashion as the experimental trials. The flow route in each practice trial was the same as that depicted in Figure 6.6. The configuration of the large tanks was true in the first trial and false in the second trial. Subjects were provided with feedback after each of these trials. Following these practice trials subjects were presented with the 240 Training trials. At the end of the Training phase subjects were given two minutes rest during which they were instructed on the screen that the next phase of the experiment (Transfer phase) would be similar to the previous phase though much longer and volume levels would vary more. After the rest period, the computer automatically presented the 400 trials in the Transfer phase.

During the rest period subjects were informed that in the second phase of the experiment the volumes of liquid in the small tanks could vary from each
other and in each problem. Subjects were told that, in order to reduce processing load, the liquid volumes would be presented sequentially and that they would control when these appeared. Subjects were then instructed on how to respond in this phase of the experiment.

Each trial began with all six tanks empty. Only the flow route was indicated with liquid in particular pipes (e.g., see Figure 6.4). Subjects were instructed to press "READY" when they had processed this information. A maximum of ten seconds was provided for this response. When "READY" had been pressed, or ten seconds had elapsed without a response, three of the four small tanks were shown containing liquid. In each trial the three tanks that were first shown containing liquid were the three that all flowed into the same large tank. For example, in Figure 6.7 the liquid in tanks 1, 2 and 3 flow into large tank A. Thus these three tanks would have been the first to show liquid in this particular configuration. Subjects were given a maximum of 20 seconds to process the configuration of the three tanks and then were required to press "READY" again. When this was pressed, or 20 seconds elapsed without a response, the remaining small tank was shown containing liquid. In the configuration depicted in Figure 6.7 this would have been tank 4. Subjects were allowed a maximum of ten seconds to process this tank and again were required to press "READY". When this was pressed, or ten seconds elapsed without a response, the entire display disappeared. The lower part of the display containing the large tanks was then redrawn. This time the large tanks were shown containing liquid. Subjects were required to press "TRUE" or "FALSE" dependent on whether the configuration of the large tanks was consistent with the configuration of the small tanks. A maximum of ten seconds was allowed for this response. When a response was made, or ten seconds elapsed without a response, feedback was provided for two seconds. Following the feedback the next item was
presented automatically. One minute rest periods were provided every 50 trials.

6.3.4 Design

In both the Training and Transfer phases of this experiment the presentation order of items with one flow route compared to items with the other, and of true and false items was random.

6.4 Results and Discussion

Data in the Training phase of Experiment 7 was analysed in six blocks of 40 trials. In the Transfer phase the data was analysed in 10 blocks of 40 trials. Only data from trials with correct responses were analysed.

General trends in the data will be presented and discussed first, followed by comparisons of learning rate between Training and Transfer performance as estimated from each of the reaction time measures.

6.4.1 General Trends

6.4.1.1 Training

A number of analyses of variance were performed on the Training data. A 6 (Training block) x 2 (Flow Route) ANOVA was used to analyse Flow-Route RTs. Two 6 (Training block) x 2 (Flow Route) x 2 (conclusion type) ANOVAs were used to analyse Conclusion RTs and Accuracy.
Chapter 6

During Training subjects showed significant improvement in all measures of performance in the simple version of the tank task. Flow-Route RTs improved from 2454 ms in Block 1 to 833 ms in Block 6 ($F(5,75) = 14.55$, $p<0.05$). Conclusion RTs improved from 1245 ms in Block 1 to 838 ms in Block 6 ($F(5,75) = 10.98$, $p<0.05$). Accuracy also improved, from 93.12% in Block 1 to 99.38% in Block 6 ($F(5,75) = 3.60$, $p<0.05$).

Flow-route had no effect on Flow-Route RTs ($F(1,15) = 1.33$, $p>0.05$). Thus subjects did not appear to process flow-route information with a bias similar to the syllogism-type bias observed in the previous experiments. However flow-route did have a significant effect on both Conclusion RTs and Accuracy. Conclusion RTs were faster when flow-route was from tank 1 to tank 4 (927 ms) than when it was from tank 2 to tank 3 (974 ms) ($F(1,15) = 5.97$, $p<0.05$). Subjects were also more accurate when flow-route was from tank 1 to tank 4 (98.02% vs. 94.79%, $F(1,15) = 10.62$, $p<0.05$). It is unclear what type of strategy for processing large tank configurations could cause this difference.

Subjects processed True large tank configurations faster than False configurations (897 ms vs. 1003 ms, $F(1,15) = 10.14$, $p<0.05$). This result is similar to the bias towards True conclusions observed in the syllogism experiments, suggesting that this may be a general feature of True/False decisions of this type. This bias, however, did not result in any Accuracy difference between the two types of configurations ($F(1,15) = 2.13$, $p>0.05$).

### 6.4.1.2 Transfer

A number of analyses of variance were performed on the Transfer data. Three 10 (Transfer block) x 2 (Flow Route) ANOVAs were used to analyse Flow-
Route RTs, 3-tank RTs and 1-tank RTs. Two 10 (Transfer block) x 2 (Flow Route) x 2 (conclusion type) ANOVAs were used to analyse Conclusion RTs and Accuracy.

During the Transfer phase, the only significant effect on Flow-Route RTs, 3-tank RTs and 1-tank RTs was that of improvement with practice. Flow-Route RTs improved from 1690 ms in Block 1 to 518 ms in Block 10 (F(9,135) = 6.13, p<0.05). 3-tank RTs improved from 3288 ms in Block 1 to 1199 ms in Block 10 (F(9,135) = 17.30, p<0.05). 1-tank RTs improved from 3020 msecs in Block 1 to 906 ms in Block 10 (F(9,135) = 20.20, p<0.05).

Conclusion RTs and Accuracy also showed significant improvement with practice. Conclusion RTs were reduced from 1629 ms in Block 1 to 873 ms in Block 10 (F(9,135) = 34.16, p<0.05) and Accuracy improved from 73.91% in Block 1 to 87.34% in Block 10 (F(9,135) = 7.34, p<0.05).

The advantage of True large tank configurations over False configurations found during Training was also observed in Transfer. True Conclusion RTs (1020 ms) were faster than False Conclusion RTs (1107 ms) (F(1,15) = 9.12, p<0.05). True configurations (91.28%) were also more accurately processed than False configurations (82.72%) (F(1,15) = 23.39, p<0.05). These results suggest that subjects retained the bias for True configurations in their processing of large tank configurations during Transfer.

One interesting result that will be presented now but discussed in further depth below concerns overall improvement during the Transfer phase of the experiment. The Total time to perform the complex version of the task was reduced from 9628 ms in Block 1 to 3496 ms in Block 10 (F(9,135) = 28.92, p<0.05). However, inspection of the data reveals a non-uniform pattern of
improvement. In the first five blocks of Transfer there was a significant overall improvement with practice \((F(4,60) = 28.77, p<0.05)\). This improvement was such that the mean Total RT of each block was less than that of the previous block. In contrast, the last five blocks of the Transfer phase showed no significant improvement with practice \((F(4,60) = 0.17, p>0.05)\). It appears that during this second half of Transfer, most of the subjects were unable to improve their performance beyond a certain point and so responded at more-or-less the same speed throughout the remainder of the experiment. This is commonly referred to as asymptotic performance. It is unlikely though that these results reflect improvement reaching asymptote. Asymptotic performance normally takes years of dedicated practice to achieve (Anderson, 1982, 1989; Newell and Rosenbloom, 1981). The subjects in this experiment practised the simple version of the task for approximately 20 minutes and only practised the complex version for just over an hour. So rather than having reached the point beyond which no benefit is gained by further practice, the subjects were more likely to have suffered from fatigue or boredom or both. Certainly many subjects complained that, despite the frequent rest periods, the experiment was too long. This possibility will be discussed in more detail below in relation to the effect of false asymptotes on the measurement of learning rates.

6.4.2 Learning Rate

Including non-zero asymptotes in power functions fitted to learning data results in power functions with increased learning rates. This can be understood by considering learning rate as a measure of how quickly performance reaches asymptote. Consider the case of two power functions, one with an asymptote of 100 ms, the other with an asymptote of 200 ms, that are fitted to the same set of practice data. The assumption underlying the
first power function is that once subjects perform with a reaction time of 100 ms, their performance will no longer improve with practice. The same assumption applies to the power function with an asymptote of 200 ms - when performance reaches this level, further practice will not result in further improvement.

The interesting difference between the two power functions is in terms of how they describe the rate of improvement in the practice data. If the data is really heading for an asymptote of 100 ms but the asymptote is not apparent (i.e., the amount of practice has not been sufficient to reach asymptote), then fitting a power function with an asymptote of 100 ms will result in an accurate estimate of the learning rate exhibited in the data. However, if a power function with an asymptote of 200 ms is fitted to this data, the function will describe the improvement as if performance in the final stages of practice is closer to asymptote than is really the case. As a result performance will appear to have reached asymptote sooner than is the case with the power function with a 100 ms asymptote. In other words, learning rate will be estimated as faster when a power function with a 200 ms asymptote is fitted to the data than when a power function with a 100 ms asymptote is used.

The general rule that can be derived from the above is that artificially high asymptotes inflate estimates of learning rate. Therefore if subjects reach a point beyond which improvement will not result from further practice for reasons other than having reached their true asymptote, such as fatigue or lack of motivation, then learning rate will appear to be faster than is actually the case. For this reason, and the fact that Total RTs revealed no significant improvement in the second half of Transfer, estimates of learning rate in this experiment will be reported for the full ten blocks of Transfer, and for the first six blocks. If artificially high asymptotes affected learning rate estimates
then the estimates based on ten blocks of practice will be higher than the corresponding estimates based on the first six blocks of practice.

6.4.2.1 Flow-Route RTs

The predicted outcome for Flow-Route RTs during Transfer was that performance would reflect the execution of old productions only and so would be best described by a version of the Old Equation (see Table 6.1). Parameters of various power functions fit to the Training and Transfer Flow-Route RTs are presented in Table 6.2. The first thing to note is that functions that were fitted directly to the data of the first six blocks of Transfer and to data of all ten blocks have learning rates which are faster than the learning rate exhibited during Training (i.e., -0.840 & -1.193 vs. -0.756). This is a result that is prohibited by the Old/New Equation if any transfer is involved between tasks. However, inspection of Figure 6.8, which displays the mean Training and Transfer data, reveals what has caused this unexpected increase in learning rate.

The mean Flow-Route RT in the first block of Transfer is substantially slower than the mean RTs of the remaining blocks, and there is also a much greater spread of reaction time values around this particular mean than is observed elsewhere during this phase. The size of the confidence interval in the first block is 966 ms, whereas in the remaining blocks the confidence intervals range in size from 88 ms to 188 ms. Confidence intervals of the magnitude observed in the first block were not observed in any of the previous experiments (the magnitude of confidence intervals observed with Premise RTs during Transfer in Experiments 1, 2, 3, 4 and 6 ranged from 212 ms to 631 ms). An inspection of the individual subjects' means that underlie this particular value revealed one subject whose mean Flow-Route RT of 8249 ms
Table 6.2: Parameters of power functions fitted to Flow-Route RTs during Training and Transfer phases of Experiment 7. Functions were fitted to all ten blocks and to the first six blocks of the Transfer data. The "Complete Data Set" refers to the Transfer data with the outlier included in the first block of Transfer. "Outlier Removed" refers to the Transfer data without this outlier. Functions labelled "Observed" were fitted directly to the data. Functions labelled "Old Equation" were fitted to values extrapolated from Training performance. Functions labelled "Old/New Equation" were fitted to values predicted by various versions of the Old/New Equation (see text for details). Functions labelled "Old/New Equation*" were fitted to values predicted by the new version of the Old/New Equation which incorporates a change in asymptote (see text for details). $r^2$ = proportion of variance in the observed reaction time values accounted for by the predicted values. rmsd = root mean squared deviation between predicted and observed reaction time values.

<table>
<thead>
<tr>
<th></th>
<th>Parameters</th>
<th>Goodness of Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>X</td>
<td>N</td>
</tr>
<tr>
<td>Training</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>295</td>
<td>2025.50</td>
</tr>
<tr>
<td>Transfer (Blocks 1-10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complete Data Set</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>350</td>
<td>711.71</td>
</tr>
<tr>
<td>Old Equation</td>
<td>15</td>
<td>766.49</td>
</tr>
<tr>
<td>Old/New Equation</td>
<td>225</td>
<td>1272.20</td>
</tr>
<tr>
<td>Old/New Equation*</td>
<td>265</td>
<td>1243.10</td>
</tr>
<tr>
<td>Outlier Removed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>215</td>
<td>727.86</td>
</tr>
<tr>
<td>Old Equation</td>
<td>15</td>
<td>766.49</td>
</tr>
<tr>
<td>Old/New Equation</td>
<td>90</td>
<td>1057.00</td>
</tr>
<tr>
<td>Transfer (Blocks 1-6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complete Data Set</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>365</td>
<td>827.38</td>
</tr>
<tr>
<td>Old Equation</td>
<td>15</td>
<td>766.49</td>
</tr>
<tr>
<td>Old/New Equation</td>
<td>225</td>
<td>1272.20</td>
</tr>
<tr>
<td>Old/New Equation*</td>
<td>265</td>
<td>1243.10</td>
</tr>
<tr>
<td>Outlier Removed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>215</td>
<td>801.32</td>
</tr>
<tr>
<td>Old Equation</td>
<td>15</td>
<td>766.49</td>
</tr>
<tr>
<td>Old/New Equation</td>
<td>90</td>
<td>1057.00</td>
</tr>
</tbody>
</table>
Figure 6.8: Mean Flow-Route RTs during Training and Transfer phases of Experiment 7. The mean of the first block of Transfer includes the outlier. The curve drawn through the Training data is the best-fit power function for this data (see Table 6.2 for equation) and has been extrapolated into the Transfer phase (labelled "Old Eqn"). The other curves drawn in the Transfer phase are versions of the Old/New Equation (see text for equations). Error bars represent confidence limits (alpha = 0.05).
was much slower than those of the rest of the group (mean = 1251 ms, range = 669 to 2400 ms). This subject’s performance during the rest of this phase was not detectably different from the performance of the rest of the group. It appears that this subject was taken aback by the change of task but quickly came to terms with the new task and continued to improve in a similar fashion to other subjects.

The effect of removing the outlying mean from calculation of the mean Flow-Route RT for the first block of Transfer is displayed in Figure 6.9. It is clear from this figure that the spread of reaction time values in this block without the outlier (confidence interval = 256 ms) is similar to the spread in the remaining blocks. The effect of removing this outlier on the estimates of learning rate is evident in Table 6.2. The power functions fit to the revised data set based on ten blocks and six blocks of practice have learning rates which are smaller than that observed during Training (i.e., -0.503 & -0.676 vs. -0.756).

The effect of an outlier on learning rate estimates underlines the importance of careful inspection of practice data. It also illustrates the fact that factors other than the inappropriate location of initial practice trials and the combination of old and new skills (see Chapter 4) can affect observed learning rate. Performance may be slowed in the initial blocks of Transfer for reasons other than a lack of appropriate productions. Factors such as disrupted attention or shock from a dramatic change in stimulus characteristics are factors that will elevate reaction time. However, subjects should be able to recover quickly from the effects of these factors. Thus the dramatic improvement that results from this recovery will be superimposed on the more gradual improvement characteristic of skill acquisition, and therefore will result in artificially high learning rates.
Figure 6.9: Mean Flow-Route RTs during Training and Transfer phases of Experiment 7. The mean of the first block of Transfer was calculated with the outlier removed. The curve drawn through the Training data is the best-fit power function for this data (see Table 6.2 for equation) and has been extrapolated into the Transfer phase (labelled "Old Eqn"). The other curve drawn in the Transfer phase is a version of the Old/New Equation (see text for equation). Error bars represent confidence limits (alpha = 0.05).
The effect of a false asymptote on learning rate was not apparent in Flow-Route RTs. The estimates of learning rate based on ten blocks of practice were in fact slower than the estimates based on the first six blocks, with and without the outlier (ten blocks: -0.840 & -0.503 vs. 1st six blocks: -1.193 & -0.676).

The improvement in Flow-Route RTs during Transfer was compared with the improvement predicted on the basis of Training performance. A second examination of improvement in Flow-Route RTs concerned the observation that, despite the removal of the outlier from the first block of Transfer, there still appeared to be a significant increase in Flow-Route RTs with the change of task. For this reason two versions of the Old/New Equation were derived, one for the data set with the outlier (\( T = 295 + 2025.5 \, P_0^{-0.756} + 719.63 \, P_n^{-0.756} \)), the other with the outlier removed (\( T = 295 + 2025.5 \, P_0^{-0.756} + 369.87 \, P_n^{-0.756} \)). In addition, the data set with the outlier exhibited an increase in asymptote compared to Training performance. Hence the new version of the Old/New Equation which incorporates a change in asymptote can be applied (\( T = 295 + 2025.5 \, P_0^{-0.756} + 55 + 675.87 \, P_n^{-0.756} \)). The improvement predicted on the basis of these versions of the Old/New Equation was compared with the Transfer data.

Table 6.2 shows that in all cases the learning rate predicted by extrapolating Training performance (i.e., Old Equation) is a lot slower than was observed during Transfer. The learning rates predicted by the various versions of the Old/New Equation appear to be closer to the observed rates. However, consideration of the extent to which the Old Equation and the Old/New Equation account for the Transfer data reveals a different result. Despite the fact that the various versions of the Old/New Equation account for a greater proportion of the variance in the Transfer data than the Old Equation,
inspection of the rmsd values in Table 6.2 and the curves in Figures 6.8 and 6.9 shows that improvement during Transfer was predicted more accurately by extrapolating Training performance than by applying the various versions of the Old/New Equation.

Figures 6.8 and 6.9 indicate that the ability of the various versions of the Old/New Equation to account for a large proportion of the variance in the Transfer Flow-Route RTs is, to a large degree, a result of the first block of Transfer. With and without the outlier, performance in this first block is substantially slower than that predicted on the basis of Training performance. It is also much slower than performance throughout the rest of this phase.

The various versions of the Old/New Equation all describe improvement from an initial slow reaction time. The Old Equation, on the other hand, describes slower improvement from a faster initial reaction time. Therefore, since most of the variance in the Transfer phase results from the initial slow Transfer performance, any function that predicts an initial slow performance and subsequent improvement (e.g., the Old/New Equation) will account for more variance than a function that describes a more gradual improvement (e.g., the Old Equation). However, it is clear from both Figure 6.8 and Figure 6.9 that extrapolating Training performance (Old Equation) predicts all but the first block of Transfer performance more accurately than any of the versions of the Old/New Equation. This result is reflected by the smaller rmsd values for the Old Equation than the Old/New Equation.

It is possible that something like the 'shock' effect mentioned above may have given rise to this set of results. Alternatively this increased reaction time during the first block of Transfer may be associated with a reorganisation of productions. As was described in §6.2.2.1, some of the productions that were executed at this stage in the simple version of the task can now be
postponed until later in the complex version. The subject with the outlying reaction time may have been affected to a large extent by either of these factors. However, the other subjects may also have been affected to some extent, resulting in an artificially elevated mean reaction time for the first Transfer block. Subjects could then have recovered quickly from this effect and then continued to improve on the Flow-Route RTs at the same rate as was observed during Training. In other words, except for the first block of Transfer, Flow-Route RTs during Transfer appear to have improved in a manner consistent with a continuation of the learning process begun in Training. This supports the prediction that at this stage of the task, subjects could perform with productions which were developed during Training.

Although the Old Equation provided the best account of Flow-Route RTs during the Transfer phase, performance in the middle of this phase appeared to be faster than was predicted by this equation. As can be seen in Figure 6.9, performance in blocks 5, 6, 7 and 8 is significantly faster than performance predicted by extrapolating Training performance. This result is consistent with performance in this phase being a result of the execution of less productions than were applied during Training. This was predicted in the introduction to this chapter, where it was suggested that some of the processing measured by Flow-Route RTs during Training is likely to be postponed to later in each trial during Transfer. As a result Flow-Route RTs during Transfer would reflect less processing during the second phase and therefore be faster than is predicted by the continued improvement of all the productions developed during Training.
6.4.2.2 3-tank RTs

The expectation for 3-tank RTs was that they would reflect the development and improvement of new productions. Thus this measure was considered to provide a crucial test of the assumption underlying the Old/New Equation that the learning rate of new skills is a constant. Therefore, according to this assumption, learning rate for 3-tank RTs should be the same as the rate observed during Training. The learning rate of Flow-Route RTs during Training was -0.756. Table 6.3 shows that the rate of improvement of 3-tank RTs was faster than this value, both when learning rate was estimated from ten blocks of practice (-1.215) and when it was estimated from the first six blocks of Transfer (-1.348). There was no evidence that this increased learning rate was a result of artificially slowed initial performance, like that found with Flow-Route RTs - performance improved gradually from the first block of Transfer. There was also no evidence that learning rate was affected by a false asymptote - learning rate was not substantially different when estimated from ten or six blocks.

A number of power functions were derived in order to provide a further test of the assumption of constant learning rate with the 3-tank-RT data. Three of these functions were particular applications of the Old/New Equation. The Old/New Equation does not predict an increase in learning rate. However, performance in the initial blocks of Transfer was significantly slower than final Training performance (Flow-Route RTs), which suggests that examining the ability of the Old/New Equation to account for the improvement in 3-tank RTs would be an interesting exercise, especially in the absence of any obvious explanation for the increased learning rate. The version of the Old/New Equation derived for this set of data is \( T = 295 + 2025.5 P_0^{-0.756} + 1991.12 P_n^{-0.756} \). In addition, two examples of the new
### Table 6.3: Parameters of power functions fitted to 3-tank RTs during Transfer phase of Experiment 7.

Functions were fitted to all ten blocks and to the first six blocks of the Transfer data. Functions labelled "Observed" were fitted directly to the data. Functions labelled "New" were designed to describe improvement of new skills and include the learning rate and asymptote observed in Flow-Route RTs during Training (see text for details). Functions labelled "Old/New Equation" were fitted to values predicted by a version of the Old/New Equation (see text for details). Functions labelled "Old/New Equation*" were fitted to values predicted by the new version of the Old/New Equation which incorporates a change in asymptote (see text for details). Functions labelled "Old Equation" were fitted to values predicted by a version of the Old Equation (see text for details). \( r^2 \) = proportion of variance in the observed reaction time values accounted for by the predicted values. \( \text{rmsd} \) = root mean squared deviation between predicted and observed reaction time values.

<table>
<thead>
<tr>
<th></th>
<th>Parameters</th>
<th>Goodness of Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>X</td>
<td>N</td>
</tr>
<tr>
<td><strong>Transfer (Blocks 1-10)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>1080</td>
<td>1444.10</td>
</tr>
<tr>
<td>New</td>
<td>295</td>
<td>2381.61</td>
</tr>
<tr>
<td>Old/New Equation</td>
<td>300</td>
<td>2468.90</td>
</tr>
<tr>
<td>Old/New Equation*</td>
<td>1065</td>
<td>1864.10</td>
</tr>
<tr>
<td>Old Equation</td>
<td>15</td>
<td>766.49</td>
</tr>
<tr>
<td><strong>Transfer (Blocks 1-6)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>1135</td>
<td>1393.90</td>
</tr>
<tr>
<td>New</td>
<td>295</td>
<td>2381.61</td>
</tr>
<tr>
<td>Old/New Equation</td>
<td>300</td>
<td>2468.90</td>
</tr>
<tr>
<td>Old/New Equation*</td>
<td>1180</td>
<td>1760.50</td>
</tr>
<tr>
<td>Old Equation</td>
<td>15</td>
<td>766.49</td>
</tr>
</tbody>
</table>
version of the Old/New Equation which incorporates a change in asymptote were derived for this data set. The version for blocks 1-10 is $T = 295 + 2025.5 \, P_0^{-0.756} + 785 + 1366.55 \, P_n^{-0.756}$. The version for blocks 1-6 is $T = 295 + 2025.5 \, P_0^{-0.756} + 840 + 1322.79 \, P_n^{-0.756}$. As before, these functions were designed to describe improvement of both old and new skills.

Another function was derived as a direct test of how well a function with the same learning rate as that observed during Training could account for the improvement in 3-tank RTs. Thus the learning rate (i.e., $c$) of this function was -0.756. This function therefore was designed to describe improvement in new skills only. This function also had the same asymptote as that observed in Flow-Route RTs during Training (i.e., $X = 295$). The reason for including this value is the assumption that it represents an estimate of the minimum reaction time to perform this type of task. Although a less-than-perfect estimate for the Transfer task, it is as good as any. The final parameter of this function was the intercept value ($N$). The best estimate for this parameter was calculated by subtracting the asymptote estimate from the mean 3-tank RT of the first block of Transfer:

\[
T = X_n + N_n \, P_n^c
\]

\[
3288.37 = 295 + N_n \, (0.739)^{-0.75588}
\]

\[
2381.61 = N_n
\]

therefore $T = 295 + 2381.61 \, P_n^{-0.756}$

Table 6.3 and Figure 6.10 illustrate how well these functions account for the 3-tank RT data. Figure 6.10 does not indicate clearly which of the functions provides the better fit to the data. However, Table 6.3 shows that the function with the same learning rate as observed in Flow-Route RTs during Training
Figure 6.10: Mean 3-tank RTs during Transfer phase of Experiment 7. Curve labelled "Old/New Eqn" represents a version of the Old/New Equation (see text for equation). Curve labelled "Old/New Eqn*" represents the new version of the Old/New Equation which incorporates a change in asymptote (see text for equation). Curve labelled "New" represents a function designed to describe improvement of new skills (see Table 6.3 and text for equation details). Curve labelled "Old Eqn" represents the power function that provided the best fit to Flow-Route RTs during Training. This curve has been extrapolated into the Transfer phase. Error bars are confidence limits (alpha = 0.05).
accounted for a greater proportion of the variance in the 3-tank RTs than either version of the Old/New Equation. In contrast, the original version of the Old/New Equation was the more accurate at predicting 3-tank RTs, as indicated by the rmsd values. None of these functions provided a better fit to the data than the functions that were fitted directly to the data.

The analysis of 3-tank RTs did not provide clear support for the prediction that improvement in this variable would reflect the development of new productions and so would be at the same rate as observed during Training. Expecting exactly the same rate was certainly optimistic and so a fairer test of this prediction was to examine a function that was designed to describe improvement of new skills in terms of how well it could account for the observed improvement in 3-tank RTs. This function was found to provide a good account of the data, but it was not substantially better than that provided by the Old/New Equation, which was designed to describe improvement of old and new skills. Therefore, neither analysis supported the assumption that all new skills will be learned at the same rate.

Unfortunately, the above result can not be considered conclusive evidence against the assumption of a constant learning rate for new skills. It could be argued that it was not appropriate to use the learning rate of Flow-Route RTs during Training as an estimate of the constant learning rate at which all new skills will be learned. As mentioned in Chapter 4, most tasks will contain some elements of old tasks. As a result it will be difficult to obtain a true measure of the rate at which absolutely new skills are learned. If this is the case then a task more like the processing of the volume levels of three tanks should have been used to estimate 'new' learning rate. However, the more similar the task is to the Transfer task, the more likely it is that there will be transfer between the tasks. This would then mean that the Transfer task
would involve old and new tasks which would then defeat the purpose of the exercise. Further discussion of this issue will be deferred to the final chapter.

6.4.2.3 1-tank RTs

The predicted outcome for 1-tank RTs during Transfer was that performance would reflect the combination of old and new productions and so be described by a version of the Old/New Equation. Table 6.4 presents parameters of various functions fitted to 1-tank RTs. From those functions fitted directly to the data it can be seen that the rate of improvement in 1-tank RTs during the Transfer phase (10 blocks: -0.547, 6 blocks: -0.581) was slower than observed in Flow-Route RTs during Training (-0.756).

There did not appear to be any effect of a false asymptote. In fact, learning rate increased slightly when it was estimated from the first six blocks of Transfer compared to the estimate from all ten blocks.

Three power functions were developed to test the prediction that the attenuation in learning rate described above is a result of the combination of old and new skills. A version of the Old/New Equation was derived from the Training Flow-Route RTs and the first block of 1-tank RTs to test this prediction directly. The equation of this function is \( T = 295 + 2025.5 P_o^{-0.756} + 1777.64 P_n^{-0.756} \). The new version of the Old/New Equation which incorporates a change in asymptote could not be applied to this data set because the asymptote of the 1-tank RTs was less than the asymptote observed in Flow-Route RTs during Training.

The second function examined with 1-tank RTs was designed to provide a benchmark for examining the ability of the Old/New Equation to account for
### Table 6.4: Parameters of power functions fitted to 1-tank RTs during Transfer phase of Experiment 7.

Functions were fitted to all ten blocks and to the first six blocks of the Transfer data. Functions labelled "Observed" were fitted directly to the data. Functions labelled "New" were designed to describe improvement of new skills and include the learning rate and asymptote observed in Flow-Route RTs during Training (see text for details). Functions labelled "Old/New Equation" were fitted to values predicted by a version of the Old/New Equation (see text for details). Functions labelled "Old Equation" were fitted to values predicted by a version of the Old Equation (see text for details). $r^2 =$ proportion of variance in the observed reaction time values accounted for by the predicted values. $\text{rmsd} =$ root mean squared deviation between predicted and observed reaction time values.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Goodness of Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$X$</td>
</tr>
<tr>
<td><strong>Transfer (Blocks 1-10)</strong></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>35</td>
</tr>
<tr>
<td>New</td>
<td>295</td>
</tr>
<tr>
<td>Old/New Equation</td>
<td>295</td>
</tr>
<tr>
<td>Old Equation</td>
<td>15</td>
</tr>
<tr>
<td><strong>Transfer (Blocks 1-6)</strong></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>0</td>
</tr>
<tr>
<td>New</td>
<td>295</td>
</tr>
<tr>
<td>Old/New Equation</td>
<td>295</td>
</tr>
<tr>
<td>Old Equation</td>
<td>15</td>
</tr>
</tbody>
</table>
the data. This function had the same learning rate as that observed in Flow-Route RTs during Training and therefore should provide a worse fit to the data than the Old/New Equation if the observed learning rate attenuation is a result of the processes that the Old/New Equation is designed to describe. This second function was derived in the same fashion as the one derived for 3-tank RTs. The learning rate and asymptote were the values observed with Flow-Route RTs during Training (i.e., $c = -0.756$, $X = 295$). The intercept was calculated by subtracting the asymptote from the mean 1-tank RT of the first block of Transfer:

$$T = X_n + N_n P_n c$$

$$=> 3020.06 = 295 + N_n (0.739)^{-0.75588}$$

$$=> 2168.14 = N_n$$

therefore $T = 295 + 2168.14 P_n^{-0.756}$

The third function examined with 1-tank RTs was an application of the Old Equation and was derived by extrapolating the Training Flow-Route RTs into the Transfer phase.

Curves generated from the three functions are displayed in Figure 6.11, along with 1-tank RTs from the Transfer phase and a curve that provided the best fit to the Training Flow-Route RTs which is extrapolated into the Transfer phase. It is clear from this figure that extrapolating Training performance (i.e., Old Equation) does not provide a good account of Transfer performance in this particular situation. This suggests that new productions other than those developed during Training were developed and improved during this phase. However, it is not clear from the figure whether 1-tank RTs were reflecting the improvement of new and old productions or just new
Figure 6.11: Mean 1-tank RTs during Transfer phase of Experiment 7. Curve labelled "Old/New Eqn" represents a version of the Old/New Equation (see text for equation). Curve labelled "New" represents a function designed to describe improvement of new skills (see Table 6.4 and text for equation details). Curve labelled "Old Eqn" represents the power function that provided the best fit to Flow-Route RTs during Training. This curve has been extrapolated into the Transfer phase. Error bars are confidence limits (alpha = 0.05).
productions. That is, it is not obvious from the figure whether the function derived from the Old/New Equation or the function with the same learning rate as observed during Training provides the best account of the 1-tank RT data.

The results in Table 6.4 cast more light on the issue. In relation to both the first six blocks of Transfer and all ten blocks, the Old/New Equation provided a superior fit to the data than the other function with the same learning rate as observed during Training. This superior fit was in terms of accounting for a greater proportion of variance in the 1-tank RT data, and, on average, predicting reaction times which were closer to the observed values. Furthermore, the Old/New Equation accounted for virtually the same proportion of variance in the data as the curves fit directly to the data, and also predicted a similar learning rate to those observed in these curves (i.e., -0.601 vs. -0.547 & -0.581).

The fact that the rate at which 1-tank RTs improved was less than that observed during Training and also was similar to the rate predicted by the Old/New Equation supports the prediction that this performance measure would reflect the combination of old and new skills. The version of the Old/New Equation examined with 1-tank RTs was derived in part from Flow-Route RTs during Training and was shown to accurately predict the observed improvement in 1-tank RTs. This result suggests that some of the productions whose operation was reflected by Flow-Route RTs during Training were incorporated into the processing of tank volumes that was reflected by 1-tank RTs. The most likely productions are those concerned with deriving expected true conclusions (i.e., those involved in goals 5, 2 and 6 in Figure 6.2). The introduction to this chapter contained a description of two other sets of productions that would be developed to cope with the section of the complex
version of the tank task reflected by 1-tank RTs (i.e., those involved in goals 9 and 5 in Figure 6.5). It would appear then that the Old/New Equation provides a good description of the improvement involved in the combination of these old and new skills.

6.4.2.4 Conclusion RTs

Conclusion RTs during the Transfer phase were expected to reflect the combination of old and new productions and so should be described best by a version of the Old/New Equation. Table 6.5 presents parameters of various power functions fitted to Conclusion RTs during Training and Transfer. Measures of learning rate during Transfer did not differ substantially as a result of estimating on the basis of six or ten blocks (i.e., -0.283 vs. -0.297). Therefore, estimates of learning rate did not appear to be affected by a false asymptote.

The rate at which Conclusion RTs improved during Transfer (-0.297 & -0.283) was substantially slower than was observed during Training (-0.705), but faster than would be predicted if Transfer performance was simply the continuation of Training performance (i.e., learning rate predicted by the Old Equation = -0.037).

The version of the Old/New Equation that was examined with Conclusion RTs was derived from Training (power function in Table 6.5) and Transfer (mean Conclusion RT of first block) performance. This equation is $T = 655 + 552.86 \ P_o^{-0.705} + 668.14 \ P_n^{-0.705}$. The learning rate that is predicted on the basis of this function (-0.584) is faster than was observed. However, Figure 6.12 shows that this equation predicts a pattern of improvement that appears very similar to what was observed. Certainly this version of the Old/New
<table>
<thead>
<tr>
<th>Parameters</th>
<th>X</th>
<th>N</th>
<th>c</th>
<th>( r^2 )</th>
<th>rmsd</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>655</td>
<td>552.86</td>
<td>-0.705</td>
<td>0.987</td>
<td>17.593</td>
</tr>
<tr>
<td>Transfer (Blocks 1-10)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>125</td>
<td>1352.10</td>
<td>-0.297</td>
<td>0.977</td>
<td>35.753</td>
</tr>
<tr>
<td>Old Equation</td>
<td>40</td>
<td>760.38</td>
<td>-0.037</td>
<td>0.898</td>
<td>367.938</td>
</tr>
<tr>
<td>Old/New Equation</td>
<td>655</td>
<td>813.18</td>
<td>-0.584</td>
<td>0.972</td>
<td>42.402</td>
</tr>
<tr>
<td>Transfer (Blocks 1-6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>15</td>
<td>1474.60</td>
<td>-0.283</td>
<td>0.972</td>
<td>38.664</td>
</tr>
<tr>
<td>Old Equation</td>
<td>40</td>
<td>760.38</td>
<td>-0.037</td>
<td>0.928</td>
<td>462.364</td>
</tr>
<tr>
<td>Old/New Equation</td>
<td>655</td>
<td>813.18</td>
<td>-0.584</td>
<td>0.964</td>
<td>43.947</td>
</tr>
</tbody>
</table>

**Table 6.5:** Parameters of power functions fitted to Conclusion RTs during Training and Transfer phases of Experiment 7. Functions were fitted to all ten blocks and to the first six blocks of the Transfer data. Functions labelled "Observed" were fitted directly to the data. Functions labelled "Old Equation" were fitted to values extrapolated from Training performance. Functions labelled "Old/New Equation" were fitted to values predicted by a version of the Old/New Equation (see text for details). \( r^2 = \) proportion of variance in the observed reaction time values accounted for by the predicted values. \( \text{rmsd} = \) root mean squared deviation between predicted and observed reaction time values.
Figure 6.12: Mean Conclusion RTs during Training and Transfer phases of Experiment 7. The curve drawn through the Training data is the best-fit power function for this data (see Table 6.5 for equation) and has been extrapolated into the Transfer phase (labelled "Old Eqn"). The curve labelled "Old/New Eqn" represents a version of the Old/New Equation (see text for equation). The curve labelled "best-fit (Tf)" represents the power function that provides the best fit to the Transfer data (see Table 6.5 for equation). Error bars are confidence limits (alpha = 0.05).
Equation accounted for a larger proportion of the variance in the data than was accounted for by the Old Equation, and predicted reaction times that were, on average, closer to the observed values than were those predicted from Training (see Table 6.5). In fact this version of the Old/New Equation provided almost as good a fit to the data as did the power function fitted directly to the data.

The Conclusion RT results are clear. The Old/New Equation was able to provide a good account of the pattern of improvement in the Transfer data. This result constitutes further support for the proposal that the Old/New Equation describes the improvement that results from the combination of old and new skills. However, the Old/New Equation did not provide a perfect account of Transfer performance. As was found in Experiments 4 and 5, the Old/New Equation predicted a learning rate that was faster than was observed. This result is puzzling considering the apparently good-fit that the Old/New Equation has with the data (see especially Figure 6.12). Again it appears that Transfer performance is affected by more factors than are included in the Old/New Equation. Unfortunately the results of the current experiment do not provide any clues as to the nature of these factors. It was not possible to apply the new version of the Old/New Equation which incorporates a change in asymptote because the asymptote for Conclusion RTs was reduced with the transition from Training to Transfer. Apparently some forms of partial transfer involve an increase in asymptote whereas others do not. Discussion of this issue will be deferred to the last chapter.

6.4.2.5 Total RTs

The predicted outcome for Total RTs during the Transfer phase was that performance would reflect the execution of both old and new productions and
so would be best described by a version of the Old/New Equation. Parameters of various power functions fitted to Total RTs during Training and Transfer are presented in Table 6.6. The fact that subjects appeared to reach an asymptote prematurely (see section 6.3.1.2 and Figure 6.13) had an obvious effect on learning rate estimates. When learning rate was measured from all ten blocks of Transfer it was faster than was observed during Training (-0.898 vs. -0.793). As mentioned earlier, this result is prohibited by the Old/New Equation if any transfer has occurred, as was expected to be the case between the two versions of the tank task. However, the effect of a false asymptote is apparent when learning rate is estimated from the first six blocks only. This estimate (-0.631) is slower than the ten-block estimate. Therefore something appears to have retarded the improvement of subjects' performance (as measured by Total RTs) in the final blocks of Transfer and this resulted in an artificially inflated estimate of learning rate.

The previously described results supported the prediction that performance of the complex version of the tank task would involve both productions developed in the context of the simple version of the task and new productions developed to cope with the additional complexity of the new task. Thus Total RTs during Transfer should reflect this combination of old and new skills, and, as predicted in the introduction to this chapter, the Old/New Equation should describe the improvement associated with this combination. Consistent with this prediction is the finding that learning rate during Transfer (based on the first six blocks) was slower than during Training (-0.631 vs. -0.793). Further examination of this prediction was conducted by deriving an applicable version of the Old/New Equation: \( T = 1030 + 2511.1 P_0^{-0.793} + 6318.12 P_n^{-0.793} \). In addition, two examples of the new version of the Old/New Equation which incorporates a change in asymptote were derived for this data set. The function for blocks 1-10 is \( T = 1030 + 2511.1 P_0^{-0.793} \).
Table 6.6: Parameters of power functions fitted to Total RTs during Training and Transfer phases of Experiment 7. Functions were fitted to all ten blocks and to the first six blocks of the Transfer data. Functions labelled "Observed" were fitted directly to the data. Functions labelled "Old Equation" were fitted to values extrapolated from Training performance. Functions labelled "Old/New Equation" were fitted to values predicted by a version of the Old/New Equation (see text for details). Functions labelled "Old/New Equation*" were fitted to values predicted by the new version of the Old/New Equation which incorporates a change in asymptote (see text for details). $r^2$ = proportion of variance in the observed reaction time values accounted for by the predicted values. rmsd = root mean squared deviation between predicted and observed reaction time values.
Figure 6.13: Mean Total RTs during Training and Transfer phases of Experiment 7. The curve drawn through the Training data is the best-fit power function for this data (see Table 6.6 for equation) and has been extrapolated into the Transfer phase (labelled "Old Eqn"). The curve labelled "Old/New Eqn" represents a version of the Old/New Equation (see text for equation). The curve labelled "Old/New Eqn*" represents the new version of the Old/New Equation which incorporates a change in asymptote (see text for details). Error bars are confidence limits (alpha = 0.05).
+ 1595 + 5063.2 \, P_{n}^{-0.793}. \) The function for blocks 1-6 is \( T = 1030 + 2511.1 \, P_{o}^{-0.793} + 250 + 6121.43 \, P_{n}^{-0.793}. \) The ability of these functions to account for the Transfer data was compared with the Old Equation.

The Old Equation did not provide a good account of Total RTs during Transfer. The learning rate that was predicted in this case (-0.080) was considerably slower than was observed (-0.631). Furthermore, as illustrated in Figure 6.13, the predicted reaction times were nowhere near the observed values. Therefore, it is obvious that performance in this phase, as measured by Total RTs, was not simply continued improvement of old skills.

The original and the new versions of the Old/New Equation accounted for similar proportions of the variance in the Transfer Total RTs. In fact these functions accounted for a similar amount of variance to the power functions that were fitted directly to the data. However, as illustrated in Figure 6.13, and indicated by the rmsd values in Table 6.6, the new version of the Old/New Equation predicted reaction times that were, on average, closer to the observed values than those predicted by the original version of the Old/New Equation. Therefore, for Total RTs, the new version of the Old/New Equation which incorporates a change in asymptote provided the better account of the combination of old and new skills.

6.4.3 Summary and Conclusions

In the introduction to this chapter, the simple and complex versions of the tank task were described along with likely strategies that would be adopted with these tasks. This description suggested that performance during Transfer would involve continued improvement of old skills and the development of new skills. The task was designed to provide separate measures of those parts
of the task that involved old skills only, new skills only, and the combination of old and new skills.

Where Transfer performance was assumed to reflect continued improvement of old skills only (i.e., Flow-Route RTs) improvement was expected to follow on from the learning function observed during Training. Except for what appeared to be a temporary slowing of performance during the first Transfer block, this prediction was supported by the results. Possible causes for this temporary slowing include a surprise reaction associated with the additional complexity of the task or the reorganisation of the goal structure of the processing strategy which postponed some components of the task to later in each trial. In addition, there was some evidence that Flow-Route RTs during Transfer reflected the execution of fewer productions than were applied during Training, which was assumed to result from this postponement of some processing to later in the task. This result was predicted on the basis of the strategies described in Figures 6.2 and 6.5.

Where Transfer performance was assumed to reflect the combination of old and new skills (1-tank RTs, Conclusion RTs and Total RTs) improvement was expected to be described by appropriate versions of the Old/New Equation. In all cases the results supported this prediction. It is interesting to note that only in Total RTs did the new version of the Old/New Equation, which incorporates a change in asymptote, provide a better account of Transfer performance than the original version of this Equation. As suggested earlier, this implies that partial transfer does not affect asymptotic performance level in all situations. However, it is also possible that Total RTs are a more accurate measure of performance in this task than either 1-tank RTs or Conclusion RTs. As with the syllogism task, subjects may carry-over processing of some parts of the task into other parts. As a result, there is no
guarantee that 1-tank RTs and Conclusion RTs always reflect the processing that is suggested by the strategy described in Figure 6.5. However, Total RTs will reflect all processing in this task. Therefore greater credence should be given to the support that Total RTs provide for the proposal that the combination of old and new skills affects asymptotic performance. This issue of the effect of the combination of old and new skills on asymptotic performance will be examined in the following two chapters.

Where performance during the Transfer phase was assumed to reflect the development and operation of new skills only (3-tank RTs), improvement was expected to be at the same rate observed during Training. This prediction was not supported by the results. This apparently contradicts the assumption that new skills are learned at some constant rate. However, an alternate explanation of the result is that the learning rate observed during Training may not have been appropriate as an estimate of the constant rate at which all new skills are learned. The fact that the learning rate estimated from the 3-tank RTs was so much faster than that observed during Training (i.e., -1.3 vs. -0.7) suggests that in the tank task it is possible to learn at a faster rate than was initially observed during Training. This issue of the appropriate measures of "new" learning rate will be examined in further depth in the final chapter.

The results illustrated how estimates of learning rate can be affected by factors other than skill acquisition. Learning rate was shown to be inflated when performance was slowed in the early blocks of Transfer for reasons that were suggested to be transient. For instance, the change in task is likely to have resulted in a change in strategy so that the order in which some goals were accomplished was altered, and this temporarily slowed performance. Learning rate estimates were also shown to be inflated by false asymptotes. These may have been caused by factors such as fatigue or changes in
motivation. The effect of these factors on the estimation of learning rate will be considered further in the final chapter in relation to the most appropriate measures of "new" learning rate.
## Chapter 7  Experiment 8

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1 Introduction</td>
<td>305</td>
</tr>
<tr>
<td>7.2 Method</td>
<td>313</td>
</tr>
<tr>
<td>7.2.1 Subjects</td>
<td>313</td>
</tr>
<tr>
<td>7.2.2 Materials</td>
<td>313</td>
</tr>
<tr>
<td>7.2.3 Apparatus and Procedure</td>
<td>313</td>
</tr>
<tr>
<td>7.3 Results and Discussion</td>
<td>314</td>
</tr>
<tr>
<td>7.3.1 Analyses of Variance</td>
<td>314</td>
</tr>
<tr>
<td>7.3.1.1 Training</td>
<td>314</td>
</tr>
<tr>
<td>7.3.1.2 Transfer 1</td>
<td>316</td>
</tr>
<tr>
<td>7.3.1.3 Transfer 2</td>
<td>317</td>
</tr>
<tr>
<td>7.3.2 Learning Functions</td>
<td>317</td>
</tr>
<tr>
<td>7.3.2.1 Premise RTs</td>
<td>318</td>
</tr>
<tr>
<td>7.3.2.2 Conclusion RTs</td>
<td>325</td>
</tr>
<tr>
<td>7.3.2.3 Total RTs</td>
<td>330</td>
</tr>
<tr>
<td>7.3.3 Summary and Conclusions</td>
<td>335</td>
</tr>
</tbody>
</table>
7.1 Introduction

As reported in Chapter 6, Experiment 7 did not provide any evidence in support of the assumption that all new skills are learned at the same rate. Some of the results of that experiment suggested that it may not have provided a suitable test of this assumption. Experiment 8 was designed to provide an alternate view on this issue. Although this experiment was not designed to test the assumption specifically, it was designed to again examine the learning functions of processes underlying performance of a task.

The derivation of the Old/New Equation as a description of the combination of old and new skills was based on a number of assumptions. One of these assumptions was the one mentioned above, that when new skills are combined with old skills they will be learned at the same rate at which the old skills were learned. A second assumption was that following this combination of the two types of skills, the old skills will continue to improve in accordance with the learning function that describes the original improvement of these skills. In other words, if a learning function for the old skills was established prior to the combination with new skills, then the continued improvement of the old skills in the context of the new task could be predicted by extrapolating the original learning function.

The major implication of the above assumption concerning the continued improvement of old skills is that these skills, once established, function independently of other skills. That is, despite old and new skills being combined to perform a task, these sets of skills will perform largely autonomously of each other. Thus, underlying the assumption that, following such a combination, old skills will continue to improve according to their previously observed learning function, is the claim that this combination will
have no effect on the performance of these skills. The productions underlying the old skills will be executed in the same way before and after the combination with new skills. Although this is not an explicit claim in the ACT* theory of skill acquisition, to a certain extent it is implicit in the ACT* account of transfer. This account states that if a production can apply, it will (Singley & Anderson, 1989). The appropriate circumstances for a production to apply are those embodied in the condition part of the productions, and this usually involves both stimulus conditions and goals. So as far as ACT* is concerned, if these conditions are similar before and after the combination of old and new skills, productions that applied before the combination will also apply after. However, the ACT* theory is not explicit as to whether the execution of these productions following transfer can be described by the same function that described performance prior to transfer.

This assumption concerning the continued improvement of old skills received indirect support from a number of the previous experiments described in this thesis. For instance, when the Experimental group of Experiment 4 were given syllogisms to solve without capitalised common elements, Conclusion RTs during Transfer were shown to improve in accordance with the learning function observed with Conclusion RTs during Training (see Figure 3.3). This result implies that in the context of the new task the old skills involved with processing conclusions continued to improve as if nothing had changed from the previous phase. Similarly, Flow-Route RTs during the Transfer phase of Experiment 7 were predicted by extrapolating the learning function of the Training Flow-Route RTs (see Figure 6.8). Thus in the context of the more complex task performed during Transfer, the old skills involved in processing flow-route information improved as if the task had not altered (although see Figure 6.7 and discussion in section 6.4.2.1 concerning temporary effects on learning rate).
The reason that the above results were said to provide indirect support for the second assumption is that, in both cases, the task component which depended on old skills was procedurally isolated from the new task components. This was because performance was measured at a stage in each task when the impact of new skills may have passed or was yet to occur. For instance, Conclusion RTs in the syllogism task were assumed to reflect processing of conclusions only and this was not expected to involve any new skills in Experiment 4. The impact of new skills was expected to be reflected in the processing of premises. Thus on each trial, the operation of new skills would have been completed by the time conclusions were presented. Similarly Flow-Route RTs in the tank task were assumed to reflect the processing of flow-route information which was not expected to involve new skills in Experiment 7. It was also expected that on each trial this processing would have been completed by the time volume level information was presented. This second source of information required the operation of new skills. Thus the operation of the old skills reflected by Flow-Route RTs would have ceased prior to the operation of the new skills.

A further test of the assumption concerning the continued improvement of old skills was made in Experiment 6. The transition from Alternating Training to Random Transfer was assumed to involve execution of the same productions in both phases despite a change in stimulus conditions. The results of this experiment suggested that the skills developed during Training improved during Transfer according to the learning function observed during Training. This therefore provides more direct support for the assumption by simulating one feature of the combination of old and new skills, that is, a change in stimulus conditions. However, an obvious feature of this combination that was not a feature of Experiment 6 was the development of new skills. As a
result the assumption that the performance of old skills will not be affected by the combination with new skills was not directly examined in that experiment.

The assumption that old skills are unaffected by a combination with new skills is relevant not only to the Old/New Equation, but also to the results which suggest that this combination leads to an increase in performance asymptote. The new version of the Old/New Equation which incorporated this change in asymptote was only one way of accounting for the change in learning function. An alternate explanation for this change is that old skills are affected by the combination with new skills. If this is the case, it would be incorrect to assume that the learning function which described improvement of old skills during Training would also describe their continued improvement during Transfer. Therefore the validity of the Old/New Equation, both the original and new versions, rests upon the assumption that skills are performed independently of each other.

To conduct a more direct test of the independence assumption underlying the Old/New Equation old skills need to be examined in the context of a task that also involves new skills. In addition the operation of the two sets of skills should not be artificially separated. This of course makes the measurement of the two sets of skills very difficult, as it was the separation of task stages in the previous experiments that enabled the measurement of the different types of skills. Thus the design of the previous experiments would appear to be inappropriate for a test of the independence assumption underlying the Old/New Equation.

A more appropriate design is suggested by a study that has already been described in Chapter 4. The study reported by Snyder and Pronko (1952) involved performing a visual-motor task with mirror-reversed vision and was
discussed in terms of a combination of old and new skills. Subjects were given a number of days practice at the task with normal vision. Then they were to practise the task with mirror-reversed vision. In the earlier discussion of this study it was not mentioned that after the mirror-reversed phase, subjects practised the task for four more days, this time with normal vision again.

In Chapter 4, performing the Snyder and Pronko task with mirror-reversed vision was described as involving a combination of old and new skills. In contrast, performing the task with normal vision was assumed to involve old skills only. Thus the three phases of the complete experiment can be considered to involve practice of old skills, practice of old and new skills, and finally further practice of old skills. This design should provide a test of the independence assumption underlying the Old/New Equation. The assumption predicts that the old skills involved in performing the task with normal vision will continue to improve throughout the last two phases of the experiment according to the learning function exhibited in the first phase. In other words, the old skills will continue to improve at the same constant rate, unaffected by the introduction of new skills. If this is the case, then performance on the task in the third phase of the experiment, where normal vision is restored, should fall along a curve extrapolated from performance in the first phase.

Figure 7.1 displays the full set of results from the Snyder and Pronko (1952) study. The two curves are the best-fit power functions for performance with normal vision and mirror-reversed vision. Both of these curves have been extrapolated in order to examine which curve provides the best account of performance in the third phase where normal vision is restored. Performance on the first day of this third phase does not appear any different from the final days of the second phase. However, performance on the three subsequent
Figure 7.1: Data reported in Snyder, F.W. & Pronko, H.H. (1952). Vision with spatial inversion. Wichita, Kansas: University of Wichita Press. Data points have been plotted on linear axes. The curves represent the best-fit power functions with the following equations:

**Normal Vision:** \( T = 72.418 \, P^{-0.100} \), \( r^2 = 0.913 \)

**Reverse Vision:** \( T = 25 + 89.381 \, P^{-0.300} \), \( r^2 = 0.938 \)
days of this third phase appears to approximate more closely the performance predicted on the basis of the original normal vision phase rather than that predicted on the basis of the mirror-reversed vision phase. Thus the results of the Snyder and Pronko study appear to support the independence assumption underlying the Old/New Equation. More specifically the results suggest that the combination of old and new skills is one where overall learning rate may be affected (i.e., the attenuation of learning rate predicted by the Old/New Equation), but the learning rate of old skills is unaffected by the change in task. The old skills continue to develop as if nothing has changed from the previous task. Furthermore, when the old skills are again observed in isolation, they emerge virtually unscathed from the combination with new skills, still improving at the same rate observed originally.

The design of the Snyder and Pronko study was incorporated in Experiment 8 to provide a further test of the independence assumption underlying the Old/New Equation, this time, with the syllogism task. The experiment involved three phases. The first phase (Training) was identical to the Training phase in Experiment 4 where subjects practised solving syllogisms with capitalised common elements. The second phase (Transfer 1) was identical to the Transfer phase of Experiment 4 where the common elements of the syllogisms were not highlighted. The third phase (Transfer 2) involved more practice with syllogisms with capitalised common elements.

If the independence assumption is correct, the skills developed in the Training phase of this experiment will continue to improve throughout the other two phases according to the learning function established during Training. In other words, despite the fact that new skills will be developed in the Transfer 1 phase to cope with the removal of the highlighting feature, the old skills will continue to improve as if the task had not changed. Indeed the parts of the
task that are processed by the old skills do not change. Therefore the learning function which describes improvement of the old skills should not change either.

If the learning function of the old skills remains the same throughout the experiment, Premise RTs during the Transfer 2 phase should be predicted by extrapolating the learning function observed with Premise RTs during Training. This is because performance during the third phase should reflect the operation of skills that were developed during the first phase and which continue to improve during the second phase. Premise RTs during Transfer 1 however, should be predicted by the Old/New Equation, because this second phase of the experiment will involve the combination of old and new skills. In contrast, Conclusion RTs throughout the two Transfer phases should be accounted for by extrapolating the learning function observed with Conclusion RTs during Training. The reason is that the same set of skills should be in operation in all three phases.

In order to provide a more comprehensive examination of the ability of the Old/New Equation to account for improvement following partial transfer, an additional function was compared with the data from this experiment. This new function is a variant of the Old Equation. Considering the changes in asymptote that have been observed following partial transfer, a change in asymptote was incorporated in the Old Equation. Therefore this new version, labelled "Old Equation + New X" (i.e., X = Asymptote), describes further improvement of old productions after the transition from Training to Transfer has caused a change in performance asymptote. Contrasting this function with the various versions of the Old/New Equation may be used to estimate the extent to which improvement during Transfer is associated with a combination of old and new skills, a change in asymptote, or both.
7.2 Method

7.2.1 Subjects

Nineteen volunteers from the University of Western Australia Psychology Department participated in this experiment for course credit or $5 per hour. Three subjects did not achieve the learning criterion of an error rate not exceeding 25% in the latter half of Training. Only the data from the remaining 16 subjects were analysed further.

7.2.2 Materials

Eight hundred and sixty-six syllogisms were used in this experiment. The two practice items and the first 576 items to be presented were the same items that were used in the Experimental condition of Experiment 4. Thus 288 items were presented during Training with common elements of each premise pair in upper case. During the first Transfer phase another 288 items were presented with all elements in lower case.

The remaining 288 items that were presented in this experiment were constructed by deriving three different combinations of the elements in the original 96 items described in Experiment 1. These new items were presented in the second Transfer phase, again with the common elements of each premise pair in upper case.

7.2.3 Apparatus and Procedure

These were identical to those described in Experiment 4 except in this experiment 866 trials were presented in three phases: the Training phase
(common elements capitalised), the first Transfer phase (all elements lower case), and the second Transfer phase (common elements capitalised). Subjects were tested individually in one session which lasted from 90 to 120 minutes. One minute rest periods were provided every 48 trials.

7.3 Results and Discussion

Analyses of variance were conducted with Premise and Conclusion RTs and Accuracy data in all three phases of the experiment. These analyses will be presented first, followed by an examination of the learning functions for the reaction time variables.

7.3.1 Analyses of Variance

7.3.1.1 Training

A number of analyses of variance were performed on the Training data. A 6 (Training block) x 2 (syllogism type) ANOVA was used to analyse Premise RTs. Two 6 (Training block) x 2 (syllogism type) x 2 (conclusion type) ANOVAs were used to analyse Conclusion RTs and Accuracy.

Mean Premise and Conclusion RTs during all phases of the experiment are displayed in Figure 7.2. It is clear from this figure that subjects showed significant improvement in these performance measures throughout the Training phase. Premise RTs were reduced from 5407 ms in Block 1 to 968 ms in Block 6 (F(5,75) = 40.00, p<0.05). Conclusion RTs improved from 1684 ms in Block 1 to 689 ms in Block 6 (F(5,75) = 18.27, p<0.05). Mean Accuracy also improved with practice, from 74.22% in Block 1 to 95.05% in Block 6 (F(5,75) = 9.61, p<0.05).
Figure 7.2: Mean Premise and Conclusion RTs during Training, Transfer 1 and Transfer 2 phases of Experiment 8.
Syllogism type again had an effect on Premise RTs. The premises of ABBC syllogisms were studied for less time than the premises of BCAB syllogisms (ABBC=2331 ms vs. BCAB=2556 ms, F(1,15) = 10.19, p<0.05). There was no such effect on Conclusion RTs or on Accuracy (for both measures F(1,15)<1). These results replicate those of Experiments 1, 2, 3, 4 and 6 and again suggest that subjects processed premises with a bias towards those with an ABBC structure.

As in the previous experiments, True conclusions were processed faster than False conclusions (True=950 ms vs. False=1054 ms, F(1,15) = 24.00, p<0.05). They were also responded to more accurately than False conclusions (True=91.84% vs. False=86.85%, F(1,15) = 10.00, p<0.05). This result replicates the bias towards True conclusions observed in the previous experiments.

7.3.1.2 Transfer 1

A number of analyses of variance were performed on the Transfer 1 data. A 6 (Transfer 1 block) x 2 (syllogism type) ANOVA was used to analyse Premise RTs. Two 6 (Transfer 1 block) x 2 (syllogism type) x 2 (conclusion type) ANOVAs were used to analyse Conclusion RTs and Accuracy.

Figure 7.2 shows that both Premise and Conclusions RTs were reduced with practice during the first Transfer phase. Premise RTs improved from 2264 ms in the first block of Transfer 1 to 1399 ms in the sixth block of this phase (F(5,75) = 10.10, p<0.05). Conclusion RTs were reduced from 778 ms in Block 1 to 660 ms in Block 6 of Transfer 1 (F(5,75) = 4.12, p<0.05). In contrast, Accuracy did not improve with practice during this phase (F(5,75) = 1.35, p>0.05).
The only other significant effects in this phase concerned True versus False conclusions. True conclusions were again processed faster than False conclusions (True=674 ms vs. False=742 ms, $F(1,15) = 11.97, p<0.05$). Conclusion RTs also showed a significant interaction between practice and True versus False conclusions ($F(5,75) = 2.49, p<0.05$), with the speed advantage of True conclusions being reduced with practice.

7.3.1.3 Transfer 2

A number of analyses of variance were performed on the Transfer 2 data. A 6 (Transfer 2 block) x 2 (syllogism type) ANOVA was used to analyse Premise RTs. Two 6 (Transfer 2 block) x 2 (syllogism type) x 2 (conclusion type) ANOVAs were used to analyse Conclusion RTs and Accuracy.

Figure 7.2 shows that the amount of improvement with practice exhibited in both Premise and Conclusion RTs was substantially less in the second Transfer phase than in the previous two phases. In fact there was no significant improvement in Conclusion RTs ($F(5,75)<1$) or Accuracy ($F(5,75)<1$). Only Premise RTs showed any significant improvement, with performance time being reduced from 878 ms in Block 1 of this phase to 745 ms in the final block ($F(5,75) = 2.70, p<0.05$). The only other significant effect in this phase was the ubiquitous speed advantage of True conclusions. Again these were processed faster than False conclusions (True= 587 ms vs. False=660 ms, $F(1,15) = 15.69, p<0.05$).

7.3.2 Learning Functions

The learning functions exhibited in Premise and Conclusion RTs will be discussed separately. Learning functions exhibited by Total RTs were also
examined, for reasons that will be discussed below, and will be presented at
the end of this section.

7.3.2.1 Premise RTs

Parameters of power functions fitted to Premise RTs in the three phases of the
experiment are presented in Table 7.1. The learning rate observed during the
Transfer 1 phase (-0.408) was slower than observed during Training (-0.877)
and faster than predicted if improvement in this phase simply continued the
improvement observed during Training (-0.273). This is the result that is
predicted by the Old/New Equation if performance during this second phase
relies on developing new productions and combining them with old
productions. Certainly there is evidence for the development of new
productions. Figure 7.3 shows that Premise RTs were slowed substantially
as a result of the switch from upper-case common elements to lower-case
common elements. This slowing is consistent with the development of new
productions.

A number of functions were compared with the Premise RTs during the
Transfer 1 phase in order to evaluate the various possible causes of the
slowing during this phase. The amount of slowing was used to derive
versions of the Old/New Equation. The equation for the original version of
this equation is \( T = 4475.7 P_o^{-0.877} + 1317.9 P_n^{-0.877} \). The equation for the
new version of the Old/New Equation which incorporates a change in
asymptote is \( T = 4475.7 P_o^{-0.877} + 600 + 742.71 P_n^{-0.877} \). In addition, two
versions of the Old Equation were compared with the data. The original
version was derived by extrapolating the function that provided the best fit to
the Training Premise RTs. The new version of the Old Equation was derived
by the same process as the original version but included a new asymptote that
Chapter 7

Table 7.1: Parameters of power functions fitted to Premise RTs during Training, Transfer 1 and Transfer 2 phases of Experiment 8. Functions labelled "Observed" were fitted directly to the data. Functions labelled "Old Equation" were fitted to values extrapolated from Training performance. Functions labelled "Old Equation + New X" were fitted to values predicted by a new version of the Old Equation which incorporates a change in asymptote (see text for details). Functions labelled "Old Equation + Tfl X" were fitted to values predicted by a version of the Old Equation which includes the asymptote observed in the Transfer 1 phase (see text for details). The function labelled "Old/New Equation" was fitted to values predicted by a version of the Old/New Equation (see text for details). The function labelled "Old/New Equation*" was fitted to values predicted by the new version of the Old/New Equation which incorporates a change in asymptote (see text for details). The function labelled "Old/New Equation-Extrap’ed" was fitted to values predicted for the Transfer 2 phase by the function that provided the best fit to the Transfer 1 data (i.e., "Observed" in Transfer 1). $r^2 = \text{proportion of variance in the observed reaction time values accounted for by the predicted values. rmsd} = \text{root mean squared deviation between predicted and observed reaction time values.}
Figure 7.3: Mean Premise RTs during Training, Transfer 1 and Transfer 2 phases of Experiment 8. The curve labelled "best-fit(Tn)" represents the best-fit power function for the Training results (see Table 7.1 for equation). This curve has been extrapolated into the other two phases (labelled "Old Eqn"). The curve labelled "Old/New Eqn" represents a version of the Old/New Equation (see text for equation). This curve has also been extrapolated into the Transfer 2 phase. The curve labelled "Old/New Eqn*" represents a new version of the Old/New Equation which incorporates a change in asymptote (see text for equation). This curve has also been extrapolated into the Transfer 2 phase. Curves labelled "Old Eqn + New X" were fitted to values predicted by a version of the Old Equation which incorporates a change in asymptote (see text for details). The curve labelled "Old Eqn + Tfl X" was fitted to values predicted by a version of the Old Equation which includes the asymptote observed in the Transfer 1 phase (see text for details). Error bars are confidence limits (alpha = 0.05).
was calculated from the difference between the mean Premise RT in the first block of Transfer 1 (2263.75 ms) and the value predicted by extrapolating Training performance (889.01 ms):

\[ T_{1st\ block\ Tfl} = X + 4475.7 P^{-0.87711} \]

=> 2263.75 = X + 4475.7 (6.314)^{-0.87711}

=> 1374.74 = X

therefore \[ T = 1374.74 + 4475.7 P^{-0.87711} \]

The values predicted by these functions are displayed in Figure 7.3 and the equations of the power functions that provide the best fit to these values are presented in Table 7.1.

Table 7.1 and Figure 7.3 show that neither version of the Old Equation provides a good account of the pattern of improvement observed during the Transfer 1 phase. These results suggest that performance during this phase involved more than improvement in old skills only. The new version of the Old Equation (i.e., Old Equation + New X), although predicting values closer to the observed values than the original version, did not vastly improve upon the account provided by this version. Therefore there was no evidence in the Premise RTs that removing the highlighting feature simply increased the performance asymptote.

In Figure 7.3 it is obvious that the function predicted by the original version of the Old/New Equation again underestimates the degree of attenuation of learning rate that is associated with partial transfer. The reduction in learning rate from -0.877 observed during Training to -0.408 during the Transfer 1 phase represents a larger attenuation than is predicted by this version of the
equation (-0.585). However, the new version of the Old/New Equation which incorporates a change in asymptote predicted a learning rate for the Transfer 1 phase that was much closer to the rate observed (-0.370). This new version also predicted values which were a lot closer to the observed values than those predicted by the original version (see Figure 7.3 and rmsd values in Table 7.1), a result similar to those observed in Experiments 4 and 5 (see §5.5). This result suggests that the transition from Training to Transfer 1 in this experiment is not only associated with a combination of old and new skills, but that this combination results in an increase in performance asymptote.

Figure 7.3 shows that in the second Transfer phase, when the common elements of premise pairs were highlighted again, Premise RTs were reduced in comparison to performance in the Transfer 1 phase. This result supports the assumption that new skills developed during the Transfer 1 phase to complement the skills developed during Training would not be executed during the Transfer 2 phase. From Table 7.1 it can be seen that learning rate was very slow during this phase compared to the previous two phases (-0.099 vs. -0.877 & -0.408). Furthermore the curve that was fitted directly to the data in this phase did not account for as much variance in the data as did the curves that were fitted to the data in the previous phases. The reason for this is that improvement in this last phase was not uniform. As is clear from Figure 7.3, each successive block of trials did not always bring a further reduction in Premise RTs.

Some of the functions that were compared to the data in the Transfer 1 phase were extrapolated into the Transfer 2 phase (see Figure 7.3). As is obvious in Figure 7.3 and Table 7.1, none of these functions (i.e., the Old Equation, both versions of the Old/New Equation, and the function that was fitted
directly to the data in Transfer 1) predicted values that were close to the observed values. The failure of the Old Equation to account for the Transfer 2 performance is of relevance to the independence assumption underlying the Old/New Equation. It was assumed that performance during this phase would rely on only the productions developed during Training. These are the old productions that were assumed to combine with additional productions during the Transfer 1 phase. The independence assumption states that this combination should have no effect on the execution of the old productions and so they will continue to improve during the Transfer 1 phase, and also the Transfer 2 phase, according to the learning function observed during Training. This function is the Old Equation, which did not provide a good account of the improvement observed during this final phase. Therefore it appears that the independence assumption may be wrong, and that the combination of old and new skills does affect the continued improvement of the old skills.

Before the independence assumption is rejected, an alternate explanation for the improvement in Premise RTs observed during the Transfer 2 phase will be examined. If the independence assumption is correct, the results of this experiment so far suggest that partial transfer results in an attenuation of learning rate, as predicted by the Old/New Equation, and an increase in performance asymptote, as described by the new version of the Old/New Equation. This suggests an explanation for why the Old Equation underestimated the Premise RTs during the Transfer 2 phase (see Figure 7.3). If some factor associated with the combination of old and new skills caused an increase in asymptote, the effect of this factor may have also influenced performance during the Transfer 2 phase. If this change in asymptote was unrelated to the continued improvement of old skills then performance during the Transfer 2 phase may be described by the equation that describes
improvement of the old skills (i.e., the Old Equation) if it incorporates a change in asymptote. An equation was derived from the Old Equation to explore this possibility. This equation included the same asymptote that was observed during the Transfer 1 phase (i.e., X = 600) because this value may represent the change associated with the combination of old and new skills. Thus this function has the following equation: \( T = 600 + 4475.7 \ p^{-0.877} \). Premise RTs predicted by this function are displayed in Figure 7.3. It is obvious from this figure that this function predicts values that are slower than the observed values. Therefore the results do not suggest that the larger asymptote observed in the Transfer 1 phase is also a feature of performance in the Transfer 2 phase. However, an alternate version of the Old Equation which incorporates a change in asymptote was derived in a similar fashion to the one derived for the Transfer 1 data. This version incorporated an asymptote which was calculated from the difference between the mean Premise RT in the first block of the Transfer 2 phase (878.47 ms) and that predicted by the Old Equation (506.58 ms):

\[
\begin{align*}
T_{1st \ block} \ &= \ X + 4475.7 \ p^{-0.87711} \\
\Rightarrow \ 878.47 \ &= \ X + 4475.7 \ (11.989)^{-0.87711} \\
\Rightarrow \ 371.89 \ &= \ X \\
\text{therefore} \quad T \ &= \ 371.89 + 4475.7 \ p^{-0.87711}
\end{align*}
\]

Values predicted by this function are also displayed in Figure 7.3. This function clearly provides the best account of the pattern of improvement during this phase. In fact it is virtually equivalent in its ability to describe the data as the function that was fitted directly to the data (see Table 7.1). Thus these results suggest an alternate interpretation of the Transfer 2 data to that which suggests that the independence assumption underlying the Old/New
Equation is erroneous. This alternate interpretation is that the old productions developed during Training are still improving during the Transfer 2 phase according to the learning function observed during Training, but now the asymptote of the function has been increased. Although this asymptote is not the same one that is observed during the Transfer 1 phase, it is possible that whatever caused the increase in asymptote from Training to Transfer 1 is also responsible for the increase in asymptote in Transfer 2. Considering the degree to which Premise RTs in both Transfer phases can be accounted for by equations which are based upon the independence assumption, it would appear that this assumption is still a reasonable one, although this conclusion rests upon being able to establish that the change in asymptote reflects a process which is independent of the continued improvement of old skills. How this might be achieved will be discussed in the final chapter.

7.3.2.2 Conclusion RTs

Parameters of power functions fitted to Conclusion RTs during the three phases of this experiment are presented in Table 7.2. Learning rate during the Transfer 1 phase (-0.093) was slower than observed during the Training phase (-0.602) and also slightly slower than that predicted by extrapolating Training performance (-0.112).

Conclusion RTs during the first block of the Transfer 1 phase were significantly slower than was predicted by extrapolating Training performance (see Figure 7.4). This suggests that, contrary to expectations, Conclusion RTs were affected by the removal of the highlighting feature of the syllogisms and so reflected to some extent the combination of old and new skills. As a result a version of the Old/New Equation was derived from Training and Transfer 1 results to describe improvement of Conclusion RTs during the
Table 7.2: Parameters of power functions fitted to Conclusion RTs during Training, Transfer 1 and Transfer 2 phases of Experiment 8. Functions labelled "Observed" were fitted directly to the data. Functions labelled "Old Equation" were fitted to values extrapolated from Training performance. Functions labelled "Old Equation + New X" were fitted to values predicted by a new version of the Old Equation which incorporates a change in asymptote (see text for details). The function labelled "Old Eqn + New X (Tf1 ext'ed)" was fitted to values predicted by the Old Equation + New X derived for the Transfer 1 phase and extrapolated into the Transfer 2 phase (see text for details). The function labelled "Old/New Equation" was fitted to values predicted by a version of the Old/New Equation (see text for details). The function labelled "Old/New Equation-Extrap'ed" was fitted to values predicted by the Old/New Equation for the Transfer 2 phase. The function labelled "Transfer 1-Extap'ed" was fitted to values predicted for the Transfer 2 phase by the function that provided the best fit to the Transfer 1 data (i.e., "Observed" in Transfer 1). $r^2 =$ proportion of variance in the observed reaction time values accounted for by the predicted values. rmsd = root mean squared deviation between predicted and observed reaction time values.
Transfer 1 phase. In addition, the fact that the effect of developing new skills appears to have been reflected by both Premise and Conclusion RTs suggested that an analysis of Total RTs would be appropriate. Such an analysis would enable the examination of the overall impact of new skills on the learning rate of old skills. This analysis will be presented in the next section.

The version of the Old/New Equation derived to describe Conclusion RTs during the Transfer 1 phase is $T = 255 + 1189.6 P^{-0.60153} + 126.7 P^{-0.60153}$. Values predicted by this function are displayed in Figure 7.4. The new version of the Old/New Equation which incorporates a change in asymptote could not be applied with this data because the asymptote observed during the Transfer 1 phase (75 ms) was less than that observed during Training (255 ms). However, the new version of the Old Equation was applied to this data set, in addition to the original version of the Old Equation. This new version incorporates an increase in asymptote which was calculated from the difference between the mean Conclusion RT in the first block of the Transfer 1 phase (778.06 ms) and the value predicted by extrapolating Training performance (392.30 ms):

$$T_{1st \ block} \text{ Tfl} \Rightarrow \begin{align*}
\Rightarrow & \quad 778.06 \quad = \quad X + 1189.6 (6.314)^{-0.602} \\
\Rightarrow & \quad 385.76 \quad = \quad X
\end{align*}$$

therefore $T = 385.76 + 1189.6 P^{-0.602}$

Values predicted by these two versions of the Old Equation are displayed in Figure 7.4. This figure shows that the new version of the Old Equation provides the best account of the data, and provides almost as good an account
Figure 7.4: Mean Conclusion RTs during Training, Transfer 1 and Transfer 2 phases of Experiment 8. The curve labelled "best-fit(Tn)" represents the best-fit power function for the Training results (see Table 7.2 for equation). This curve has been extrapolated into the other two phases (labelled "Old Eqn"). The curve labelled "Old/New Eqn" represents a version of the Old/New Equation (see text for equation). The curve labelled "Old/New Eqn*" represents a version of the Old/New Equation which incorporates a change in asymptote (see text for equation). The curve labelled "Old Eqn + New X (T11)" was fitted to values predicted for the Transfer 1 phase by a version of the Old Equation which incorporates a change in asymptote (see text for details). All of these curves have been extrapolated into the Transfer 2 phase. The curve labelled "Old Eqn + New X (Tf 2)" was fitted to values predicted for the Transfer 2 phase by the version of the Old Equation which incorporates a change in asymptote (see text for details). Error bars are confidence limits (alpha = 0.05).
as the power function fitted directly to the data (see Table 7.2). This result suggests that the slowing in the initial blocks of this phase was not a result of additional processing but instead was due to an increase in asymptote. The cause of this change in asymptote is not clear. However, it may be associated with the change in asymptote observed with Premise RTs, and this is assumed to be a result of the combination of old and new skills. Hence, even though the Conclusion RTs do not directly reflect a combination of old and new skills, it may be that the change in asymptote associated with the combination of old and new skills in the processing of premises reflects a general process which can affect a wide range of performance operations. Thus the change in asymptote observed with Conclusion RTs may reflect some general effect on cognitive functioning. This issue will be pursued further in the final chapter.

In the second Transfer phase there was no significant improvement in Conclusion RTs with practice. This is reflected by the poor fit obtained by fitting a power function directly to the data (see Table 7.2). In spite of this, the four functions analysed in the Transfer 1 phase were extrapolated into the Transfer 2 phase to examine how well they can account for the data. In addition, the new version of the Old Equation which incorporates a change in asymptote was applied to this data. The change in asymptote was calculated from the difference between the mean Conclusion RT observed in the first block of the Transfer 2 phase (643.17 ms) and the value predicted by extrapolating Training performance (266.67 ms):

\[
T_{1st \ block \ Tr} = X + 1189.6 \ p^{-0.602}
\]

\[
\Rightarrow \quad 643.17 = X + 1189.6 \ (11.989)^{-0.602}
\]

\[
\Rightarrow \quad 376.50 = X
\]
therefore \[ T = 376.50 + 1189.6 P^{-0.602} \]

Values predicted by all of these functions have been plotted in Figure 7.4. The best accounts of the data were provided by the new version of the Old Equation derived from this data and by the new version of the Old Equation which was derived from the Transfer 1 data and extrapolated into the Transfer 2 phase. The learning functions predicted by these two equations are virtually indistinguishable (see Figure 7.4 and Table 7.2).

In summary, the main result that was observed with the Conclusion RTs is that when the highlighting feature was removed from the premises in the Transfer 1 phase, Conclusion RTs were slowed. The main conclusion that was reached with respect to this slowing is that rather than reflecting a combination of old and new skills, as was the case with Premise RTs, it appears to indicate an increase in asymptote. This change in asymptote was evident not only throughout the Transfer 1 phase, but persisted through the Transfer 2 phase, when the highlighting feature was reinstated.

7.3.2.3 Total RTs

The combination of Premise RTs and Conclusion RTs to form Total RTs reveals a pattern of results similar to those observed with Premise RTs. Learning rate during the Transfer 1 phase (-0.382) was slower than during the Training phase (-0.748) and faster than predicted by the Old Equation (-0.231). This is the type of result that is predicted by the Old/New Equation when performance reflects the combination of old and new skills.

Two versions of the Old/New Equation were derived for the Transfer 1 data. The equation of the original version is \[ T = 30 + 5863.2 P_o^{-0.74797} + 1480.04 \]
The equation for the new version of the Old/New Equation which incorporates a change in asymptote is $T = 30 + 5863.2 P_{0.74797} + 1050 + 467.17 P_{0.74797}$. In addition, two versions of the Old Equation were derived for this data set. The original version was derived by extrapolating Training performance. The new version incorporated a change in asymptote, and this was calculated from the difference between the mean Total RT during the first block of the Transfer 1 phase (3041.81 ms) and the value predicted by extrapolating Training performance (1477.51 ms):

$$T_{1st\ block\ TR1} = X + 5863.2 P^{-0.74797}$$

$$=> 3041.81 = X + 5863.2 (6.314)^{-0.74797}$$

$$=> 1564.30 = X$$

therefore $T = 1564.30 + 5863.2 P^{-0.74797}$

Values predicted by these functions are displayed in Figure 7.5. As is obvious in this figure and from Table 7.3, the new version of the Old/New Equation which incorporates a change in asymptote provides the best account of the Total RTs during the Transfer 1 phase. This replicates the Premise RTs result and supports the observation that partial transfer results in an attenuation of learning rate and an increase in performance asymptote.

In the Transfer 2 phase, reaction times were again significantly faster than during the Transfer 1 phase and significantly slower than predicted by extrapolating Training performance (see Figure 7.5). There was significant improvement during this phase with practice (Total RT in Block 1 = 1521.73 ms vs. Total RT in Block 6 = 1367.64 ms, $F(5,75) = 2.37, p<0.05$) although Figure 7.5 illustrates how small and non-uniform in direction this improvement was. Furthermore, the power function that provided the best fit
Figure 7.5: Mean Total RTs during Training, Transfer 1 and Transfer 2 phases of Experiment 8. The curve labelled "best-fit(Tn)" represents the best-fit power function for the Training results (see Table 7.3 for equation). This curve has been extrapolated into the other two phases (labelled "Old Eqn"). The curve labelled "Old/New Eqn" represents a version of the Old/New Equation (see text for equation). The curve labelled "Old/New Eqn*" represents a new version of the Old/New Equation which incorporates a change in asymptote (see text for equation). Both of the Old/New Equation curves have been extrapolated into the Transfer 2 phase. Curves labelled "Old Eqn + New X" were fitted to values predicted by a version of the Old Equation which incorporates a change in asymptote (see text for details). The curve labelled "Old Eqn + Tf1 X" was fitted to values predicted by a version of the Old Equation which includes the asymptote observed in the Transfer 1 phase (see text for details). Error bars are confidence limits (alpha = 0.05).
### Chapter 7

#### Parameters Goodness of Fit

<table>
<thead>
<tr>
<th>Parameters</th>
<th>X</th>
<th>N</th>
<th>C</th>
<th>r²</th>
<th>rmsd</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>30</td>
<td>5863.20</td>
<td>-0.748</td>
<td>0.989</td>
<td>228.335</td>
</tr>
<tr>
<td>Transfer 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>1080</td>
<td>1917.50</td>
<td>-0.382</td>
<td>0.998</td>
<td>14.392</td>
</tr>
<tr>
<td>Old Equation</td>
<td>20</td>
<td>1514.80</td>
<td>-0.231</td>
<td>0.949</td>
<td>1191.490</td>
</tr>
<tr>
<td>Old Equation + New X</td>
<td>215</td>
<td>2844.90</td>
<td>-0.110</td>
<td>0.949</td>
<td>390.920</td>
</tr>
<tr>
<td>Old/New Equation</td>
<td>20</td>
<td>2975.00</td>
<td>-0.435</td>
<td>0.998</td>
<td>481.826</td>
</tr>
<tr>
<td>Old/New Equation*</td>
<td>220</td>
<td>2811.80</td>
<td>-0.204</td>
<td>0.989</td>
<td>118.990</td>
</tr>
<tr>
<td>Transfer 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>0</td>
<td>1505.30</td>
<td>-0.060</td>
<td>0.771</td>
<td>28.281</td>
</tr>
<tr>
<td>Old Equation</td>
<td>20</td>
<td>936.31</td>
<td>-0.136</td>
<td>0.736</td>
<td>581.149</td>
</tr>
<tr>
<td>Old Equation + New X</td>
<td>15</td>
<td>1516.60</td>
<td>-0.079</td>
<td>0.736</td>
<td>35.527</td>
</tr>
<tr>
<td>Old Equation + Tf1 X</td>
<td>10</td>
<td>1994.00</td>
<td>-0.059</td>
<td>0.736</td>
<td>471.261</td>
</tr>
<tr>
<td>Transfer 1-Extrap'ed</td>
<td>0</td>
<td>2017.50</td>
<td>-0.050</td>
<td>0.742</td>
<td>500.979</td>
</tr>
<tr>
<td>Old/New Equation-Extrap'ed</td>
<td>30</td>
<td>1271.00</td>
<td>-0.158</td>
<td>0.741</td>
<td>309.306</td>
</tr>
<tr>
<td>Old/New Equation*-Extrap'ed</td>
<td>135</td>
<td>1963.60</td>
<td>-0.070</td>
<td>0.738</td>
<td>548.676</td>
</tr>
</tbody>
</table>

Table 7.3: Parameters of power functions fitted to Total RTs during Training, Transfer 1 and Transfer 2 phases of Experiment 8. Functions labelled "Observed" were fitted directly to the data. Functions labelled "Old Equation" were fitted to values extrapolated from Training performance. Functions labelled "Old Equation + New X" were fitted to values predicted by a new version of the Old Equation which incorporates a change in asymptote (see text for details). Functions labelled "Old Equation + Tf1 X" were fitted to values predicted by a version of the Old Equation which includes the asymptote observed in the Transfer 1 phase (see text for details). The function labelled "Old/New Equation" was fitted to values predicted by a version of the Old/New Equation (see text for details). The function labelled "Old/New Equation*" was fitted to values predicted by the new version of the Old/New Equation which incorporates a change in asymptote (see text for details). The function labelled "Old/New Equation*-Extrap'ed" was fitted to values predicted by the Old/New Equation for the Transfer 2 phase. The function labelled "Old/New Equation*-Extrap'ed" was fitted to values predicted by the Old/New Equation* for the Transfer 2 phase. The function labelled "Transfer 1-Extrap'ed" was fitted to values predicted for the Transfer 2 phase by the function that provided the best fit to the Transfer 1 data (i.e., "Observed" in Transfer 1). \( r^2 = \) proportion of variance in the observed reaction time values accounted for by the predicted values. \( \text{rmsd} = \) root mean squared deviation between predicted and observed reaction time values.
to this data did not account for a large proportion of the variance compared to functions in the previous two phases (i.e., 0.771 vs. 0.989 & 0.998).

All of the functions that were compared with the data in the Transfer 1 phase, except the new version of the Old Equation which incorporates a change in asymptote, were extrapolated into the Transfer 2 phase. Table 7.3 shows that these functions were not able to account for any more variance in the data than the function fitted directly to the data. Figure 7.5 shows that none of these four functions provided a particularly good fit to the data either.

Two different versions of the Old Equation were derived with a change in asymptote and compared to the Total RTs during the Transfer 2 phase. One of these was simply the equation which described improvement of Total RTs during Training with the asymptote observed during the Transfer 1 phase (i.e., X = 1080 ms) and so had the equation \( T = 1080 + 5863.2 P^{-0.74797} \). The second of these functions included an asymptote that was calculated from the difference between the mean Total RT in the first block of the Transfer 2 phase (1521.60 ms) and the value predicted by extrapolating Training performance (914.61 ms):

\[
T_{\text{1st block}} = T_{\text{2nd block}} = X + 5863.2 P^{-0.74797}
\]

\[
1521.60 = X + 5863.2 (11.989)^{-0.74797}
\]

\[
606.99 = X
\]

therefore \( T = 606.99 + 5863.2 P^{-0.74797} \)

Values predicted by both of these equations are displayed in Figure 7.3. It is clear in this figure that only the version that incorporated an asymptote suggested by the Transfer 2 data (i.e., Old Equation + New X) provided a
good account of this data. In fact, the power function fitted to the values predicted by this equation is very similar to the power function fitted directly to the data (see Table 7.3).

The fact that the Old Equation was not able to provide a good account of Total RTs during the Transfer 2 phase calls into question the validity of the independence assumption, that old skills continue to improve according to one function during and after the combination with new skills. However, as was observed with Premise RTs, a new version of the Old Equation which incorporates a change in asymptote provided a good account of performance during this phase. This result suggests that it is possible that old skills do continue to improve according to one function but that the combination with new skills somehow affects the asymptote of this function.

7.3.3 Summary and Conclusions

Experiment 8 was designed to provide a further test of the ability of the Old/New Equation to account for improvement that accompanies the combination of old and new skills. In general the Old/New Equation was able to predict qualitative changes in learning rate but not the amount by which learning rate would be affected. The predicted learning rates were always faster than the observed rates. The new version of the Old/New Equation, which incorporates a change in asymptote, was seen to provide a much better account of the improvement patterns observed during the Transfer 1 phase in both Premise and Total RTs than the original version. This result replicates the results discussed in §5.5 with respect to Experiments 4 and 5. Thus it appears that following partial transfer, learning rate is attenuated and performance asymptote is increased. Possible causes for this effect on asymptotes will be considered in the final chapter.
Another major aim of Experiment 8 was to test an assumption underlying the two versions of the Old/New Equation. Under this assumption, when old and new skills are combined in a new task, the old skills will improve as if nothing has changed. In other words, improvement of old skills will be described by extrapolating the learning function that described the improvement of these skills prior to the combination. Furthermore, if the new skills are no longer used, the old skills will continue to improve according to the original learning function. This last prediction was tested in this experiment.

Initially the results of Experiment 8 did not appear to support the prediction that performance in the third phase of the experiment would be accounted for by extrapolating from Training performance. It was expected that the Old Equation would account for performance during the Transfer 2 phase as indicated by all three measures, but in all cases the observed values were consistently slower than the predicted values. However, it was discovered that the best account of performance in this phase was provided by a version of the Old Equation which included a slower performance asymptote. This suggests that old productions may function autonomously of the development of new productions but that something associated with the combination of old and new skills affects the minimum execution time of these productions (i.e., the asymptote).

In summary, although the independence assumption underlying the Old/New Equation did not receive unqualified support from the results of Experiment 8, it does appear at least to be a reasonable assumption. This conclusion, together with the ability of the new version of the Old/New Equation to account for improvement following partial transfer, suggest that the Old/New Equation and the model of skill acquisition underlying it provide a reasonable
description of the processes underlying the combination of old and new skills. However, there is at least one feature of this combination of old and new skills that is not a property of the learning model associated with the Old/New Equation, and that is its apparent effect on performance asymptote. This issue will be pursued in the following concluding chapter.
### Chapter 8  General Discussion

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.1 Introduction</td>
<td>339</td>
</tr>
<tr>
<td>8.2 The Transfer of Cognitive Skill</td>
<td>339</td>
</tr>
<tr>
<td>8.3 The Shape of Learning Functions following Partial Transfer</td>
<td>343</td>
</tr>
<tr>
<td>8.4 Partial Transfer and Changes in Asymptote</td>
<td>350</td>
</tr>
<tr>
<td>8.4.1 Stimulus Processing and the Change in Asymptote</td>
<td>350</td>
</tr>
<tr>
<td>8.4.2 Partial Transfer and a Controlling Process</td>
<td>353</td>
</tr>
<tr>
<td>8.4.3 Anderson's Account of the Power Law Revisited</td>
<td>355</td>
</tr>
<tr>
<td>8.5 Task Learning Functions as Summary Functions of Production Improvement</td>
<td>356</td>
</tr>
</tbody>
</table>
8.1 Introduction

In this chapter the experiments reported in this thesis will be discussed with respect to the major aims of the thesis. A number of implications for theories of skill acquisition and the training of cognitive skills will be identified in this discussion.

This thesis had two major aims. Part 1 was concerned with the determinants of the transfer of cognitive skill. Part 2 was concerned with examining the effect of transfer on the shape of learning functions. Achieving these aims led to a number of conclusions about transfer. However, these conclusions have in turn raised further questions. These conclusions and their associated questions will be discussed in this chapter.

8.2 The Transfer of Cognitive Skill

In Part 1 of this thesis, the ACT* theory of skill acquisition and transfer was examined in terms of how well it can account for a transfer phenomenon called the contextual interference effect. The ACT* account was contrasted with the intratrial processing account suggested by Carlson and Yaure (1990). The ACT* account was seen as the superior account because it was able to account for more of the experimental results. Furthermore, the hypothesis that transfer is dependent on common productions, which is the heart of the ACT* account, was able to provide a parsimonious connection between differences observed during Training and differences observed during Transfer. In contrast, the intratrial processing account was not able to predict Training differences and so could not associate these with Transfer differences.
With respect to the ACT* account, performance time differences observed during Training were assumed to reflect strategy differences associated with the different Training conditions. That is, the fewer productions executed to perform a task, the faster will be the performance time. However, the conditions in which performance could proceed with fewer productions during Training were seen to involve strategies that were not well-equipped for performance in other conditions. It turned out that the productions that were not executed during Training were the ones that were required in the transfer situations observed. This resulted in partial transfer to the new situation, which was observed as a slowing in performance. Thus the relative differences in performance time during both Training and Transfer could be accounted for by the number of productions executed in each phase. This supported the hypothesis that productions are the units of skilled performance and that the extent of transfer between tasks is determined by the number of common productions executed in the performance of these tasks.

Support for the ACT* account came from its ability to predict systematic deviations in performance during Transfer from that predicted by learning functions established during Training. The ACT* account of the power law of learning has as a major parameter the number of productions being executed in the performance of a task. Performance times which differed from those predicted by extrapolating from Training performance could be accounted for by differences in the number of productions executed during Training and Transfer. This result constitutes further support for the notion that productions, or some sort of identical elements in Thorndike's sense, underlie transfer.

An interesting implication arises from the above conclusion. As was mentioned in Chapters 1 and 2, if the degree of transfer between two tasks is
determined by the extent to which they share identical elements, this appears to preclude any economies of training. In other words, unless a training condition encourages the development of particular skills that are required for performance with a transfer task, little transfer can be expected. Therefore it would appear that no form of training on a particular task could provide a substitute for actual practice with that task. According to Singley and Anderson (1989) though, identifying production rules as the elements underlying skilled performance and transfer provides considerable scope for transfer between tasks. The reason is that productions are abstract rules which may be executed in conditions that are different to those in which they were developed. However, as was demonstrated clearly in Experiments 1 and 3, the abstract nature of productions is not sufficient to enable efficient performance in situations which actually require additional skills.

An interesting feature of Experiments 1 and 3 was that subjects in the Blocked Training and Highlight Training conditions had the same opportunity as Random Trained subjects to learn the critical skill of identifying syllogism type by locating common elements. However particular features of the stimulus conditions encouraged subjects to take advantage of surface information in the syllogisms to solve the syllogisms (e.g., repetition of syllogism type in Blocked Training, highlighted common elements in Highlight Training). As a result these subjects did not develop productions that were critical for efficient performance with syllogisms presented in a random order. The adverse effect on the performance of these subjects was obvious. Hence possessing abstract productions which underlie performance of the greater part of a task was not enough to guarantee efficient performance when the task was such that additional operations were also required. So it would seem that the best form of training to ensure maximal transfer is one where the task being practised is as close as possible to the transfer task. A
related conclusion is that any feature of a training schedule that makes performance easier, such as the highlighting in Experiment 3, may encourage the development of performance strategies which rely on this feature. This is because subjects appear to learn only what is required for efficient performance in the current task. If a transfer task does not include this feature then transfer to this task may only be partial. The reason for this is that the development of a skill necessary for efficient performance in the transfer task is not encouraged by the feature which facilitates performance during training. The adverse effect of such a performance environment has recently been considered with respect to the design of airplanes (Stockton, 1988). Concern has reportedly been expressed that, as computers take over more of the flight control, airplanes may become too easy to fly and, as a result, pilots may lose or never develop the ability to react appropriately in situations for which computer flight systems are not programmed.

The above suggestion as to the best method of training to maximise transfer may be inappropriate in some situations. For example, some tasks will require that components be well-practised before other components are attempted. Thus flight trainees are given practice at manoeuvering a plane during flight before take-offs and landings are attempted. Certainly in such situations only partial transfer from the part task to the whole task is expected. However, although the experiments reported in Part 2 of this thesis demonstrated that recovery from the deficit in performance associated with partial transfer is possible, the combination of well-practised skills with new skills was seen to have long-lasting effects on the progress of this improvement. Thus these effects should also be considered in the development of efficient training programmes.
8.3 The Shape of Learning Functions following Partial Transfer.

The discussion on the shape of learning functions in Chapter 4 began with Anderson's (1982) account of the power law of learning. In this account learning rate appears to be a constant for each person. However, considering that variations in learning rate have been observed within subjects, it seems erroneous to suppose that learning rate is constant.

A model of learning was proposed in Chapter 4 which was designed to account for differences in learning rate that have been reported in previous studies. These studies were also described in Chapter 4. The learning model suggested that learning rate can be affected by a combination of old and new skills. Before the validity of this model is discussed, a number of other factors which can affect learning rate should be considered. One of these factors was obviously operating in Experiment 7 when Flow-Route RTs in the first block of Transfer were slowed considerably compared to later performance in this phase. This resulted in a very fast overall learning rate. When the effect of performance in the first block was taken into account the learning rate estimate was reduced considerably. At least two mechanisms for this effect on learning rate were suggested. One involved a surprise reaction by the subjects invoked by the change in stimulus conditions. The other suggestion concerned a reorganisation of the goal structure underlying performance of the task. Both of these factors would have only temporary effects on reaction time, but their effects on learning rate, as measured by fitting power functions to reaction times throughout the observation period, would be pronounced. A similar effect on the estimation of overall learning rate by a temporary slowing effect on reaction time was apparently evident with Total RTs during the Transfer phase of Experiment 7. Subjects appeared
to stop improving in the last half of this phase, possibly because of boredom or fatigue. This had the effect of raising the collective performance asymptote to an artificially high level. In turn this had the effect of raising the learning rate as estimated by fitting a power function to the data. There is no doubt then that a number of factors contribute to the learning rate that is observed in a particular situation. This suggests that the account of the power law of learning suggested by Anderson requires refinements specifying how the parameters underlying the account (see §4.2) are determined by subject and task factors. One such refinement should involve specification of how the combination of old and new skills can affect learning rate.

The model that was proposed in Chapter 4 to account for deviations in learning rate is concerned with more systematic effects on learning rate than those caused by such factors as motivation, fatigue or surprise. This model predicts that when old and new skills are combined to perform a task, learning rate in this task is slower compared to the rate at which the old skills were originally learned. This attenuation was shown to be a natural result of combining two power functions which describe improvement in the separate (i.e., old and new) components of the task. This combination of two power functions was formalised as the Old/New Equation. The nature of this equation suggested that the amount by which learning rate would be attenuated by the combination of old and new skills would be moderated by two factors: (1) the relative number of processing steps involved in these skills, and (2) the relative amounts of practice each of these skills had prior to their combination. These properties of the Old/New Equation were seen to provide a good account of some of the differences in learning rate that have been reported in previous studies.
In a very general sense, the results of the experiments reported in Part 2 of the thesis were supportive of the predictions made on the basis of the Old/New Equation. However, the main failing of this equation was that it predicted an attenuation of learning rate which was less than that observed. In an effort to account for this failing of the Old/New Equation it was discovered that in a number of cases partial transfer was associated with not only a reduction in learning rate, but also an increase in performance asymptote. The Old/New Equation was modified to incorporate an increase in asymptote and this new version provided a better account of the data, predicting learning rates much closer to those observed. Therefore it appears that the combination of old and new skills attenuates learning rate and increases the asymptotic level of performance.

The validity of the Old/New Equation as a model of improvement that follows partial transfer rests on the validity of two assumptions that underlie this equation. The first of these assumptions states that new skills will improve at the same rate as was observed with old skills, that is, the rate observed when old skills were themselves new. Experiment 7 was designed as a test of the validity of this assumption. However, this test involved some problems. The major problem concerned the most appropriate measure of 'new' learning rate. The major stumbling block in establishing such a measure is that, for adults, virtually no task contains only new components. Thus the easiest way around this problem was to examine the learning rate in two tasks (i.e., Training and Transfer tasks) that involved similar proportions of old and new skills. Similar learning rates in these tasks would go a long way towards validating the assumption. However the results of Experiment 7 did not support this assumption, as the learning rate during Transfer was observed to be much faster than observed during Training. In support of the assumption though, it can always be argued that the measure of 'new' learning rate
established with the Training task was not appropriate because it may have involved a different proportion of old and new skills. This suggests then that the assumption cannot be falsified, as a failure to support the assumption could always be attributed to a difference in the proportion of old and new skills in the two tasks examined. However, this is not the case. There are at least three ways in which the assumption can be falsified.

The first method of attempting to falsify the hypothesis concerning 'new' learning rates is to utilise a two group design. In this design an Experimental group performs with a Training task and then performs with a Transfer task which has as little as possible in common with the Training task. A Control group performs with the Transfer task only. If the assumption of a constant learning rate for new tasks is correct, then learning rates in the Transfer task should be comparable for the two groups. However the problem with such a design is that learning rates between groups vary considerably on the same task, regardless of experimental manipulation. For example, consider the results of Experiments 4 and 8, where conditions during Training were identical. The Experimental group of Experiment 4 had a learning rate of -0.703 with Premise RTs and -1.209 with Conclusion RTs during Training. This contrasts with the learning rates of the subjects in Experiment 8: -0.877 with Premise RTs and -0.602 with Conclusion RTs. Thus between-subjects comparisons of learning rates are not a reliable method of evaluating the assumption concerning 'new' learning rates.

A second method of evaluating the 'new' learning rate assumption is to continue to use within-subjects comparisons of learning rates and to ensure that the measure of 'new' learning rate is made with a task that involves the same proportion of old and new skills as the Transfer task. The best method of ensuring this is to have the old and new skills as similar as possible in the
two tasks. However, the greater the similarity between the two tasks, the
greater the chance of there being transfer from one task to the other, which,
according to the Old/New Equation, will affect learning rate. So increasing
the precision of the measure of 'new' learning rate in this way defeats the
original purpose. Thus testing the assumption with adults should not be
considered viable unless two tasks can be conceived which involve only new
components. This suggests a third method of testing the assumption. This
involves examining the acquisition of skills by infants or young children. The
possibility of developing tasks for these subjects which involve only new
components is far more likely than with adults. Therefore the issue of the rate
at which new skills are learned should be deferred until a more appropriate
testing situation is developed.

The second assumption that underlies the Old/New Equation was termed the
independence assumption. This assumption states that old skills are
unaffected by a combination with new skills. This means that old skills will
continue to improve according to the same function that described the initial
learning of these skills. Experiment 8 was designed to test this assumption by
comparing the learning function observed prior to a combination with new
skills (i.e., in the Training phase) and the function that described performance
when conditions were returned to the original conditions and the new skills
were no longer required to be executed (i.e., in the Transfer 2 phase). If the
independence assumption is correct, it would be expected that the former
learning function would predict the latter learning function. The results
indicated that, to a certain extent, this was the case. However, the function
observed during the Transfer 2 phase was different to the one predicted by the
Training function in one important respect. The Transfer 2 function possessed
a larger asymptote than was observed during the Training phase, even though
the task conditions were identical. Apparently this increase in asymptote was
related to the increase in asymptote observed during the Transfer 1 phase. Therefore it was assumed that whatever caused the increase in asymptote associated with the combination of old and new skills was also responsible for the increased asymptote in the Transfer 2 phase. Thus it was concluded that the independence assumption was a reasonable one if it could be established that the change in asymptote was a reflection of a process independent of the continued improvement of old skills. This point is discussed below (see §8.4).

The major findings of the experiments reported in Part 2 of the thesis are that the independence assumption was validated to a certain extent, that there are considerable methodological difficulties associated with evaluating the assumption concerning 'new' learning rate, and that the new version of the Old/New Equation which incorporates a change in asymptote provides a good account of the pattern of improvement that follows partial transfer. Therefore it appears that this version of the Old/New Equation is a reasonable model of the improvement that follows partial transfer and the processes underlying this improvement. Thus the Old/New Equation provides one account of what causes variation in learning rates within subjects. As Fitts and Posner (1967, p. 19) suggest, except for infants, very few tasks involve only new components. So every task that an adult performs will involve a combination of various skills with varying amounts of practice and which are controlled by varying numbers of productions. According to the Old/New Equation, this will result in varying learning rates for each task for the same subjects, despite the fact that there may be only one basic learning rate underlying all improvement.

The Old/New Equation leads to a number of implications for skills training. The first is that learning a task by practising some components and then
adding further components later in training - a progressive-part training method - will result in an overall learning rate which is slower than if the task is practised as a whole from scratch. The slower rate comes mainly from the phase of training where old and new skills are combined. However, this may not always be a disadvantage. For instance, it may be inefficient to practise some tasks as a whole from the beginning for a number of reasons. One is that the complexity of some tasks may overwhelm the processing capacities of trainees, making any improvement difficult (Anderson, 1982; Woltz, 1988). A second reason is that some tasks involve expensive consequences of failure at the whole task (e.g., flying a plane). In both of these situations, practising some components of a task to a level of proficiency before proceeding to other components of the task has more pragmatic value than a training method designed to maximise learning rate.

A second implication of the Old/New Equation for training concerns the increase in asymptote that is associated with the combination of old and new skills. At present it is not clear whether this means that a progressive-part training method will always result in a performance asymptote which is slower than one resulting from whole-task training. In other words, will the best performance resulting from progressive-part training always be slower than that resulting from whole-task training? The reason that this issue is not clear at present is that the cause of the increase in asymptote is unknown. Two possible explanations will be considered below. One is that the asymptote increase is just an artifact of the change in task. In other words, the asymptote is task-specific, possibly related to the amount of information processed in the task. If this is the case, then progressive-part training and whole-task training should result in the same asymptote. The second possible explanation of the asymptote increase is that the combination of old and new skills invokes an integration process that has its own inherent performance
asymptote, and this adds to the asymptote associated with performing the task components. If this is true, then progressive-part training will always result in a slower performance asymptote than whole-task training.

8.4 Partial Transfer and Changes in Asymptote

As was suggested at the beginning of this chapter, the experiments in this thesis have provided some answers to the questions examined. However, these answers have in turn raised further questions. For example, the fact that the version of the Old/New Equation which incorporates an increase in asymptote provides a good model of partial transfer raises the question of what causes the change in asymptote. Anderson (1989b) suggested that asymptotic performance represents the minimum time to process stimuli and execute a response. This suggestion does not appear to be supported by a change in asymptote, particularly in tasks which are similar in terms of the types of stimuli to be processed and responses to be executed (i.e., the syllogism tasks). Therefore either Anderson's interpretation of asymptotic performance is wrong or the combination of old and new skills in some way affects the minimum time to process stimuli and execute responses. Two possible explanations for the change in asymptote will be considered below.

8.4.1 Stimulus Processing and the Change in Asymptote

One possible explanation for the increase in asymptote associated with partial transfer is suggested by the results of Experiment 7. In this experiment, partial transfer in some of the conditions did not result in a change in asymptote. This observation may indicate that the change in asymptote is peculiar to partial transfer in some tasks only. For instance, there may be some feature of the syllogism task not present in the tank task which affects
the performance asymptote. The only condition in which an increase in asymptote was observed in the syllogism task involved Highlight Training and Random Transfer. This condition was assumed to involve a change in processing strategy from Training to Transfer. That is, new productions were assumed to be developed during Transfer to complement the old productions. The processing strategy developed during Training was assumed to involve detecting the direction of the diagonal of common elements which were highlighted, and then processing the uncommon elements. It was assumed that this strategy would be modified during Transfer to involve locating the common elements by processing each element of the premises one by one, and then proceeding as during Training. Therefore more stimulus processing is performed during Transfer than during Training. Considering Anderson's interpretation of asymptotic performance, this assumed strategy change should result in an increased asymptote. In other words, no matter how much practice subjects have at performing the Transfer task, they will still have to process each of the premise elements in each trial. Thus the minimum time to perform the task must be slower than in the Training task, where less information must be processed to solve the syllogism. In contrast, with the tank task, it could be argued that the stimuli to be processed during Training and Transfer did not differ substantially in the conditions where an increase in asymptote was not observed (i.e., 1-tank RTs and Conclusion RTs). Therefore, given Anderson's interpretation of asymptotic performance, increases in asymptotes would not be expected in these conditions.

There are problems with the hypothesis that the observed increases in asymptotes are related to changes in the amount of information processed in a task. Total RT in the tank task was the only measure where an increase in asymptote was associated with partial transfer. It would be difficult to argue that this measure reflected an increase in the amount of information processed
from Training to Transfer but that 1-tank RTs and Conclusion RTs did not. Thus the Total RTs do not support this explanation of the increase in asymptote. Considering the suggestion that Total RT was probably the most stable measure of performance in the tank task, it would appear to be premature to pay too much credence to this 'amount of information' hypothesis.

The 'amount of information' hypothesis is also unable to account for the increase in performance asymptote from the Training phase to the Transfer 2 phase of Experiment 8. In these two phases the stimulus conditions were identical. Thus the same amount of information should have been processed in both phases and therefore no effect on asymptote should have resulted. However, the 'amount of information' hypothesis could account for this result if it was assumed that the strategy change from Training to Transfer 1 affected the strategy used during the Transfer 2 phase. In other words, having developed the strategy of processing all of the premise elements during the Transfer 1 phase, subjects continued to use, to some extent, a similar strategy during the Transfer 2 phase. As a result, subjects may have processed more of the premise elements than was necessary and than were processed during Training (but less than during Transfer 1 as reaction times were faster during the final phase). This would have increased the minimum processing time and therefore the performance asymptote.

Certainly the disruptive effect of integrating old and new skills has been reported before. For example, Miller and Paredes (1990) found that students learning to multiply performed addition slower during this period than before this period. This result was interpreted by the authors as indicating that learning to multiply was interfering with the well-established skill of performing addition. Miller and Paredes went on to suggest that "within a
domain of knowledge, learning new skills may often require one to rethink or reorganise previous knowledge, leading in turn to a temporary disruption of performance. Finding that such disruptions occur as one acquires new knowledge suggests that new knowledge is being incorporated into a structure containing the previous skill." (1990, p. 239). Such an integration of new knowledge with old knowledge may explain the increase in asymptote from the Training phase to the Transfer 2 phase in Experiment 8. If during the Transfer 1 phase the new skills for locating common elements were integrated with the old skills for performing the rest of the syllogism task, subjects in this experiment may have found it difficult to execute only the old skills during the Transfer 2 phase when these were all that were required. As a result, more information was processed during the Transfer 2 phase than during Training, leading to a slower minimum performance time.

The results of the experiments reported in this thesis do not enable a decision on the validity of the 'amount of information' hypothesis as an explanation of the asymptote effect. Therefore this decision should be deferred until further experimental evidence is obtained. Such evidence would come from manipulating the amount of information to be processed in order to examine the effect on performance asymptote.

8.4.2 Partial Transfer and a Controlling Process

A second possible explanation for the increased asymptote associated with partial transfer involves a process which controls the integration of old and new skills. This explanation suggests that when this process is operating an additional minimum time is added to total performance time. In other words, just as basic stimulus processing and response execution are thought to
involve minimum performance times, so too could a process which integrates two or more skill modules involve such an asymptote.

It is not clear whether the integration process explanation of the asymptote effect can account for the observation of an increased asymptote in the Transfer 2 phase of Experiment 8. In this phase, the new skills developed during the Transfer 1 phase were assumed to no longer be executed, so no integration process should have been operative. However, an increased asymptote was also observed with Conclusion RTs during both of the Transfer phases in Experiment 8. Conclusion RTs were assumed to reflect the execution of old skills only, and this was supported by the fact that the Old/New Equation did not provide a good account of transfer in this measure. However, something increased the performance asymptote in this measure following the combination of old and new skills in Premise RTs. Thus if the effects of the integration process are not specific to the locus of cognition where old and new skills are combined, but are more widespread, then this hypothesis may prove a valid one. However, the mechanism by which such a process could have the observed effect is not obvious, although such a control process has obvious similarities to executive or metacognitive processes that have recently been suggested as important components of general cognitive functioning and therefore intelligence (e.g., Brown, 1978; Carroll, 1981; Sternberg, 1985). Therefore, before this hypothesis warrants further speculation, evidence is required to establish whether it provides a reasonable explanation. Such evidence may come from experiments which compare the performance asymptote of a task practised only as a whole with the asymptote of this task when it involves the combination of old and new skills. The integration hypothesis predicts that the asymptote should be greater in the second case.
8.4.3 Anderson's Account of the Power Law Revisited

Whatever the status of the above explanations for changes in asymptote, the results of the experiments reported in Part 2 of this thesis suggest that something is amiss with Anderson's (1982) account of the power law of learning. In particular this account clearly ignores the importance of asymptotic performance in describing improvement. It will be recalled from Chapter 4 (see § 4.2) that Anderson's account does consider asymptotes in the derivation of a power function that describes learning. However, in the following combination of two power functions which describe the effects of algorithmic improvement and strengthening on performance time:

\[ TT = (N^* + N_0P^{-f})(C + AP^{-g}) \] (3)

the asymptotes of these two functions, \( N^* \) and \( C \), are both assigned zero values in order to arrive at the following simple form of a power function:

\[ TT = N_0AP^{-(f+g)} \] (4).

Equation 4 was further simplified to:

\[ T = N_0P^C \] (5),

a power function with an asymptote of zero. The problem with this derivation is that when data is described by a power function with a non-zero asymptote, the asymptote apparently has no theoretical relation to the derivation, in contrast to the other parameters of the function. Instead the asymptote is simply "tacked" on to Equation 5 in order to improve the fit to the observed data. Therefore it appears that the assumption that \( N^* \) and \( C \) are both zero is apparently not a valid one. Certainly Anderson's own definitions of these parameters suggest their roles in skill acquisition should not be ignored: \( N^* \) represents the minimum number of productions that constitute the optimal procedure for performing a task, and \( C \) represents the minimum time for execution of a certain number of productions. It is obvious then that both of these parameters are of considerable relevance to the above discussion.
concerning changes in asymptote. Therefore the more complete expansion of 
Equation 3 -

\[ TT = CN^* + CN_0P^{-f} + N^*AP^{-g} + N_0AP^{-(f+g)} \]  

- may need to be considered in further examinations of the asymptote 
question.

8.5 Task Learning Functions as Summary Functions of 
Production Improvement

Although the cause of the increase in performance asymptote associated with 
partial transfer is not clear, the results of the experiments reported in Part 2 of 
the thesis do allow some firm conclusions. One is that the parameters of 
learning functions, such as learning rate and asymptotes, are by no means 
constants of the cognitive systems of subjects. It is clear that these can be 
fected by the combination of old and new skills, and it is also reasonable to 
suggest that they are also affected by factors such as motivation, fatigue and 
surprise. Thus, as suggested above, there is evidence that is counter to what 
is implied by Anderson's (1982) account of the power law of learning, that 
the rate at which performance on a task improves is a constant of the learning 
system.

The fact that the rate at which performance on a task improves can be a 
function of previous experience with components of the task is not 
surprising. Certainly it would be expected that learning a task would be 
facilitated if a trainee is familiar with parts of the task. However, what is 
important to realise is the apparent distinction between the learning of a task 
and the learning of components underlying performance of the task. 
Experiments in Part 2 of this thesis demonstrated that improvement on a task 
could be accounted for by considering the relative contributions of task
components that had a history of performance and task components that had not been practised before. The Old/New Equation, which was designed to describe such a situation, was able to predict that such a combination of old and new skills would affect learning rate. This effect involved an attenuation of learning rate compared to the rate at which the old task components were originally learned. Thus the learning rate of a 'new' task was shown to be different to that of a previously observed new task, that is, the task in which the old task components were learned. The significance of this finding is that the rate of improvement on a task is determined by the history of the components which underlie performance of the task, rather than experience with the task itself.

The ability of the Old/New Equation to account for improvement on tasks which involved the combination of old and new skills centred on the assumption that the learning rate of new skills would be the same as the rate at which the old skills were originally learned. Even though this assumption did not receive direct support from the experiments reported in this thesis, it appears a reasonable assumption given the superior ability of the Old/New Equation compared to other functions to account for improvement following partial transfer. Thus the variability in task learning rates appears to result from the combination of underlying skills with varying application histories that all improve at the same rate. Therefore, contrary to what is implied by Anderson's account of the power law of learning, the rate of improvement on a new task is not directly determined by parameters of the cognitive system. Instead it is the rate of improvement of productions which underlie performance of the task that is determined by the parameters of the cognitive system. Only when performance on a task relies on the execution of completely new productions will the task learning rate be the same as the cognitive system's learning rate. However, given that for adults most tasks
involve the execution of productions that vary in the extent to which they have been practised, improvement on a new task is unlikely to ever be at the same rate at which the underlying components improve.

The picture of skill acquisition that emerges from the above discussion of learning rates is one where cognitive representations such as production rules are the basic units of skilled performance. When a number of new productions are executed together to perform a task, performance will improve according to a power function that has a learning rate determined by the parameters of the cognitive system. If the task conditions applicable to the successful execution of these productions continue to be present, then performance on the task will continue to improve according to the same power function. This will be the case even if task conditions change, as long as the appropriate conditions for the execution of the productions remain in the stimulus environment or are produced by the execution of other productions. Thus collections of productions can develop the appearance of skill modules, where changes in task conditions do not affect their execution, nor the pattern of their improvement. New productions may develop alongside these already well-practised productions as task conditions dictate. These new productions will improve according to a different power function, although one with a learning rate again determined by the parameters of the cognitive system. In other words, new productions will improve at the same rate as the one which describes the improvement of old productions. Thus old and new productions will improve together according to their own learning functions. Although these learning functions will have the same basic learning rate, the momentary learning rates of the two sets of productions will be different because these productions are at different points along their learning functions. Performance on the task as a whole will improve at a rate that is not the same as the learning rate of new productions. In fact, the learning rate
on the total task will be a function of the combination of the separate learning functions which describe the improvement of the different sets of productions which underlie performance on this task. Thus the power function which describes improvement on the total task can be seen as an aggregate of the learning functions of the components which underlie performance of this task. In other words, just as performance time on the total task is a function of the time to execute all of the productions which contribute to the overall performance, so is the learning function for performance on the total task a function of the learning functions describing improvement in all of these productions.

Although the simple addition of two power functions which describe improvement in old and new skills provides a reasonable account of changes in learning rate, more complicated tasks than those examined in this study will no doubt involve more complicated combinations of functions. Even in the present study such a situation is conceivable. In Experiment 1, Blocked Training was suggested to involve the execution of some productions on every trial (e.g., those concerned with processing the uncommon elements of premise pairs in order to derive an expected true conclusion) and some productions only once per block of trials (e.g., those concerned with identifying syllogism type). The latter set of productions therefore would be practised less than the former set and so improvement in these two sets of productions should be described by different power functions. The learning function which described improvement during Blocked Training would therefore have been a summary of at least two functions. This simple example suggests that the large variation that is observed in learning rates for different tasks, both between and within subjects, may be a result of examining performance on tasks which involve various combinations of productions with varying practice histories. This apparently simple explanation for
variations in learning rate belies the complexity which underlies the learning of a task. For example, it would be an immensely difficult exercise to determine the relative contributions to the overall pattern of improvement in a task that are made by, on one hand, the number of productions involved in various task components and, on the other hand, the practice histories of these productions.

The fact that new tasks rarely involve the execution of new productions only calls into question the ability of simple power functions to account for improvement with practice. Although such functions may be able to describe the improvement that is observed, the parameters of such functions will not be accurate reflections of features of the components underlying performance. For instance, in Anderson's (1982) account of the power law of learning, the coefficient of a power function (i.e., \( N \) in \( T = NPC \)) was said to be proportional to the number of productions executed in the performance of a task. However, this will only be the case if all of the productions being executed are new. If some of the productions are old, and therefore are further along their learning functions than the new productions, then this coefficient will underestimate the number of productions being executed. Similarly, Newell and Rosenbloom's (1981) revised version of the simple power function (i.e., \( T = N (P + E)^c \)) may be able to account for improvement on a task that relies on the execution of old productions only, but all of these productions must have been practised to the same extent for the parameters of this function to be informative (i.e., \( E \) must be the same for all of the productions). Therefore when a task involves the execution of productions with varying practice histories, as most tasks will, the best way to describe improvement on the task is to combine separate power functions which describe the improvement of the various sets of productions. The difficulty however is in estimating a trainee's experience with particular task
components and the number of productions that are executed to perform these components. The degree to which such an exercise is undertaken will depend on the accuracy that is desired from a function designed to describe improvement on the task.

In conclusion, the experiments reported in this thesis have demonstrated that, despite the variability inherent in the parameters of learning functions, with careful investigation the variability of these parameters can provide insights into the processes underlying skill acquisition and transfer, and more generally, the nature of cognitive functioning.
References


