Induction in Fluid Intelligence: Knowledge, Novelty, Learning and Proactive Interference

Myvan Bui

B Econ (Soc Sci) (Hons I, Psych)

School of Psychology
University of Sydney
Australia

A thesis submitted in fulfilment of the degree of
Doctor of Philosophy

31 March 2010
Copyright 2010
Myvan Bui
All Rights Reserved.
STATEMENT OF ORIGINALITY

This is to certify that this thesis has not been submitted for a higher degree to any other university or institution.

The source of the information contained herein is original and is solely the work of the author, except as otherwise indicated in the text.

Myvan Bui

PhD Candidate
I dedicate this to my husband, Chris

(Harrow, Chrissy! 😊)
ACKNOWLEDGEMENTS

There are a number of people that I would like to thank. First and foremost I would like to thank my supervisor Damian Birney who consistently went above and beyond his role as supervisor. Thank you for your enthusiasm, energy, compassion and integrity. I would also like to thank Jens Beckmann and Amirali Minbashian (especially for their comments and assistance towards the end of my candidature when I needed it the most), and everyone part of the “Personality and Individual Differences Lab” (Sydney University), “Accelerated Learning Lab” (AGSM, UNSW), Insight SRC and “Student Room 310 Antarctica” (Sydney University) for various, numerous (including humorous) things.

I would like to thank my parents Huyen Nhung, Hieu Cuong, and my little sister Thien. Thien, thank you for your ability to make me laugh and love life no matter what the situation. I would also like to thank and acknowledge my extended family and friends, especially Thea Ormerod, Janet Watson Kruse and Barry Cook who inspired me to hope, love, have faith and to work harder for the right things.

Lastly, but most ardently, I would like to thank and acknowledge my husband Chris. Thank you for working so hard to support us – allowing us to buy the nice things in life like food, shelter, clothing; and of course, thank you for your love and ridiculously ridiculous sense of humour.
# TABLE OF CONTENTS

## CHAPTER 1: INTRODUCTION

1.1 Can Cognitive Psychology Tell Us Why People Differ in Differential Psychology’s Gf? ........1

1.2 The Problem: Cognitive Theories Conflict with the Conceptualisation of Gf .........................2

1.3 A Possible Solution: The Learning Hypothesis ....................................................................3

1.4 Overview of the Thesis .........................................................................................................5

## CHAPTER 2: INTELLIGENCE & FLUID INTELLIGENCE

2.1 Introduction .........................................................................................................................8

2.2 The Definition of Intelligence – or Lack Thereof .................................................................8

2.3 The Differential Approach – The Origins of Gf .................................................................11

2.3.1 Criticisms of the Differential Approach .......................................................................17

2.4 Going beyond Circularity and Reification ..........................................................................18

2.5 Cognitive-Differential Approach .......................................................................................19

2.6 Empirical Findings and Unknowns about Gf: Reasoning, WMC and Learning .............21

2.6.1 A Neglected Process: Learning ...............................................................................26

2.7 The Difficult Task of Linking Ability to Processes using Theory .................................27

2.8 Summary ..........................................................................................................................30

## CHAPTER 3: REASONING & INDUCTION

3.1 Introduction .........................................................................................................................33

3.2 What is Reasoning .............................................................................................................33

3.4 Reasoning Theories in Psychology ..................................................................................35

3.4.1 Spearman’s Cognitive Theory of “g” (Eduction of Relations and Correlates) .............37

3.4.2 Sternberg’s Unified Theory of Reasoning (Conceptual Knowledge) .........................39

3.4.3 Holland et al.’s Theory of Induction (Learning and Conceptual Knowledge) ..............44

3.5 Theoretical Similarities and Contrasts: Implications for General Processes and Individual Differences involved in Gf tasks ...............................................................................46

3.6 Central Thesis and General Predictions: Knowledge, Novelty, Learning and Combating Proactive Interference are all involved in Gf Task Performance ..............................................49

3.7 Summary ..........................................................................................................................55

## CHAPTER 4: EXPERIMENT 1

4.1 Introduction .........................................................................................................................57

4.2 Learning and Raven ..........................................................................................................58

4.3 Raven’s Advanced Progressive Matrices ..........................................................................59

4.4 Search for Sources of Individual Differences in Raven Performance ...............................62
4.4.1 Amount of Information ........................................................................................................ 62
4.4.2 Learning Opportunity ........................................................................................................ 63
4.4.3 Perceptual Organisation ..................................................................................................... 64
4.4.4 Rule Type ........................................................................................................................... 67
4.5 Combining the Evidence on Sources of Individual Differences ............................................. 70
4.6 Competing Hypotheses and Specific Predictions ................................................................... 73
4.7 Method ................................................................................................................................... 80
  4.7.1 Participants ....................................................................................................................... 80
  4.7.2 Cognitive Tasks ................................................................................................................ 81
  4.7.3 Procedure ........................................................................................................................ 83
4.8 Results and Discussion ........................................................................................................... 84
  4.8.1 Descriptive Statistics ....................................................................................................... 84
  4.8.2 Preliminary Analyses ....................................................................................................... 87
  4.8.3 Analysis of Learning Sets (Manipulation Check) ............................................................. 88
  4.8.4 Analysis of Transfer Set .................................................................................................. 92
  4.8.5 Analysis of Transfer Set when Actual Learning is taken into Account ............................ 97
4.9 General Discussion ................................................................................................................ 104
  4.9.1 The Effects of the D2 Manipulation on Group Means .................................................... 104
  4.9.2 The Effects of the D2 Manipulation (Individual Differences taken into Account) .......... 106
  4.9.3 Summary and Conclusion ............................................................................................... 107
  4.9.4 Limitations and Further Research .................................................................................. 108
CHAPTER 5: EXPERIMENT 2 ........................................................................................................ 110
  5.1 Introduction ......................................................................................................................... 110
  5.2 Issues Raised by Experiment 1, Chapter 4 ........................................................................... 111
  5.3 Within-Item Reasoning ....................................................................................................... 114
    5.3.1 Within-Item Rule Discovery and Across-Item Learning: Similarities and Differences ... 118
  5.4 Application of Prior Knowledge, Across-Item Learning or Within-Item Rule Discovery: It’s
      Probably All Three? ............................................................................................................. 118
  5.5 The Cost of Learning: The Proactive Interference Hypothesis ............................................. 121
  5.6 Learning and Proactive Interference: Two Sides of the Same Coin? ................................. 122
  5.7 Summary of Aims and General Hypotheses ....................................................................... 124
  5.8 Teasing Apart Prior Knowledge, Across-Item Learning and Within-Item Rule Discovery: The
      “Modified Sweller & Gee Task” ......................................................................................... 126
  5.9 A Methodological Issue: The Marker Task and Its Presentation Order............................. 130
  5.10 Experiment 2: A Pilot Study ............................................................................................ 131
    5.10.1 Specific Predictions .................................................................................................... 132
5.10.2 Method ............................................................................................................................. 133
5.10.3 Cognitive Tasks .................................................................................................................. 134
5.10.4 Procedure ........................................................................................................................ 138
5.11 Results & Discussion .............................................................................................................. 138
  5.11.1 Descriptive statistics ........................................................................................................ 139
  5.11.2 Learning and Proactive Interference at the Group Level ................................................. 141
  5.11.3 Examinations of Individual Differences .......................................................................... 143
  5.11.4 MSG Task and Significance Testing ................................................................................ 149
5.12 Preliminary Conclusions ......................................................................................................... 155

CHAPTER 6: EXPERIMENT 3 ...................................................................................................... 159
  6.1 Experiment 3: The MSG Task and Number Series ............................................................ 159
  6.2 Method .................................................................................................................................. 160
  6.3 Results & Discussion ............................................................................................................. 162
    6.3.1 Descriptive Statistics ....................................................................................................... 163
    6.3.2 Correlations ..................................................................................................................... 166
    6.3.3 Non-parametric Comparisons of Gf Ability Groups (Based on NSEH) ......................... 168
    6.3.4 Hierarchical Linear Modelling with NSEH ................................................................. 173
    6.3.5 Examination of NSH High Performers ........................................................................... 183
  6.4 Summary and General Discussion ....................................................................................... 185

CHAPTER 7: EXPERIMENT 4 ...................................................................................................... 190
  7.1 Experiment 4: The MSG Task and Letter Series ............................................................... 190
  7.2 Method .................................................................................................................................. 191
  7.3 Results & Discussion ............................................................................................................. 193
    7.3.1 Descriptive Statistics ....................................................................................................... 194
    7.3.2 Correlations ..................................................................................................................... 197
    7.3.3 Non-parametric Comparisons of Gf Ability Groups (Based on LSEH) ......................... 199
    7.3.4 Hierarchical Linear Modelling with LSEH ................................................................. 203
    7.3.5 Examination of LSH Higher Performers ....................................................................... 213
  7.4 General Discussion and Summary of Results for Experiments 2, 3 and 4 ....................... 215
    7.4.1 Implications for Processes Involved in Gf and Individual Differences ....................... 221
    7.4.2 Significance and Limitations of Experiments 2, 3 and 4 ............................................. 223

CHAPTER 8: DISCUSSION & CONCLUSION .............................................................................. 226
  8.1 Summary of Aims and Hypotheses ....................................................................................... 226
  8.2 Summary of Findings ........................................................................................................... 227
  8.3 Implications for Processes Involved in Gf tasks ................................................................. 229
8.3.1 Complex Items: Rule Discovery and Learning Processes ................................................. 229
8.3.2 What About Early Simple Items? ...................................................................................... 231
8.3.3 What about Proactive Interference? ................................................................................... 232
8.4 Implications for Cognitive Theories ......................................................................................... 233
8.5 Implications for Gf .................................................................................................................... 234
8.6 Conclusion ................................................................................................................................ 235

CHAPTER 9: REFERENCES ........................................................................................................ 236

APPENDIX A .............................................................................................................................. 241

A.1 Items Generated for the Raven Learning Sets......................................................................... 241
A.2 Characteristics of Learning Set items used in each condition and their equivalents in the original Raven (Advanced Progressive Matrices)........................................................................... 247
A.3 Transfer Set item characteristics ............................................................................................. 248
A.4 Regression model for variables predicting performance on Learning Set................................ 248
A.5 Working out for probabilities of chance level performance for Raven’s APM Transfer Set... 249
A.6 Regression model for variables predicting performance on Transfer Set............................... 249

APPENDIX B .............................................................................................................................. 250

B.1 MSG Items with all of their correct target-answer pairs and rules......................................... 250
LIST OF FIGURES

Figure 4.1. An example to illustrate the format of Raven test items.................................61

Figure 4.2. An item that illustrates a simple example of the correspondence finding/abstraction requirement and the distribution of two rule.............................................65

Figure 4.3. Predicted group means on the transfer items for the three conditions, as predicted by the three competing hypotheses.................................................................78

Figure 4.4. Predicted correlations between the transfer items and a Gf marker, for the three conditions, as predicted by the three competing hypotheses........................................79

Figure 4.5. Predicted differences in means on the Transfer Items for High Gf and Low Gf groups, for the three conditions, as predicted by the three competing hypotheses.....79

Figure 4.6. Point estimates of mean Learning Set performance for those 1 standard deviation below and those 1 standard deviation above the mean on Number Series (Gf), in the three conditions........................................................................................................89

Figure 4.7. A comparison of proportion correct for each Learning Set item in the t conditions and the Raven Manual. Circled points indicates item containing the D2 rule...........91

Figure 4.8. Point estimates of mean Transfer Set performance for those 1 standard deviation below and those 1 standard deviation above the mean on Number Series (Gf), in the three conditions........................................................................................................95

Figure 4.9. Three graphs showing point estimates of mean Transfer Set performance for those 1 standard deviation below, above and at the mean on Number Series (Gf) at 1 standard deviation below and above the mean for Learning Set performance (From Model 3, Table 4.6) for each condition, No, Limited and Most.................................100

Figure 5.1. Screen shot of an example item from the instructions that participants get for the task.........................................................................................................................128

Figure 5.2. Trimean MSG scores across the task for Experimental groups, by LSEH ability groups..............................................................................................................147

Figure 5.3. Trimean MSG scores across the task for Experimental groups, by NSR ability groups..............................................................................................................148

Figure 5.4. Trimean MSG score on Complex and PI items for Experimental and Control Conditions, by LSEH ability grouping.................................................................150

Figure 5.5. Trimean MSG score on Complex And PI items for Experimental and Control Conditions, by NSR ability groups..........................................................................150

Figure 6.1. Trimean MSG scores on Complex and PI items, for Experimental and Control Conditions, by NSEH ability groups.........................................................................171
Figure 6.2. Trimean MSG scores across the task for Experimental groups, by NSEH ability groups.................................................................172

Figure 6.3. Predicted probability of success (Correct = 1) as a function of Condition, NSEH and Attempt, on Complex Item, based on Model 1 from Table 6.7...............................179

Figure 6.4. Predicted probability of success (Correct = 1) as a function of Condition, NSEH and Attempt, on Complex Item, based on Model 1 from Table 6.8..............................183

Figure 6.5. Trimean MSG scores on Complex and PI items for Experimental and Control Conditions, for those in the top 5% of NSH. .................................................................185

Figure 7.1. Trimean MSG scores on Complex and PI items for Experimental and Control Conditions, by LSEH ability groups.................................................................202

Figure 7.2. Trimean MSG scores across the task for Experimental groups, by LSEH ability groups.................................................................202

Figure 7.3. Predicted probability of success (Correct = 1) as a function of Condition, LSEH and Attempt, on Complex Item, based on Model 1 from Table 7.7.................................209

Figure 7.4. Predicted probability of success (Correct = 1) as a function of Condition, LSEH and Attempt, on Complex Item, based on Model 1 from Table 7.8.................................213

Figure 7.5. Trimean MSG scores on Complex and PI items for Experimental and Control Conditions, for those in the top 5% of LSH.................................................................214
LIST OF TABLES

Table 2.1 Broad Factors and their Narrow Abilities, from the Cattell-Horn-Carroll (CHC) theory of cognitive abilities. From Alfonso et al. (2005).................................15

Table 3.1 Example “mediators” from Sternberg’s (1986) Unified Theory of Reasoning.......42

Table 4.1 Demographic information.................................................................................85

Table 4.2 Descriptive statistics for tasks used in the study, for each condition...............86

Table 4.3 Correlations between tasks used in the study, for each condition...............87

Table 4.4 Differences among conditions on Learning Set performance based on estimated marginal means (with Number Series as a covariate).................................90

Table 4.5 Means for individual Transfer Set items and their tests against chance level.....93

Table 4.6 Summary of hierarchical regression analyses for variables predicting Transfer Set performance in the three conditions.........................................................99

Table 4.7 Estimates for three simple slopes of Transfer Set performance as a function of Learning Set performance, at three different levels of Number Series (Gf) performance in the Most condition...........102

Table 5.1 MSG item rules based on the conditions in which they appeared...............136

Table 5.2 Demographic information.................................................................................139

Table 5.3 Descriptives for Gf marker tasks......................................................................139

Table 5.4 Descriptive statistics for scores on the Modified Sweller & Gee task............140

Table 5.5 Spearman’s rho correlations between Gf marker tasks and MSG items...........144

Table 5.6 Cross-tabulation of participants in each Gf ability group...............................145

Table 5.7 Descriptive statistics and F-tests for LSEH (for Low and High LSEH groups) and NSR (for Low and High NSR groups), across the three conditions.................146

Table 5.8 Descriptive statistics and Mann-Whitney tests for MSG task scores across conditions, within Gf ability groups............................................................151

Table 6.1 Demographic information.................................................................................163

Table 6.2 Descriptive statistics for Gf marker tasks................................................................164

Table 6.3 Descriptive statistics for Modified Sweller & Gee task....................................165
Table 6.4 Spearman’s rho correlations between Gf marker tasks and Modified Sweller & Gee items .................................................................................................................................167

Table 6.5 Descriptive statistics and F-tests for Low NSEH and High NSEH groups, on NSEH across conditions ..................................................................................................................169

Table 6.6 Descriptive statistics and Mann-Whitney tests for MSG task scores across conditions, within Gf ability groups .................................................................................................................169

Table 6.7 Estimated parameters of the Hierarchical Linear Model for performance on the Complex Item ............................................................................................................................................176

Table 6.8 Estimated parameters of the Hierarchical Linear Model for performance on the PI Item ...............................................................................................................................................181

Table 7.1 Demographic information .................................................................................................................................194

Table 7.2 Descriptive statistics for Gf marker tasks ..................................................................................................................195

Table 7.3 Descriptive statistics for Modified Sweller & Gee task .................................................................................................196

Table 7.4 Spearman’s rho correlations between Gf marker tasks and Modified Sweller & Gee Items ...............................................................................................................................................198

Table 7.5 Descriptive statistics and F-tests for Low LSEH and High LSEH groups, on LSEH across conditions ..........................................................................................................................................200

Table 7.6 Descriptive statistics and Mann-Whitney tests for MSG task scores across conditions, within Gf ability groups ..................................................................................................................................200

Table 7.7 Estimated parameters of the Hierarchical Linear Model for performance on the Complex Item ...............................................................................................................................................206

Table 7.8 Estimated parameters of the Hierarchical Linear Model for performance on the PI Item ...............................................................................................................................................211
ABSTRACT

The main aim of this thesis was to examine whether learning processes occur in fluid intelligence (Gf) tasks, whether it is essential for them to occur for induction to take place and whether they contribute to individual differences in performance. In mainstream differential research, Gf is conceptualised as a factor important in induction tasks that are considered novel and context-free (Cattell, 1963, 1987). Thus, performance has typically been assumed to be uninfluenced by previous acquisitions of knowledge structures. Sources of individual differences in Gf task performance have been attributed to working memory capacity (WMC), particularly individual differences in the ability to combat proactive interference. In contrast, the cognitive reasoning literature associates induction with the use of prior conceptual knowledge. A middle-ground position is that Gf tasks may require learning to occur across the task, which would draw upon WMC. That is, individual differences in Gf task performance may be due to knowledge learnt across the task, rather than knowledge brought to the task. Gf items have traditionally been presented in easy-to-hard order but easier items may unintentionally provide learning opportunity for harder items. This would contradict both classic and modern test theories which make the assumption that items within a task are independent of each other.

The learning hypothesis was explored in the current work along with the issue of whether it is possible to reliably solve complex Gf items without some relevant, prior knowledge. Also, the distinction between within-item induction and across-item learning was investigated, along with the relationship between across-item learning and proactive interference. An experimental-differential approach was used to manipulate learning opportunity within Gf tasks in four experiments.

The first experiment examined whether learning takes place in Raven’s Advanced Progressive Matrices (Raven, 1962) and if so, to what extent this learning is a source of individual differences. Specifically, whether rule learning within the task is necessary for abstraction to take place and whether those of higher Gf ability learn faster than those of lower Gf ability.

The next three experiments examined the distinction between knowledge that may be brought to the task, learning that occurs across multiple items in the task and induction within a single item that may be independent of any prior knowledge including knowledge learnt across the
task. The effect of proactive interference as a consequence of learning and knowledge was also investigated. The experiments examined which of these are relevant to general performance (i.e., common to everyone) and which contribute to individual differences. Learning-opportunity was manipulated in a task from the cognitive reasoning literature – the *Modified Sweller and Gee (MSG) Task*. Traditional Series Completion tasks were used as Gf markers and data analyses employed Hierarchical Linear Modelling (HLM).

The advantage of the MSG Task is that it has qualities typical of Gf tasks but unlike conventional Gf tasks, it is able to assess within-item induction in isolation from any potential influences from across-item learning. This is because it involves multiple attempts within each item with feedback, allowing single items to be administered reliably. When across-item learning opportunity is absent, the MSG Task is able to provide an estimate of participants’ within-item induction success through the number of attempts they need within a single item. The amount participants learn across items can be approximated by comparing performance on items preceded by learning opportunity (i.e., easier items with similar rule-types), with items not preceded by learning opportunity. Lastly, the effects of proactive interference can be evaluated by comparing performance on items preceded by interference (i.e., items with different rule-types) with those that are not preceded by interference.

Overall, it was found that with no learning opportunity leading up to novel items (to provide relevant prior knowledge), solution was nearly impossible for all participants. When learning opportunity was provided, all participants were able to greatly improve their performance but those of higher Gf improved more. It was concluded that while Gf tasks appear visually novel, they must contain a combination of familiar elements in earlier items (which make use of knowledge that participants bring to the task) and novel elements in later items (which require the use of knowledge that must be learnt from earlier items); and those of higher Gf perform better on Gf tasks, at least partly because they are able to benefit more from the learning opportunity provided by earlier items. It was found that proactive interference affects all participants when they learn from prior items. However, insufficient evidence was found to suggest that the ability to combat proactive interference contributes to individual differences in performance.
CHAPTER 1

INTRODUCTION

1.1 Can Cognitive Psychology Tell Us Why People Differ in Differential Psychology’s Gf?

The main thesis of this dissertation can be stated quite simply: While prior experience and knowledge may be important to inductive reasoning, what contributes to individual differences in Gf is dynamic and context-free learning.

The overarching aim of this thesis is to gain a better understanding of Gf ability. In the differential literature, the quintessential nature of intellectual functioning has often been associated with general fluid intelligence (Gf) (Cattell, 1963; Gustafsson, 1988). While the Gf construct has reliably been identified using correlational methods such as factor analysis, its exact nature is still rather poorly understood (Carlstedt, Gustafsson, & Ullstadius, 2000). Much research effort has been devoted to understanding Gf itself and what causes people to differ on Gf. Traditionally, Gf has been defined as the ability to reason and solve novel problems that do not rely extensively on an explicit base of knowledge such as that which might be developed from schooling or previous experience (Carpenter, Just, & Shell, 1990). This definition is based on the observation that tasks which load on Gf are often abstract, inductive reasoning tasks that require participants to induce novel rules that govern the relationship between abstract stimuli (such as geometric shapes, numerals or letters). This thesis aims to gain a better understanding of Gf, through investigating why people differ on Gf tasks.
A proper explanation of why people differ on Gf tasks would require a description of the processes involved in performing the tasks and the ways in which these processes differ in different people (Borsboom, Mellenbergh, & van Heerden, 2003; Floyd, 2005; Schlinger, 2003). Many researchers have recommended that differential psychology take note of process theories from cognitive psychology (Cronbach, 1957; Deary, 2001; Lohman, 2000).

However, combining the two approaches is not without some complications. Cognitive psychologists have been concerned with how people reason in general and have developed theories that account for processes that are common to all people (Mackintosh, 1998). Studies and theorising have been aimed at isolating reasoning processes common to all. In contrast, the abilities identified by differential psychology are based on individual differences (variation in performance). Thus, theories from cognitive psychology are not easily mapped onto the Gf construct.

### 1.2 The Problem: Cognitive Theories Conflict with the Conceptualisation of Gf

The difficulty of using theories from cognitive psychology to explain differences in performance on Gf tasks is illustrated by cognitive theories of induction, which seem to conflict with differential psychology’s conceptualisation of Gf. Cognitive theories of inductive reasoning that are broad enough to be applicable to all Gf tasks, emphasise the importance of prior conceptual knowledge in the inductive process (Holland, Holyoak, Nisbett, & Thagard, 1989; Sternberg, 1986). For example, according to Holland et al. (1989), induction cannot take place without previous knowledge. Due to the nature of inductive reasoning (i.e., reasoning in the absence of complete information), the solution to inductive problems can only be considered plausible (rather than correct) at best. Thus, whether a solution could be characterized as plausible can be determined only with reference to the
current knowledge of the person. Hence, induction is highly context dependent because it is
guided by prior knowledge. This conflicts with the differential literature’s conceptualisation
of Gf as a context-free ability to solve novel problems (particularly inductive problems) that
do not rely extensively on an explicit base of knowledge.

1.3 A Possible Solution: The Learning Hypothesis

It is possible to reconcile the two disparate views of induction and Gf, if cognitive processes
involved in Gf tasks are conceptualised as three broad, non-mutually exclusive types:

1) Processes that occur.

2) Processes that need to occur for induction to take place.

3) Processes that contribute to individual differences.

With regard to “processes that occur”, it is possible that not all processes that occur in Gf
tasks need to occur for induction to take place nor do they all contribute to individual
differences in performance on that task. For example, during the course of solving a Gf task,
participants may be reminded of similar tasks. This process does not need to occur for
induction to take place nor would it contribute to performance on that task.

Similarly, with regard to “processes that need to occur”, it is possible that not all processes
that need to occur for induction to take place, contribute to individual differences in
performance. For instance, the use of conceptual knowledge may be required in induction but
it may not contribute to individual differences in Gf task performance and hence, would not
be considered a defining feature of the Gf construct (which is defined by variation in
performance).
Instead, what contributes to individual differences may be learning – a dynamic, fluid, context-free process. We argue that learning processes as a source of individual difference in Gf has been undervalued in mainstream research. That is, we will argue that research from disparate fields converge on the notion that what contributes to individual differences in Gf may be learning – a dynamic, fluid, context-free process. We also propose that learning processes that occur across items within Gf tasks may act as the mediator between the novelty in Gf items and the knowledge needed to solve them. If correct, Gf (being an individual differences construct) could still legitimately be conceptualised as fluid and context-free.

A major focus of this thesis will be to empirically determine whether learning occurs in Gf tasks, whether it needs to occur for induction to take place and whether it contributes to individual differences. There is certainly evidence in both the differential literature (Carlstedt et al., 2000; Verguts & De Boeck, 2002a, 2002b) and the cognitive literature (Anderson, 1993; Gick & Holyoak, 1983; Holland et al., 1989) that learning is important in induction. However, there is also research that suggests that learning does not occur in Gf tasks (Alderton & Larson, 1990; Sternberg et al., 2002), that it does not need to occur (Sternberg, 1986; Sternberg & Berg, 1986) and that it does not contribute to individual differences in performance on Gf tasks (Unsworth & Engle, 2005a). Furthermore, both classic and modern test theory make the assumption that conditional on person ability and item difficulty, items within a task are independent of each other. That is, items can in principle be presented in any order without changing the nature of the construct being measured. However, if learning occurs from one item to another, this would contradict the common test-theory assumption of item independence (D. P. Birney & Sternberg, 2006). Thus, the questions of whether learning occurs within Gf tasks and how it contributes to performance in Gf tasks are worth investigation.
1.4 Overview of the Thesis

We will be using a combined cognitive-differential approach to manipulate learning opportunity within Gf tasks in four experiments. Specifically, we will be using and synthesising various theories of induction to gain a better understanding of how learning processes might operate in Gf tasks and to derive hypotheses about the outcome of learning manipulations in Gf tasks. Cognitive theories may give us insights into the cognitive processes involved in Gf task, but they are hypotheses that need to be empirically tested, particularly because it is not clear how or if they apply to reasoning within Gf tasks. That is, it is not clear if cognitive theories can explain performance on Gf tasks generally, at the mean group level, before individual differences are taken into account. Theories designed to explain performance on tasks from the cognitive literature may not be applicable to tasks from the differential literature. Furthermore, cognitive theories do not explicitly pinpoint processes that may be sources of individual differences. Thus, it may be necessary to make significant modifications/extensions to these theories for them to be able to explain mean performance on Gf tasks and individual differences in performance on Gf tasks.

Chapter 2 of the thesis will be a review of the literature (theories, findings and issues) relevant to the study of Gf and its processes. This will include an outline of the differential approach to the study of intelligence (which gave rise to the Gf construct) and the cognitive-differential approach (which arose out of the aim to uncover the cognitive processes involved in ability constructs such as Gf). The reasons for why experimental manipulations and broad, comprehensive cognitive theories are needed to make the link between processes (from cognitive psychology) and Gf comprehensible will be discussed.

In Chapter 3, three broad, comprehensive cognitive theories of induction will be outlined, each with different degrees of emphasis on knowledge and learning. The implications of
these theories for the processes that may be involved in Gf tasks will be discussed. Our
general hypotheses for the thesis (about how learning may be involved Gf tasks) will be
outlined in detail.

Chapter 4 will examine whether learning takes place in Raven’s Advanced Progressive
Matrices (a ubiquitous Gf task) and if so, to what extent learning is a source of individual
differences (Experiment 1). Specifically, we will examine whether rule learning within the
task is necessary for abstraction to take place, and whether those of higher Gf ability learn
faster than those of lower Gf ability. Implications for competing accounts of induction will be
discussed. That is, discussion will revolve around how well the classical conceptualisation of
Gf, knowledge-based theories of induction, and learning-based theories of induction are able
to explain the results at the level of group differences and at the level of individual
differences.

Chapter 5 will consist of one pilot study (Experiment 2) and Chapters 6 and 7 will consist of
two larger studies (Experiments 3 and 4, respectively). These studies examine a number of
issues including those that are raised by the results of Experiment 1, Chapter 4. These
chapters more explicitly examine the distinction and similarities between the learning that
may occur across items within a task, knowledge that may be brought to a task and reasoning
within an item that may be independent of any outside knowledge or across-item learning.
These chapters also examine the relationship between learning and item-order effects and
proactive interference (considered a consequence of learning). The studies examine which of
these are relevant to performance and individual differences in the Number Series and Letter
Series tasks (ubiquitous Gf tasks) and a modified Gf-like task from Sweller and Gee (1978),
which has characteristics ideal for exploring the questions of interest – the “Modified Sweller
and Gee” (MSG) Task.
Lastly, Chapter 8 integrates the main issues addressed in each of the other chapters. Particular emphasis will be given to the implications of the empirical findings for how Gf should be conceptualised in relation to induction and the involvement of learning, use of knowledge and the combating of proactive interference.
CHAPTER 2
INTELLIGENCE & FLUID INTELLIGENCE

2.1 Introduction

Gf arose out of research aimed at understanding human intelligence. Since then, much research effort has been devoted to understanding Gf itself and what causes people to differ on Gf tasks. This chapter outlines the theoretical context in which Gf and the studies of its processes developed.

The differential research approach (from which the concept of Gf developed) emerged from early aspirations to more systematically understand the organization (structure) of abilities that may be associated with intelligent functioning. However, the differential approach was unable to identify the reasons why people differ on Gf tasks and this led to the cognitive-differential approach. This chapter outlines these two approaches, including what they reveal about Gf and what we still do not know. It starts with a discussion of the shortcomings of conceptualizations of intelligence which are based on impressions and assumptions about the nature of intelligence. The chapter concludes with a discussion of the reasons as to why more comprehensive theorizing and theory testing is needed before processes identified in the cognitive literature can be meaningfully linked to abilities from the differential literature.

2.2 The Definition of Intelligence – or Lack Thereof

“Looked at in one way, everyone knows what intelligence is; looked at another way, no one does...people all have conceptions [of intelligence]...but no one knows for certain what it actually is.” – Sternberg (Sternberg, 2000).
Many debates and confusion in intelligence research arise as a result of the lack of agreement over how intelligence should be defined and conceptualised. Hence, a discussion of this issue and any implicit assumptions about what is meant when the term “intelligence” is used is important for any study that is about intelligence or intelligence-related concepts.

Studies of lay conceptions of intelligence (for example see, Sternberg, 2000), where participants are asked what they consider to be “intelligence”, have revealed that what people understand intelligence to mean varies from person to person and from culture to culture, and even within cultures. Where commonalities exists, it is in terms of an emphasis on reasoning and problem solving (Sternberg, 2000).

There is equal disagreement amongst researchers of intelligence. Various vague definitions have been put forward (see Sternberg & Berg, 1986 for a collection of examples). Adaption to the environment, basic mental processes and higher order thinking such as the ability to reason, think abstractly, learn, problem solve and make decisions, are often mentioned as being attributes of intelligence. That is, dominant attempts at definitions of intelligence tend to be attempts at specifying what attributes are necessary for a person to be considered intelligent (Neisser et al., 1996). Intelligent people are considered to possess the attributes of being able to reason, think abstractly, learn etc.

A criticism of this approach of specifying attributes is put forward by Schlinger (2003) who believes such attempts at defining intelligence results in erroneous, circular explanations of intelligent behaviour. An example of such circular reasoning is as follows. 1) A person displays a behaviour society regards as intelligent – such as successfully solving a reasoning problem. 2) The explanation for the intelligent behaviour is that the person possesses intelligence. However, this type of explanation simply takes the name or label given to a
behaviour (i.e., “intelligent” behaviour) and converts it into the explanation of that very same behaviour (i.e., intelligence was the reason for that intelligent behaviour – the act of successfully solving a reasoning problem). According to Schlinger (2003), this type of conceptualization of intelligence is a reification - intelligence is incorrectly assumed to be an attribute or essence possessed by a person that determines their behaviour; and when intelligence as an essence or attribute is assumed, researchers then feel that a definition of the essence or quality can be formulated. The end result is that the definitions formulated are just based on the behaviour observed in the first place. For example the definition of intelligence as “capacity to learn” is often formulated by teachers trying to explain why some are better at learning than others. This is circular. While this method of formulating definitions is practiced by both lay people and researchers alike (see Sternberg, 2000; Sternberg & Berg, 1986 for examples) it does not get us very far.

Instead, Cattell (1987) and Schlinger (2003) advocate that researchers must discover the definition through experimentally analysing behaviour and looking for regularities, order and structure. Indeed, understanding behaviour in the scientific sense means being able to specify the historical and contemporary conditions or variables necessary for its occurrence (i.e., cause and effect relations). This may be what is needed to get at the essence of intelligence, to formulate a real definition that goes beyond circular reasoning. If so, formulating a definition of intelligence would involve an iterative process of the formulation of theories and empirical study. This may be true because so far, a satisfactory (comprehensive) definition of intelligence still seems to be “a work in progress”.

The next sections outline the two major areas of work that have been done in the field of intelligence research – “differential” research and “cognitive-differential” research. Historically, the main approach to the study of intelligence has been the differential approach.
It is the approach from which Gf originally emerged, but it has certain limitations. While it has been able to identify and categorize ways in which people differ in performance on various Gf tasks (i.e., to identify abilities), it has not been as successful at identifying the causes as to why people differ in these abilities. The cognitive-differential approach arose as an attempt to address this limitation. These two approaches will be discussed in more detail in the next sections.

2.3 The Differential Approach – The Origins of Gf

“We know how to measure something called intelligence, but we do not know what has been measured” – Brody (2000).

The differential approach is also known as the “individual differences” approach and the “psychometric” approach. It is often called the psychometric approach because it is based on the idea that basic mental abilities can be discovered and classified through psychological tests and psychometric analysis. It is often called the differential or individual differences approach because the (implicit) assumption is that between-individual variation in task performance is caused by individual variation in one or more underlying abilities (Borsboom, Mellenbergh, & van Heerden, 2004) and if the selection of tasks is conducted in an appropriate way, abilities can be identified by factors in a factor analysis.

The differential approach began with Charles Spearman. Spearman (1932) examined a large amount of data from various studies that used mental abilities tasks and found that almost all correlations were positive. Using factor analysis he found that a general factor (which he called “g”) could be extracted from these tasks. However, tasks were not perfectly correlated and tasks that were similar to each other (in some way) correlated more highly with each other. Hence, Spearman concluded that task performance was determined by a g factor and
factors specific to those tasks (which he called specific factors “s”). The g factor has been interpreted by many as being statistical evidence for the existence of a “general intelligence” (for examples see Gottfredson, 1997; Jensen, 1998). That is, it has been assumed by many researchers that it is intelligence that causes the positive correlations between mental abilities tasks.

Spearman’s (1932) g was challenged when Thurstone (1938; Thurstone, 1947) developed a method of factor analysis different to Spearman’s method. Thurstone argued that factors should be rotated to “simple structure” which resulted in a factor structure in which tasks loaded highly on a single factor and had near zero loadings on other factors. Thurstone’s analyses led him to conclude that there were several primary ability factors, including verbal comprehension, number facility, spatial reasoning, memory, deduction, and inductive abilities (Brody, 2000), rather than just a single general factor g. However, the difference in Spearman’s and Thurstone’s results was largely due the use of different methods of factor analysis; because mathematically, there is very little difference in the amount of variability captured by the two approaches.

According to Brody (2000), correlation matrices of ability measures can usually be interpreted as containing a g factor that accounts for about 50% of the covariance in the matrix. Clearly, g does not cover all the covariance in the correlation matrix. Narrow ability measures must be postulated to better account for relationships in the matrix.

The psychometric models currently most commonly endorsed have a hierarchical structure (Davidson & Downing, 2000) that can reconcile Spearman’s (1932) and Thurstone’s (1938; Thurstone, 1947) models. This type of structure places one or more general factors at the top of the hierarchy and delegates specific factors to lower levels. That is, when first order factors
are observed to correlate with each other, one or more second order factors are introduced to explain the intercorrelations. Third order factors might also be introduced to account for intercorrelations between second order factors. These are even more general than the second order factors and appear above them in a hierarchical model. The higher the factor on a hierarchical model, the further removed it is from people’s actual performance on psychometric tasks (Davidson & Downing, 2000). It is important to note that based on statistical criteria alone, the different models are almost indistinguishable. Thus, preference for the number of levels in the hierarchy is based on theoretical stance.

Currently, the hierarchical Cattell-Horn-Carroll (CHC) theory (Flanagan, McGrew, & Ortiz, 2000; McGrew, Flanagan, Keith, & Vanderwood, 1997) is the most supported psychometric theory of the structure of cognitive abilities (Alfonso, Flanagan, & Radwan, 2005). It builds upon the integrated empirical research and theorising of Raymond Cattell, John Horn and John Carroll (Neisser et al., 1996; Alfonso et al., 2005) and includes evidence from developmental, neurocognitive, and outcome criterion studies (Alfonso et al., 2005; Horn & Blankson, 2005; Neisser et al., 1996).

Cattell (1963; Cattell, 1987) put forward a theory of cognitive abilities that was based on two main factors, fluid intelligence (Gf) and crystallized intelligence (Gc). In this Gf-Gc theory, Gf was conceptualised as a factor that represents what is common to reasoning tasks and tasks that require adaption to new situations. Gf was considered to be influenced by incidental learning and biological factors. Gc was conceptualised as a factor that contributes to performance on tasks that require acquired knowledge and is largely influenced by acculturation. Horn (1968; Horn & Blankson, 2005) and Carroll (1993) independently added to Cattell’s model other factors that they thought should be considered distinct from Gf and Gc. Although the factors added by each were not identical, they were very similar. The main
difference between Horn and Carroll’s models was that Carroll’s model contained a higher order $g$ factor whereas Horn (and Cattell’s) model did not. Descriptions and comparisons of these factors and models can be found in (Alfonso et al., 2005).

In an attempt to resolve the differences between the models, McGrew (2005) proposed an integrated Gf-Gc theory which is now known as the Cattell-Horn-Carroll (CHC) theory. CHC theory consists of seventy narrow abilities (first order factors), ten broad cognitive abilities (second order factors) and no general ability factor.

The results from the differential approach are considered first approximations to the description and organization of human cognitive abilities (Horn & Blankson, 2005). After various empirical research studies and model refinement by multiple researchers, the ten broad factors and narrow abilities from CHC theory (listed in Table 2.1) are currently the most widely used description and organization of human cognitive abilities (Alfonso et al., 2005). It is subject to further empirical research and model refinement, and is thus unlikely to be the final list. An issue of contention is the omission of $g$ in the CHC theory (Alfonso et al., 2005). The importance and necessity of $g$ has been a constant source of disagreement amongst differential researchers. Nevertheless, researchers largely agree on the existence of the ten broad cognitive abilities, particularly Gf.
Table 2.1  
Broad Factors and their Narrow Abilities, from the Cattell-Horn-Carroll (CHC) theory of cognitive abilities. From Alfonso et al. (2005).

<table>
<thead>
<tr>
<th>Broad Factors</th>
<th>Narrow Abilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fluid Intelligence (Gf)</td>
<td>Induction</td>
</tr>
<tr>
<td></td>
<td>General Sequential (deductive) Reasoning</td>
</tr>
<tr>
<td></td>
<td>Quantitative Reasoning</td>
</tr>
<tr>
<td></td>
<td>Piagetian Reasoning</td>
</tr>
<tr>
<td></td>
<td>Speed of Reasoning</td>
</tr>
<tr>
<td>Quantitative Knowledge (Gq)</td>
<td>Mathematical Knowledge</td>
</tr>
<tr>
<td></td>
<td>Mathematical Achievement</td>
</tr>
<tr>
<td>Crystallized Intelligence (Gc)</td>
<td>Language Development</td>
</tr>
<tr>
<td></td>
<td>Lexical Knowledge</td>
</tr>
<tr>
<td></td>
<td>Listening Ability</td>
</tr>
<tr>
<td></td>
<td>General Information</td>
</tr>
<tr>
<td></td>
<td>Information about Culture</td>
</tr>
<tr>
<td></td>
<td>General Science Information</td>
</tr>
<tr>
<td></td>
<td>Geography Achievement</td>
</tr>
<tr>
<td></td>
<td>Communication Ability</td>
</tr>
<tr>
<td></td>
<td>Oral Production and Fluency</td>
</tr>
<tr>
<td></td>
<td>Grammatical Sensitivity</td>
</tr>
<tr>
<td></td>
<td>Foreign Language Proficiency</td>
</tr>
<tr>
<td></td>
<td>Foreign Language Aptitude</td>
</tr>
<tr>
<td>Reading &amp; Writing (Grw)</td>
<td>Reading Decoding</td>
</tr>
<tr>
<td></td>
<td>Reading Comprehension</td>
</tr>
<tr>
<td></td>
<td>Verbal Language Comprehension</td>
</tr>
<tr>
<td></td>
<td>Cloze Ability</td>
</tr>
<tr>
<td></td>
<td>Spelling Ability</td>
</tr>
<tr>
<td></td>
<td>Writing Ability</td>
</tr>
<tr>
<td></td>
<td>English Usage Knowledge</td>
</tr>
<tr>
<td></td>
<td>Reading Speed</td>
</tr>
<tr>
<td>Short-Term Memory (Gsm)</td>
<td>Memory Span</td>
</tr>
<tr>
<td></td>
<td>Working Memory</td>
</tr>
<tr>
<td></td>
<td>Learning Abilities</td>
</tr>
<tr>
<td>Visual Processing (Gv)</td>
<td>Visualization</td>
</tr>
<tr>
<td></td>
<td>Spatial Relations</td>
</tr>
<tr>
<td></td>
<td>Closure Speed</td>
</tr>
</tbody>
</table>
· Flexibility of Closure
· Visual Memory
· Spatial Scanning
· Serial Perceptual Integration
· Length Estimation
· Perceptual Illusions
· Perceptual Alternations
· Imagery

· Auditory Processing (Ga)
  · Phonetic Coding
  · Speech Sound Discrimination
  · Resistance to Auditory Stimulus Distortion
  · Memory for Sound Patterns
  · General Sound Discrimination
  · Temporal Tracking
  · Musical Discrimination and Judgment
  · Maintaining and Judging Rhythm
  · Sound-Intensity/Duration Discrimination
  · Sound-Frequency Discrimination
  · Hearing and Speech Threshold factors
  · Absolute Pitch
  · Sound Localization

· Long-Term Storage & Retrieval (Glr)
  · Associative Memory
  · Meaningful Memory
  · Free Recall Memory
  · Ideational Fluency
  · Associational Fluency
  · Expressional Fluency
  · Naming Facility
  · Word Fluency
  · Figural Fluency
  · Figural Flexibility
  · Sensitivity to Problems
  · Originality/Creativity
  · Learning Abilities

· Processing Speed (Gs)
  · Perceptual Speed
  · Rate-of-Test-Taking
  · Number Facility

· Decision Speed/Reaction Time (Gt)
  · Simple Reaction Time
  · Choice Reaction Time
  · Semantic Processing Speed
2.3.1 Criticisms of the Differential Approach

A criticism of the differential approach is as follows: factors such as Gf are used as explanations for people’s performances differing in consistent ways on certain tasks, yet the very same task performances are used to calculate correlations that constitute evidence for the existence of the factor (Schlinger, 2003). This is similar to the circular-reasoning criticism of definitions of intelligence mentioned at the start of this chapter.

The problem is that constructs such as Gf can be reified and become empty, circular explanations. Explanations for performance on Gf tasks often involve the assumption that people vary in the amount of Gf ability that they possess. Those who are better at Gf tasks are said to possess more Gf ability but typically, the only evidence that they possess more Gf ability is that they are better at Gf tasks. In contrast, a proper explanation would involve some statement that would clarify why people differed in performance, beyond referring to a construct whose existence is based on the very same observed differences (Borsboom, Mellenbergh et al., 2003). While it should be acknowledged that the differential approach is able to describe and categorize ways in which people differ in mental performance (it identifies abilities through factors), it does not identify the causes for the differences between people (Borsboom et al., 2003; Horn & Blankson, 2005). A proper explanation of why people differ on the tasks would require a description of the processes involved in performing the tasks (which load on the factors) and the ways in which these processes differ in different people (Borsboom et al., 2003; Floyd, 2005; Schlinger, 2003).
2.4 Going beyond Circularity and Reification

At this point, it is worth explicating the definitions of "abilities", "factors" and "processes", and how they differ. Going beyond circularity and reification requires knowing the difference between abilities, factors and processes.

*Ability:* Carroll (1993) defines ability in terms of a potential for performance that varies from person to person. The existence of an ability is inferred from a collection of behaviours (e.g., test item responses) on which individuals vary systematically in efficiency and accuracy. An individual’s ability level is then conceptualised in relative terms as a comparison between his or her performance with that of an appropriate “standardization” group (Floyd, 2005). The discovery of an ability only tells us that people differ on a group of behaviours, not why they differ.

*Factor:* A factor is a latent trait derived via factor analysis which reflects systematic covariation in a group of individuals’ performances across multiple tasks of a defined class (e.g., Gf tasks). It is argued that factor analysis distils out the systematic variation common to a group of tasks from the systematic variation that is unique to each individual task, therefore producing a “purer assessment” of the ability. That is, factor analysis more accurately pinpoints where people differ – not why they differ (Carroll, 1993).

*Process:* Processes lie at a lower, more fundamental conceptual level than abilities. Carroll (1993) defines a process as any action or series of actions where something is operated upon to produce some result. Thus, a cognitive process is one in which mental representations are operated upon to produce either some new representations or a response (such as behaviour). Cognitive processes may be viewed as hypothetical constructs since they are unseen and we can only infer their existence from responses on tasks (Floyd, 2005).
The outcome of processing is what is captured in measures of abilities. That is, cognitive abilities represent individual differences in the complete series of cognitive processes that have been instantiated to arrive at a response.

Various researchers have commented that a large proportion of task-development research is characterized by an absence of theories of processes, instead focusing on convergent and divergent (correlational) validation strategies (D. P. Birney & Bowman, 2009; Borsboom et al., 2004; Schlinger, 2003). However, if one is to avoid the potential reification criticism of the psychometric approach, a greater understanding of the cognitive processes that are reflected in ability measurements is needed (Borsboom et al., 2004; Floyd, 2005; Schlinger, 2003; Spearman, 1932). Understanding the reasons as to why people differ on an ability level requires an understanding of the processes involved in that ability. It is not enough to simply know that the ability exists or that it is more or less similar to other abilities (which is the information that convergent and divergent correlational strategies provides).

The following section outlines the cognitive-differential approach to the study of intelligence. This is the approach that in various ways incorporates theories about cognitive processes.

### 2.5 Cognitive-Differential Approach

The cognitive-differential approach takes an ability from the differential approach and describes individual differences in that ability in terms of differences in information processing capacities or strategies. It borrows heavily from theorizing done in the cognitive science literature and is based on two assumptions (Mayer, 1992). Firstly, it assumes that there exists a cognitive architecture composed of a system of constructs such as long term memory, working memory, short term memory, and processes for acting on information represented in these subsystems. Together, they are hypothesized to be the building blocks of
intelligent behaviour. Secondly, it assumes individual differences - that people may differ with respect to the amount of capacity or efficiency of these subsystems and that these differences are the basis for differences in abilities.

The cognitive-differential approach has resulted in a varied set of studies ranging from research that is predominately psychometric focused on the one hand, to research which is mostly experimentally focused on the other (see Lohman, 2000; Mackintosh, 1998 for comprehensive reviews). One feature that the different studies seem to have in common is that they apply theories and methods from the cognitive psychology literature (particularly the information processing branch) to tasks which have been modelled as Gf-loaded psychometric tasks.

Experimentally-focused cognitive-differential studies have largely been concerned with detailed study of specific tasks in much the same way that cognitive psychology has been. Although many different tasks have been studied, the most common studies investigate tasks that contain analogy problems, matrix problems, series completion, and classification problems, with the focus of investigation being on one task at a time (for examples see Arendasy & Sommer, 2005; Carpenter et al., 1990; DeShon, Chan, & Weissbein, 1995; Embretson, 1998; Mulholland, Pellegrino, & Glaser, 1980; Primi, 2002; Quereshi, 2001; Richardson, 1991; Simon & Kotovsky, 1963; Unsworth & Engle, 2005b; Verguts, De Boeck, & Maris, 2000; Vigneau, Caissie, & Bors, 2006). Some common methodologies include computer simulation of problem solving behaviours, analysis of task components, examination of strategies and strategy shifting, examination of response errors and task difficulty, and retrospection and think-aloud reports. More than one of these methods are often used in a single study and often to examine changes in behaviour/performance due to experimental manipulations of task requirements. These changes in behaviour/performance
are used to make inferences about the nature of the cognitive processes involved. Sometimes changes in the task and behaviour/performance are also compared with changes in the task’s correlation with the Gf factor or marker task. Increases in correlations suggest that the inferred processes are more likely to be determinants of Gf. In some studies, individual differences are explored though comparison of high and low ability groups, either statistically or through experimental groups.

The advantage of cognitive-differential approach is that it examines individual differences and cognitive processes simultaneously. However, its shortcoming is that it is not much more informative than the differential approach on its own, when solely correlational/psychometric methods are used. Such methods merely discover the link between two constructs – albeit one is from the cognitive literature (such as WMC) and one is from the differential literature (such as Gf). The result is that our understanding of how the “process” (such as WMC) influences the ability remains limited (D. P. Birney & Bowman, 2009). We shall elaborate upon this in the next section.

2.6 Empirical Findings and Unknowns about Gf: Reasoning, WMC and Learning

Studies from the differential and cognitive-differential approach have revealed many things about the nature of Gf. Results from several differential studies show that the relationship between Gf and g is very strong. Hence, they have sometimes been regarded as equivalent for theoretical and practical purposes (Gustafsson, 1988). That is, Gf seems to have a prominent place in general intellectual functioning.

Also, there is consensus that Gf is measured in tasks that require reasoning and is related to the ability to indentify relationships, comprehend implications and draw inferences (Horn &
Blankson, 2005; Lohman, 2001; Saladin, 2007). This consensus is based on inferences about the nature of tasks that load on Gf. Measures that clearly statistically define Gf seem to require (in one way or another), reasoning. In the CHC hierarchical model (see Table 2.1), Gf sits above induction, general sequential (deductive) reasoning, quantitative reasoning, Piagetian reasoning and speed of reasoning (Flanagan et al., 2000; McGrew et al., 1997).

Gf seems to be particularly prominent in (what at least appears to be) novel, context-free reasoning tasks. Gf is conventionally characterized by, “The use of deliberate and controlled mental operations to solve novel, “on-the-spot problems” (McGrew, 2005 p.151, italics added). According to Cattell (1971) Gf taps “the level of complexity of relationships which an individual can perceive and act upon when he doesn’t have recourse to answers to such complex issues already stored in memory” (p.99). Indeed, tasks used to measure Gf are often abstract tasks that require participants to induce seemingly novel rules that govern the relationship between abstract stimuli (such as geometric shapes, numerals or letters).

Also, Gf tasks tend to contain complex items with multiple rules, steps or numbers of stimuli that have to be attended to simultaneously. As the complexity of the task increases so does its correlation with Gf (Gustafsson, 1988). For this reason (the need to maintain multiple rules, steps and stimuli), many researchers from the cognitive-differential approach have linked Gf to Working Memory Capacity (WMC) - a construct originally investigated by cognitive researchers. While there is some disagreement over the definition of WMC, it has been defined as the simultaneous storage and processing of information, and the controlled sustaining of attention in the face of interference and distraction (Engle, Kane, & Tuholski, 1999; Miyake & Shah, 1999).
Kyllonen and Christal (1990) found that Gf tasks correlated very highly with WMC tasks. They looked at the factor structure of a group of traditional psychometric ability tasks together with a varied collection of WMC measures developed from a general cognitive architecture theory. They found that the ability tasks tended to load on one factor and the WMC tasks on another, but these factors were very highly correlated with each other. Since this finding, more psychometric evidence has emerged to support the claim that WMC and Gf are highly related (Buehner, Krumm, & Pick, 2005; Kane, Hambrick, & Conway, 2005; Oberauer, Schulze, Wilhelm, & Suss, 2005).

Studies that show that WMC and Gf are highly related have been informative because the typical operationalisation of WMC and Gf are not very similar. WMC tasks do not involve higher level reasoning (if they involve reasoning then they should not be considered WMC tasks) (Lohman, 2000). In contrast, the quintessential element of Gf tasks is that they involve reasoning (Halford, Bain, Mayberry, & Andrews, 1998; Horn & Blankson, 2005; Sternberg, 1986). Yet, despite this difference, WMC and Gf tasks are highly correlated.

While these studies have been informative (that is, they reveal that there is a relationship between WMC and Gf), they do not answer why there is a relationship. It is however possible to speculate what the nature of the relationship might be. Common explanations for the WMC-Gf link focus on within-item processes. For example, Gf items often contain large numbers of steps, stimuli and sub-results that must be stored and processed simultaneously, and those with larger WMC may be more successful at this (Carpenter et al., 1990; Embretson, 1995; Mackintosh, 1998; Primi, 2002). That is, WMC constraints may limit one’s performance on individual items and thus contribute to individual differences in overall scores on Gf tasks.
However, some researchers (Salthouse & Pink, 2008; Unsworth & Engle, 2005a; Verguts & De Boeck, 2002b) have found that the relationship between WMC tasks and Gf tasks are fairly constant, regardless of the amount of information that exist within the Gf items. Hence, Salthouse and Pink (2008) concluded that the relationship between Gf and WMC could not be attributed to individuals performing better on Gf items being capable of processing and storing more within-item information than individuals who perform poorer on Gf items.

Verguts and De Boeck (2002b) conducted a study where the amount of information was kept low within each item but subjects still had to remember multiple rules over the task. They found that despite the low working memory load within items, overall performance still correlated with a WMC measure. Furthermore, items with similar rules presented consecutively were easier than items where dissimilar rules were presented in alternating sequence (Verguts & De Boeck, 2002a). Based on these findings, they argued that rules become “primed” and the amount of priming is a factor of individual differences related to WMC. They concluded that WMC contributes to individual differences in performance over items and that across-item learning may be the link between WMC and performance.

In another study, Carlstedt, Gustafsson and Ullstadius (2000) presented Gf tasks in two different sequences: one where the same kind of items were presented together (traditional homogenous sequence) and one where items from different tasks were presented alternately (heterogeneous sequence). Carlstedt et al. (2000) had predicted that the heterogeneous presentation would be the better measure of Gf because they speculated that the switching between tasks would increase the complexity of the battery, compared to the homogenous presentation. To their surprise, the homogeneous sequence was the better measure of Gf. Carlstedt et al. suggested that the homogenous sequence may have provided an opportunity to learn from the earlier items, which may have then made solving later items easier and that
higher Gf subjects benefited more from this learning opportunity than lower Gf subjects. Hence, regardless of whether the link between WM and Gf may be due to learning, learning seems to be an important component of Gf.

In contrast to the learning hypothesis, Unsworth and Engle (2005a) argue that what is important to performance in Gf tasks is the ability to control attention, especially under conditions of distraction and interference. Hence, they believe Carlstedt et al.’s (2000) findings were due to the homogenous tasks creating a condition where there was more proactive interference (PI) from previous solutions. That is, in the homogenous condition, participants had to control their attention and try to block out or inhibit (irrelevant) solutions from previous items, and it is was for this reason that the homogenous items were a better measure of Gf.

In summary, what we know so far about Gf is that measures which clearly define Gf seem to require (in one way or another), reasoning, particularly reasoning in novel situations. Also, Gf tasks tend to contain complex items (with multiple rules, steps or stimuli) and as the complexity of tasks increase, so do their correlations with Gf. Furthermore, Gf shows consistently strong relationships to WMC but the reason for this is unclear. Three explanations have been put forward for this link. The most common explanation is people with higher WMC can keep in memory many elements and therefore, would be better at storing sub-results within an item, needed for solving a single item. Another is that people with higher WMC can store many solution principles across items and use them to solve harder items; that is, they are better reasoners because they are better learners across items, within a single task. The third explanation is those with higher WMC are better able to combat proactive interference from solutions to older items and hence, are able to concentrate
better on current items. Whilst these explanations may not be mutually exclusive, the connection between WMC and Gf is still unclear and warrants further investigation.

2.6.1 A Neglected Process: Learning

The learning hypothesis (as mentioned in the previous section) is an interesting one. Intuitively, learning seems to have a strong relationship with intelligence. Originally, intelligence tasks (in the western world) were designed to identify children who were unable to profit from experience (Ferretti & Butterfield, 1992). When Raymond Cattell (1963; 1987) first put forward his Gf-Gc theory, Gf was conceptualised as a factor that loads on reasoning tasks and tasks that require adaptation to new situations and influenced by incidental learning. That is, in past theorising, learning had strong ties with intelligence and Gf.

However, in mainstream differential research, performance on Gf tasks is usually assumed to be uninfluenced by previous acquisitions of knowledge structures and learning (Gustafsson, 1988; Richardson, 1991). The emphasis of Gf’s definition has been the ability to reason and solve novel problems that do not rely extensively on an explicit base of knowledge such as that which might be developed from schooling or previous experience (Carpenter, Just, & Shell, 1990). Indeed, tasks used to measure Gf are often abstract tasks that require participants to induce novel rules that govern the relationship between abstract stimuli (such as geometric shapes, numerals or letters). In CHC theory, “Learning Abilities” are not listed under Gf, but are listed as narrower abilities under Short-Term Memory (Gsm) and Long-Term Storage and Retrieval (Glr) (Alfonso et al., 2005) (see Table 2.1), and are considered poorly defined by existing research (McGrew, 2005).

Learning ability is rarely investigated explicitly in the individual differences literature on intelligence or Gf (Carlstedt et al., 2000). One reason for this may be the overly restrictive
and theoretically arbitrary assumption of many theories of intelligence, that cognitive abilities are stable and mostly immutable (Birney & Sternberg, 2006). Learning and change do not fit easily into such conceptualisations. Another is that it has been commonly assumed that conventional Gf tasks do not allow for feedback nor opportunity to learn and hence, would not involve any dynamic processes such as learning (Sternberg et al., 2002). According to Carroll (1993), it has often been proposed that an important aspect of intelligence is the ability to learn but methodological issues and insufficient data has made it difficult to convincingly demonstrate this relation.

The possibility that learning processes are involved in Gf will be considered in more detail in Chapter 3. While there might be little explicit consideration of learning (and contrary to common assumptions regarding the conceptualisation of Gf), solution of Gf tasks may actually require learning processes. Not only may learning be a required process in Gf measures, it may be a source of individual differences (Carlstedt et al., 2000); and if what researchers such as Verguts and De Boeck (2002b) are saying is correct, then learning may even be the link between WMC and Gf. However, before empirical investigations concerning how learning processes may be involved in Gf can be coherently and efficiently conducted, comprehensive process theories are needed to guide our investigations.

2.7 The Difficult Task of Linking Ability to Processes using Theory

As mentioned previously, after the link between Gf and WMC was established a number of studies emerged aimed at understanding why there is connection between Gf and WMC. Looking at the bigger picture, such studies are more fundamentally aiming at connecting ability to processes. They are trying to identify the processes involved in Gf and sources of individual differences that contribute to differences in performance on Gf tasks. However, linking ability constructs (such as Gf) to cognitive processes (such as WMC) requires
substantial theories (Deary, 2001) which then need to be tested. Since the reason for the strong relationship between Gf and WMC is not clear, this suggests that there is a gap in theorizing and theory testing in the literature.

A unified cognitive-differential approach has long been advocated as a necessary path for the advancement of our understanding of constructs such as Gf. Many researchers have recommended that differential psychology take note of process theories from cognitive psychology (Cronbach, 1957; Deary, 2001; Lohman, 2000). Traditionally, studies of Gf in the differential literature have been more pre-occupied with its factor structure than with processes that may be involved in solving Gf items. Descriptions of Gf in the differential literature developed largely as a result of visual examinations of items that load on Gf and the introspection of researchers (for an example see Cattell, 1987). Meanwhile, in the cognitive literature, there have been studies that formulate theories and examine cognitive processes involved in reasoning tasks. Some of these works examine tasks that are similar to (and sometimes the same as) those found in the Gf literature. Hence, it would be sensible to assume that insights from the cognitive literature may be helpful in telling us more about processes involved in Gf.

However, combining the two approaches is not without some complications. Cognitive psychologists have not been concerned with why and in what ways people differ in their ability to reason effectively (Mackintosh, 1998). Studies are based on theories of how people reason in general and they typically focused on isolating reasoning processes common to all. In contrast, the abilities identified by differential psychology are based on individual differences (variation in performance) Thus, theories from cognitive psychology are not easily mapped onto the Gf construct.
Another problem is that for practical reasons, the cognitive literature has mainly focused on studies of single tasks (Sternberg, 1986). The detailed investigation of processes involved in the study of single tasks can result in a very large amount of data related to the processes involved in solving *that particular task*. It is not clear how much of the data from single task studies can be generalized across tasks, persons, time and situations (Galotti, 1989; Lohman, 2000). Also, the processes that contribute to individual differences may be a small fraction of the processes involved in the solving of the whole task – and the former may be hard to untangle from the latter (Lohman, 2000). Certainly, establishing an understanding of the processes entailed in a solution is an involved and complex activity. This is in part because there may be: 1) variability in the processes different individuals use to solve the same item, 2) variability in the processes the same individual uses to solve different items in the same task, and 3) variability in the processes used by individuals of different ability (Borsboom et al., 2003; Floyd, 2005). Ultimately we should be most concerned with processes that are common to both performance on Gf tasks and everyday life tasks, since the whole point of studying ability tasks is that they are assumed to tell us something about performance in the real world (Galotti, 1989). Nevertheless, establishing a detailed understanding of processes which can be generalized across families of tasks, persons, situation, time and which represent sources of individual differences on these tasks, may be a very delicate balancing act (Lohman, 2000).

A practical starting point would be to first focus on comprehensive, general, cognitive theories which cut across Gf tasks. General theories can help us to better focus on general processes – that is, processes common to everyone in all Gf tasks as well as everyday life tasks. The next step would be to empirically test the cognitive theories to see if they apply to (can explain) performance on Gf tasks at the mean group level. The step after that would be to identify sources of individual differences (if any), associated with those processes.
Cognitive theories are usually not explicitly concerned with processes that may be sources of individual differences. Thus, hypotheses should be formulated about which processes mentioned in the cognitive theories might actually contribute to individual differences in performance on Gf tasks. These hypotheses should then be empirically tested (using an experimental-differential approach) to see if the processes in question actually differentiate those of high and low Gf abilities.

Since there is consensus that Gf is measured in tasks requiring reasoning, general cognitive theories of reasoning would be a good starting point for investigations into Gf. Some potential candidate theories will be introduced in Chapter 3. Particular focus will be given to theories that address learning processes because learning processes have largely been neglected in cognitive-differential research but as argued here, has been shown to have links with WMC and past theorizing regarding intellectual functioning.

2.8 Summary

Schlinger (2003) argues that the use of “intelligence” arose out of attempts to explain certain behaviours such as differences in performance on certain tasks (e.g. reasoning tasks). However, successful performance on such tasks erroneously became the definition for intelligence (e.g. ability to reason). The result was that we did not know much more about why people differed in performance on various tasks.

The differential approach to the study of intelligence led to the discovery and classification of various abilities, such as Gf. The limitation of the differential approach is that the use of abilities to explain differences in performance also does not tell us much about why people differ. Understanding why people differ on performance requires a study of cognitive
processes that contribute to the individual differences that define abilities. This has been the subject of investigation in the cognitive-differential literature.

Understanding what Gf is exactly and why people differ on Gf tasks is largely still research in progress. The differential and cognitive-differential approaches have revealed that Gf is measured in tasks that require reasoning in seemingly novel problems and has strong links with WMC. However, the reason for the link is not clear.

Three major explanations have been put forward. The most common explanation is that when solving a Gf item, people with higher WMC can keep many elements in memory and therefore, are better at storing sub-results and information from the processing of features needed within an item, to solve the item. Another is that people with higher WMC can store many solution principles over multiple items and use them to solve harder items; that is, they are better reasoners because they are better learners. The third explanation is that those with higher WMC are better able to combat proactive interference from solutions to older items and hence, are able to concentrate better on current items.

The learning hypothesis is an interesting one because despite learning having been linked with Gf in original conceptualisations of Gf, it has since become somewhat neglected in Gf and intelligence research. The possibility that learning processes are involved in Gf will be considered in more detail in Chapter 3.

Fully understanding why people differ in performance on Gf tasks requires an understanding of the cognitive processes involved in such tasks but making the link between processes and abilities requires substantial theory and theory testing. Since the reason for the link between Gf and WMC is not clear, there must be a gap in theorizing and theory testing in the literature. The differential approach has largely been devoid of process theories while
cognitive research has developed various processes theories. However, cognitive research has largely not developed theories that contain the level of detail that can be generalized across Gf tasks and take into account the individual differences that are inherently important to ability constructs. A practical starting point to address this is to first focus on comprehensive, general, cognitive theories which cut across reasoning tasks and then to use these theories to target empirical study to identify sources of individual differences which are relevant to Gf tasks. The next chapter focuses on three such theories.
CHAPTER 3
REASONING & INDUCTION

3.1 Introduction

The overarching aim of this thesis is to gain a better understanding of Gf through investigating why people differ on Gf tasks. The cognitive reasoning literature consists of a large body of work on cognitive processes, including reasoning tasks which share features with Gf tasks. Using reasoning theories from the cognitive reasoning literature to uncover processes involved in Gf tasks seems like a sensible thing to do but unfortunately, this is not a straightforward matter. As touched upon in Chapter 2, there is a lack of unity in the cognitive literature due to the task-specific nature of research studies. This creates an obstacle because it is not clear which theories are relevant and generalizable to Gf tasks. Another obstacle is that cognitive reasoning theories are generally more concerned with how people reason in general and not with pinpointing why people differ on tasks. We will address these obstacles by outlining three broad, general theories of reasoning and then using these theories to form the basis for our hypotheses about individual differences in Gf tasks. These hypotheses will be focused on learning processes. This will be preceded by a general discussion of the nature of reasoning which will serve to highlight some reasons why only a few comprehensive, unified theories of reasoning currently exist.

3.2 What is Reasoning

Reasoning has been defined as any process of drawing a conclusion from a set of premises (Blackburn, 1996). Various conceptualizations of reasoning have evolved within the field of formal logic (Ladriere, 2003). However, the bodies of literature on reasoning from schools of
philosophy and logic are quite complex and lengthy, and beyond the scope of this thesis. We limit ourselves to theories that come more directly from the discipline psychology.

In psychology, two prominent, broad categories of reasoning have received the most attention: deductive reasoning and inductive reasoning (Bell & Staines, 2001). Deductive reasoning has been described as a way of reasoning that relates two or more general known concepts or conditions to a specific case. For example, if all birds build nests and a magpie is a bird, then a magpie will build a nest. In contrast, inductive reasoning uses a specific observation to reach a general conclusion. For example, if a child puts her hand into a bag of marbles and withdraws three pieces, all of which are red, she may conclude that all the marbles are red. There are widely accepted standards available for evaluating the quality of deductive reasoning, but evaluating induction is matter of debate (Nickerson, 2004). In deduction, a conclusion follows directly from the premises. A valid deduction is one in which it is impossible to assert the premises and deny the conclusion without contradiction. Valid inductive reasoning only requires that based on the premises, the conclusion is highly probable.

Arguably, inductive reasoning is more relevant to real life than deductive reasoning (Nickerson, 2004). Most of the reasoning problems from life are not solvable simply by a series of deductions from the information given, since they usually do not come with all the information that is essential for their solutions (Galotti, 1989; Nickerson, 2004). Typically one must make assumptions or look to known examples as a guide, which in essence is inductive reasoning (Nickerson, 2004). Tasks that have the highest g and Gf loadings are also considered to be inductive reasoning tasks (Snow, Kyllonen, & Marshalek, 1984). Hence, studying inductive Gf tasks may be a good place to start in identifying general processes common to Gf tasks as well as everyday life functioning – because ultimately, the whole
point of studying ability tasks is that they are assumed to tell us something about performance in the real world (Galotti, 1989). Later chapters in this thesis will be empirical studies of inductive Gf tasks.

3.4 Reasoning Theories in Psychology

While there is agreement on the general definition of reasoning (outlined above), it is not comprehensive enough for research purposes. In a review of the literature, Galotti (1989) argues that some of the most important issues impeding progress in reasoning research includes lack of agreement over definitions of terms. Boundaries for usage of the term “reasoning” remains unclear, which makes it very difficult for readers to know if any two investigators are studying a common entity and hard to meaningfully interpret patterns of performance across different lab reasoning tasks. Nickerson (2004) points out that in the psychology literature, reasoning has a variety of connotations that differ in their inclusiveness. Some authors do not make distinctions between reasoning and other constructs such as thinking, problem solving and decision making (Leighton, 2004). Even when theorists consider reasoning to be a particular type of thinking, its definition is rather broad. For example, Wason and Johnson-Laird (1972) see no clear boundaries surrounding the study of reasoning; whether someone is drawing conclusions from premises according to traditional Aristotelian laws of logic, solving a crossword puzzle, planning to buy a house or finding the best route of travel, they are all considered to be engaged in reasoning.

Galotti (1989) also argues that narrow definitions would not be desirable either. For instance, reasoning could be defined as “thinking according to the theorems of a logical system”, decision making could be defined as “weighting and combining probabilistic information in such a way as to rank alternatives”, and thinking could cover both of those constructs as well as other tasks in which information is used or combined. While consistent use of such
definitions would lead to clearer communication amongst researchers, Galotti argues that if reasoning were defined so narrowly, then the relationship between reasoning and everyday functioning (where rules of logic may not apply) would be obscured. Thus, it is a fine balance between developing a sufficiently broad definition of reasoning that is able to extend to ordinary life reasoning tasks and guarding against being unclear about the phenomenon in an attempt to be inclusive (Galotti, 1989).

Galotti (1989) proposes her own definition based on her summary of the literature. She attempts to be both precise and inclusive and the result is that her definition seems a little vague and tautological, as highlighted by the many instances of her use of phrases such as: “may, but need not”,

“[Reasoning is defined as] mental activity that consists of transforming given information (called the set of premises) in order to reach conclusions. This activity must be focused on at least one goal (but may be focused on more than one). The activity must not be inconsistent with systems of logic when all of the premises are fully specified, although there may not always be an applicable system of logic to govern specific instances of reasoning. The activity may or may not be self-contained; that is, people may implicitly or explicitly add to, subtract from, or otherwise modify any or all of the premises supplied. When original premises are modified, the final conclusion must be consistent with the modified premises. The activity may, but need not, involve the breaking of mental set. The conclusions may, but need not, be startling or non-obvious at the outset of the activity. The conclusion may, but need not, be deductively valid” (p.333 italics added).
In order for hypotheses about Gf processes to be formulated and tested, detailed, precise and sufficiently general conceptualizations of reasoning are needed but such conceptualizations are not common in the literature. Successfully formulating such definitions is quite a balancing act and not an easy task (as can be seen from (Galotti, 1989). Furthermore, most comprehensive reasoning theories also focus solely on deduction (for examples, see (Braine, 1990; Evans, Newstead, & Byrne, 1993; Johnson-Laird, 1983; Manktelow, 1999; Rips, 1994) which is not suitable for our purposes because our focus is on induction.

Three attempts that focus on induction that are comprehensive and general are Spearman (1932) from the differential literature, Holland, Holyoak, Nisbett and Thagard (1989) from the cognitive literature and Sternberg (1986) from the cognitive-differential literature. In these theories of reasoning, attempts are made at specifying general processes that could be common to Gf tasks and everyday real-life functioning. The theories will be outlined in turn.

3.4.1 Spearman's Cognitive Theory of “g” (Eduction of Relations and Correlates)

Spearman’s (1932) cognitive theory of \( g \) is a prominent, cognitive theory of intelligence from the differential literature; prominent, because it is one of the earliest. While it was not meant to be a theory of reasoning per se and it is about \( g \), rather than Gf, it is still important to our discussion for a number of reasons. Firstly, as mentioned in Chapter 2, Gf has strong links with \( g \). The results from several studies show that the correlational relationship between Gf and \( g \) is so strong that theoretically and for practical purposes, the two constructs could be regarded as equivalent (Gustafsson, 1988). Secondly, matrices tasks (highly Gf loaded tasks) were originally constructed to put Spearman’s theories into testable form (Eysenck, Fulker, & Eysenck, 2006). Also, like Gf which loads predominantly on tasks that require reasoning, Spearman’s \( g \) loadings were highest in complex mathematical, abstract and verbal reasoning tasks and lower in motor skills and repetitive tasks. Lastly, Cattell (1963) was building upon
Spearman’s work when he came up with his conceptualization of Gf. Hence, Spearman’s theory is an important starting point in any discussion about cognitive processes and Gf.

Spearman’s (1932) theory of $g$ came about when he examined a large amount of data from various studies that used mental abilities tasks and found that almost all correlations were positive. Using factor analysis he found that a general factor (which he called “$g$”) could be extracted from these tasks. The $g$ factor has been interpreted by many as being statistical evidence for the existence of a “general intelligence”, since intelligence may be the cause of the positive correlations between mental abilities tasks. Spearman was never happy with this explanation because it was not clear what “intelligence” actually meant.

Spearman (1932) wanted to know exactly what $g$ was and how universal it might be. He thought that in order to investigate this, one would need to study the complete system of fundamental processes of cognition. Such knowledge did not exist during his time (nor does it exist now), so he attempted to outline the processes himself (Spearman, 1932).

He came up with what he called the three laws of neogenesis which he thought were fundamental processes of cognition:

- **Law 1:** Knowing of our own experience – a person has more or less power to observe what goes on in their mind. They feel, strive and know, and *know* that they feel, strive and know.

- **Law 2:** Eduction of relations – seeing first order, direct relationships between things. When a given person has in mind any two or more items of mental content (whether perceived or thought of), they have more or less power to bring to mind any relations that essentially exist between them.
• **Law 3:** Eduction of correlates - given an item of mental content and a relation (such as one inferred from the eduction of relations), one is able to bring to mind a correlate to the mental content. That is, a mental content that is connected to the original mental content by the relation.

Spearman (1932) concluded that the laws of neogenesis (the fundamental processes of cognition) were largely about spotting relations. He further concluded that the spotting of relations is what is common to everyday functioning and ability tasks with high $g$ loadings, because $g$ was found in all tasks where you have to look for relations (even simple ones) and the spotting of relations (he argued) was crucial to everyday functioning. Some of the types of relations he identified and argued were prevalent and fundamental to everyday functioning were the relations of: evidence, likeness, conjunction, space, time, identity, attribution, causation, and constitution. Lastly, he concluded that $g$ emerged in factor analysis because of individual differences. That is, people differed in their ability to implement the three laws of neogenesis, particularly the eduction of relations and correlates.

### 3.4.2 Sternberg’s Unified Theory of Reasoning (Conceptual Knowledge)

While acknowledging that reasoning research has been focused on narrow tasks, Sternberg (1986) argues that there are unities in reasoning that transcend problem form and content which become apparent by looking across many tasks. According to Sternberg, these unities allow reasoning to be defined as, “...the *controlled* and *mediated* application of three processes - *selective encoding*, *selective comparison*, and *selective combination* - to *inferential rules*” (p. 281, italics added).

Each of the key terms in Sternberg’s (1986) definition will be explained. *Selective encoding* refers to the activity of distinguishing relevant from irrelevant information. The selected
information is then stored in working memory. *Selective comparison* refers to the activity of deciding what mentally stored information is relevant for solving a problem. Solution of most problems requires retrieval of declarative or procedural knowledge or both, from long term memory. *Selective combination* is the activity of putting together information that has been selectively encoded and compared (the information to be combined is stored in working memory). The nature of *inferential rules* in Sternberg’s theory is not very clear. Sternberg states that,

“The word *rules* is used broadly to incorporate heuristics, mental guidelines, algorithms, and the like…different types of problems require different types of rules. Rules may even change within problem type as a function of problem content, resulting in effects of problem content as well as problem type in task performance. It would not be possible to specify here the complete set of rules required for all possible kinds of reasoning problems. Indeed, it is not clear that it is even possible to do this because as new reasoning tasks are invented, new rules may become applicable” (p.288).

Sternberg (1986) seems to use “inferential rules” as an umbrella term for any processes involved in reasoning other than selective encoding, combination and comparison. However, he does give two examples of types of inferential rules which he calls procedural rules and declarative rules, which are based on procedural and declarative knowledge. Procedural rules are (primarily a function of problem format/type) and include things that one needs to do to solve the problem (e.g. inferring relations between stimuli, mapping relations between concepts, applying relations etc) and solution strategies. Declarative rules (primarily a function of problem content) specify the possible relations allowed to form the basis of a given problem. This *does not* mean that the answer to the problem must already be known, because reasoning is only involved if processing is done in a controlled, effortful way (as
opposed to automatic). However, it does mean that the reasoner must possess general, conceptual procedural and declarative knowledge of relevance to the problem. For example, in a geometric matrix problem, two figures might be related by the *addition or subtraction* of a part of a shape or a *gradual change in size* of a shape. The declarative rules in this case are the concepts of addition or subtraction and gradual change in size. One has stored mentally the possible declarative rules that serve as a basis for inductive relations and calls upon these rules in solving problems. Selective encoding, comparison and combination processes determine the procedural and declarative rules to use.

The last term in Sternberg’s (1986) definition of reasoning that needs to be clarified is *mediators*. Mediators refer to intervening variables that increase or decrease the availability or accessibility of the inferential rules for use in a particular problem. Examples from Sternberg 1986) are given in Table 3.1. While selective encoding, comparison and combination processes determine the inferential rules to use, mediators effects accessibility and availability of rules.
Table 3.1

<table>
<thead>
<tr>
<th><strong>Contextual probability</strong></th>
<th>When the probability for application of certain kinds of inferential rules is influenced by the problems that appear in a given problem set. Affects accessibility to rules.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Entrenchment</strong></td>
<td>Inferential rules are more easily applied if their use is entrenched within people’s general experience. Affects accessibility to rules.</td>
</tr>
<tr>
<td><strong>Prior knowledge</strong></td>
<td>Inferential rule may not be usable if it is unknown to the problem solver. E.g. Lexical problems are not solvable if one does not know the meaning of parts of the word. But, lack of prior knowledge does not make solving of the problem impossible, as you might be able to infer rule without knowing it in advance. Affects availability.</td>
</tr>
<tr>
<td><strong>WM load</strong></td>
<td>The difficulty of applying a set of inferential rules to a given reasoning problem is often affected by the individual's working memory capacity. Problems with more elements and intermediate steps than a person can handle will affect their ability to access relevant rules.</td>
</tr>
</tbody>
</table>

In the theory, selective encoding, selective comparison, selective combination, inferential rules and mediators are applicable to the definition of reasoning only when executed in a *controlled* fashion. The greater the degree of automatization, the smaller is the degree of reasoning considered to be involved. Anything that can be answered fast and automatically is not reasoning. Hence, a task is a reasoning task if and only if its solution involves the *mediated and controlled application of inferential rules for purposes of selective encoding, selective comparison, or selective combination*. In other words, the controlled use of inferential rules for any one of the three kinds of selective functions, as mediated by the
mediating variables described earlier, is necessary and sufficient to define a task as involving reasoning. The degree of involvement of reasoning is inversely related to the degree to which the application of the kind of process is automatized in the particular instance. Thus, the extent to which a task is a reasoning one is a function of an interaction between person and the task and depends on the extent to which selective encoding, selective comparison and combination are involved and done in controlled way.

It is not clear from Sternberg’s (1986) writing whether the three processes are applied to the rules, or whether the rules are applied to the processes. For instance, sometimes he describes reasoning as being,

“…the controlled and mediated application of three processes - selective encoding, selective comparison, and selective combination - to inferential rules” (p. 281).

At other times he describes reasoning as,

“…mediated and controlled application of inferential rules for purposes of selective encoding, selective comparison, or selective combination” (p.293).

Arguably, the rules and processes may be applied to each other iteratively.

With regard to inductive reasoning in particular (as opposed to deductive reasoning), Sternberg (1986) considers it to mainly involve selective encoding and comparison, rather than selective combination. Selective encoding is needed to distinguish relevant from irrelevant information. The selected information is then stored in working memory. Selective comparison is needed to decide what information stored in long term memory might be relevantly applied to what is stored in short term memory for the purposes of solving the problem. The solution of most inductive problems is thought to require the retrieval of declarative or procedural knowledge or both, from long term memory.
3.4.3 Holland et al.’s Theory of Induction (Learning and Conceptual Knowledge)

Holland et al.’s (1989) theory (or “framework”, as they describe it) is comprehensive and takes up several hundred pages. Only a brief description of relevant parts will be attempted here. While Holland et al.’s framework is relevant to other types of reasoning, its main focus is a description and explanation of induction. This is suitable for our purposes because Gf tasks are mainly induction tasks.

Interestingly, Holland et al.’s (1989) notion of induction is inextricably linked with learning in that they define induction as all inferential processes that expand knowledge in the face of uncertainty. That is, the reasoning process is not something that is done in isolation. It is inextricably linked to the expansion of knowledge or learning. Although this is not a unique view in the cognitive reasoning literature, it is a fairly recent view (Anderson, 1993). In fact, Holland et al. assume what is essential to induction are goals, previous knowledge and learning. This is because in their theory, the plausibility of possible predictions in inductive reasoning can only be determined with reference to the relevant goals and the knowledge that the reasoner already possesses. The reasoner is characterized as possessing a cognitive information-processing system that forms predictions based on prior knowledge that the system possesses and receives feedback about its successes/failures in attaining its goals. The feedback then determines if the knowledge should be confirmed or modified. That is, the system learns, “The study of induction, is the study of how knowledge is modified through its use” (Holland et al., 1989, p.5).

Specifically, the central hypotheses and assumptions they make about induction are:

1) Induction is directed (goal driven) by problem-solving - inductions are based on the need to solve a problem.
2) Knowledge is represented by rules.

3) Rules are used to construct mental models. Mental models are where newly acquired information (stored in working memory) and older information (from long term memory) is integrated in various ways (in working memory) to “model” (represent) the problem and to produce predictions about the solution.

4) During induction, multiple rules are triggered and compete with and complement each other to play a part in representing the problem.

5) The induction process is based on feedback regarding the success or failure of predictions generated by the system. Correct predictions strengthen rules and incorrect predictions trigger a weakening or revision of the rules. Stronger rules will tend to win the process in point 4) above.

6) Things learnt about the problem are stored in memory in rule clusters. These clusters are called “schemas” and they can be accessed as large units that serve to generate future plausible inferences and problem solutions.

7) When no solution can be found, the system tries to transfer other known rules associated with similar, better understood situations to the current problem. That is, it uses analogies.

That is, induction is seen as goal driven, hypothesis testing about the environment (or problem). Hypotheses are based on previous knowledge. When predictions fail, this becomes a problem that the system solves by modifying the incorrect knowledge and creating new rules as hypotheses. Concepts with shared properties are activated, thus providing analogies for use in problem solving and rule generation. Induction difficulties arise when there is no previous knowledge or analogies to draw from or when new rules fail to compete effectively with old, well-established intuitive rules that are entrenched.
In Holland et al.’s (1989) theory, there is no major distinction between learning as a result of induction and traditional conceptualizations of learning (such as learning from education). The only difference is that education is seen as rules directly inserted from the outside, as opposed to rules that the system has to generate through induction. These inserted rules are assumed to act very much like induced rules. That is, they will enter into competition for the right to represent the environment along with the rest of the rules in the system’s possession. Educational learning difficulties are considered to result from the inability of the new inserted rules to compete effectively with old, well-established rules.

### 3.5 Theoretical Similarities and Contrasts: Implications for General Processes and Individual Differences involved in Gf tasks

Many researchers have recommended that differential psychology refer to process theories from cognitive psychology to gain a better understanding of processes involved in abilities (Cronbach, 1957; Deary, 2001; Lohman, 2000). Tasks that clearly define Gf seem to require reasoning (in one way or another), particularly inductive reasoning. Hence, theories of inductive reasoning such as those of Sternberg (1986) and Holland et al. (1989) may provide potentially plausible accounts of Gf because they are task-general theories. Spearman’s (1932) cognitive theory of g from the differential literature is also a task-general one. However, some obstacles to a merger persists. For instance, each of these theories emphasize different processes and with the exception of Spearman, they are not very explicit about processes that contribute to individual differences.

The theories of Spearman (1932), Sternberg (1986) and Holland et al. (1989) have different emphasis on what induction involves. Sternberg’s theory emphasizes the use of prior conceptual knowledge, while Holland et al.’s conceptualization includes knowledge, but emphasizes learning. Spearman’s theory emphasizes the eduction of relations and correlates,
which seem to occur in isolation from previous knowledge and learning. Only Spearman’s original conceptualization of $g$ can be viewed as being consistent with the dominant, novelty view of Gf because in his theory, those of higher ability will be better able to elude relations and correlates, even in novel situations. In contrast, Sternberg’s theory emphasises the controlled application of knowledge-based inferential rules for the purposes of selective encoding (of new information) and comparison (with information stored in long term memory). That is, reasoning is guided by prior knowledge. The theory does not say that the reasoner must have knowledge of the answer, since this would not be considered reasoning. Still, this could be seen as being inconsistent with the novelty view of Gf because some prior knowledge must be brought to the problem\(^1\). Holland et al.’s theory is consistent with Sternberg’s theory and inconsistent with the novelty view of Gf, because it is also a knowledge-based theory of reasoning, where induction is seen as being guided by prior conceptual knowledge. It differs to Sternberg’s, in that it also includes learning processes as part of the inductive act. Holland et al.’s theory states that induction is a process of learning and applying knowledge-based inferential rules.

The emphasis on knowledge in the theories of Sternberg (1986) and Holland et al. (1989) may lead one to question whether their theories may be applicable to general reasoning processes (that is, uniform processes which are applicable to all) involved in Gf tasks. Furthermore, with the exception of Spearman (1932), the cognitive theories under discussion

\(^1\) Sternberg does try to placate the novelty view to some extent. His theory states that it is not essential to know the rule to start with, since lack of prior knowledge does not make solving of the problem impossible. You might be able to infer the rule, without knowing it in advance. It is not clear what he means by “infer” in this latter sense (when knowledge is not available) because reasoning in his theory relies on the use of declarative and procedural rules (that is, knowledge). He does state that there are other types of inferential rules, but does not state what they might be. Hence, from his theory, it is not clear what reasoning without knowledge would be like.
are not very explicit about processes that contribute to individual differences in Gf task performance. According to Spearman, an individual difference in performance is due to people differing in their ability to carry out the three laws of neogenesis, especially the eduction of relations and correlates. While general reasoning processes are described in some detail in Sternberg & Holland et al., it is not clear which processes distinguish those of high Gf from those of low Gf.

Holland et al. (1989) do briefly mention a potential source of individual differences. They claim that individual differences in induction are due to the accumulation of goal directed rules (conceptual knowledge). The size and type of the store of rules will differ from individual to individual, depending on their experience. This focus on knowledge as a source of individual difference in induction is at odds with the conceptualization of Gf in differential literature which sees Gf as the ability to reason and solve novel problems undistorted by individual differences in knowledge (Roberts & Stevenson, 1996). In fact, if individual differences in induction are due to the accumulation of knowledge, then Gf would be a context-dependant construct and closer in nature to Gc.

As for Sternberg (1986), while his focus is on outlining a unified theory of reasoning and is not especially concerned with individual differences in reasoning, he does refer to potential sources of individual differences. Like Holland et al. (1989), Sternberg hypothesizes that the possession of the relevant knowledge-based inferential rules would influence reasoning performance. An additional factor in performance is the individual’s working memory capacity, because according to the theory, problems with more elements and intermediate steps than a person can handle will affect their ability to access relevant rules. That is, working memory capacity is only an issue because it potentially limits access to relevant knowledge.
In summary, the cognitive theories of Sternberg (1986) and Holland et al. (1989) emphasize the importance of knowledge to reasoning in general, as well as refer to knowledge as a source of individual differences. Yet, somewhat paradoxically, Gf is conceptualised in the differential literature (for instance by Carroll, 1993; McGrew, 2005; Spearman, 1932) as a context-free ability that is relatively uninfluenced by past experience and knowledge. The follow sections further explore the potential synergy in these theories as they relate to Gf.

3.6 Central Thesis and General Predictions: Knowledge, Novelty, Learning and Combating Proactive Interference are all involved in Gf Task Performance

We propose a general hypothesis that may reconcile the need for knowledge (emphasized in the cognitive reasoning theories as being essential to induction) with the observation in the differential literature that Gf is involved in tasks that are considered novel. The hypothesis is based on the premise that there is a difference between processes involved in reasoning that contribute to individual differences in reasoning performance, and processes involved in reasoning that do not contribute to individual differences. Knowledge may be involved and important in reasoning in Gf tasks, but it may not contribute to individual differences in performance in Gf tasks. Instead, we propose that what contributes to individual differences in Gf tasks may be learning – a dynamic, fluid and context-free process. We also propose that learning may act as the mediator between the novelty in Gf items and the knowledge needed to solve them. If correct, Gf (being an individual differences construct, captured by variability in performance) could still legitimately be conceptualised as fluid and context-free. That is, we hypothesize that while prior experience and knowledge may be important to inductive reasoning, what contributes to individual differences in Gf task performance (and hence the Gf factor), may be dynamic and context-free learning.
Indeed, Holland et al. (1989) considers learning to be the central process in inductive reasoning but did not consider it a source of individual differences. We test the hypothesis that it is. In addition to this overarching and encompassing thesis, we test four general intrinsically-related hypotheses:

**Knowledge, Novelty and Learning**

**Hypothesis 1:** Complex Gf items contain novel relationships that are very difficult (or near impossible) to induce without guidance from some prior conceptual knowledge. This prior knowledge is (gradually) acquired by attempting earlier, easier, items which are more familiar. Thus, learning is a necessary process in Gf tasks.

**Knowledge:** Inductive Gf tasks typically involve items that contain multiple elements (e.g., numbers, shapes, colours, etc), whose relationships are governed by specific rules that the reasoner must induce (“the eduction of relations” in Spearman’s (1932) terms). The reasoner is then required to find/choose another element that would be consistent with the rule (“the eduction of correlates” in Spearman’s terms).

Gf tasks are presented in easy-to-hard format, primarily to motivate participants. Early in the task, Gf items conventionally consist of simple relationships to define the rules that the reasoner must induce. Arguably, because these are simpler relationships, they may also be relatively familiar relationships that one may have encountered in school or in life. It is possible that on these items, we get some guidance from prior knowledge brought from outside of the test situation (Holland et al., 1989). As one gets further into the task, the relationships tend to become more complicated. This has the effect of increasing the level of difficulty. The most popular (sometimes post-hoc) explanation for this is that more complicated rules have many more components and therefore place greater demands on
working memory capacity (Halford, Cowan, & Andrews, 2007; Halford, Wilson, & Phillips, 1998). However, this increase in complexity may also have the effect of increasing the novelty of the underlying relationships and hence the novelty of the item.

Learning & Novelty: Passmore (1935) highlights why learning may be very important in Gf tasks, “Things have various relations or associations, the associations we select being determined by the forces operating in us, and their past experience. Some relations are more generally observed than others, but since there are always several relations existing objectively between things, it is quite arbitrary to settle on one of them as being the sort of association an intelligent person would make” (p. 287).

That is, the correct answer to a problem that might be used to measure Gf may not always be clear. The test makers’ idea of the correct response is typically based upon a novel rule he or she constructs to govern how elements in an item are related to each other. Consider an example: The Number Series Completion item – 1 3 4 7 11 18 ? – requires one to determine the number which comes next in the series. The rule used to develop this item is to sum adjacent terms. That is, the sum of 1 and 3 is 4. The sum of 3 and 4 is 7; the sum of 4 and 7 is 11, and so on. Applying this particular rule indicates the next element should be “29”. However, how do we score a participant who gives the response “19” along with the statement: “the numbers are increasing, so any number larger than the last will be correct”? The reasoning is valid, it is just not the one intended.

The point here is that the response that is considered correct is often the one that conforms to the rule that the test maker had in mind. However, it is possible that there are responses that are considered “incorrect” by the test makers, but nevertheless can still be based on consistent rules that the test makers had never considered. Even when a single response might be
considered correct, different people might give different but equally valid reasons. Indeed, the rules that are included in Gf tasks are often quite novel (arbitrary) (Carpenter et al., 1990). Thus, before a difficult Gf item can be solved, the reasoner may have to first know (or learn) what relations or rules are considered legitimate by the test maker – they must narrow in on the intended relations and reduce the initially arbitrariness. Earlier problems in Gf tasks which have simpler rules give participants insight into what relations are considered legitimate in the harder problems. That is, while the novelty of Gf tasks may serve to equate pre-existing declarative knowledge, learning during the task may be necessary to reason effectively. Without within-task-learning, reasoning may simply involve a pseudo-guessing trial-and-error approach.

An often untested assumption is that people’s performance on later, more complicated items would be the same regardless of whether they had seen the earlier items. However, if Gf tasks are largely novel and inductive reasoning requires guidance from knowledge, they may be harder or even unsolvable without earlier, more familiar items to act as guides. That is, contrary to common assumptions, earlier Gf task items may act as opportunities to learn about the novel relationships in later, more difficult items.

**Learning and Gf**

**Hypothesis 2:** Learning is a source of individual differences in Gf task performance.

That is, those of higher Gf ability benefit more from within-task learning opportunity than those of lower ability.

If Hypothesis 1 is correct, then learning would be a necessary process required for performance on Gf tasks. Hypothesis 2 further predicts that those of higher Gf benefit more from the learning in Gf tasks.
As mentioned in Chapter 2, empirically, Gf shows consistently strong relationships to WMC and the reason for their link may be learning. Verguts and De Boeck (2002a; Verguts & De Boeck, 2002b) put forward the explanation that people with higher WMC can store many solution principles over items and use them to solve harder items. That is, they are better reasoners because they are better learners across the task. This is consistent with Holland et al.’s (1989) hypothesis that induction involves inferential processes that expand knowledge in the face of uncertainty. Furthermore, the empirical work of Carlstedt et al. (2000) suggest that a sequence of homogenous items is a better measure of Gf than a sequence of heterogeneous items; and this may be because homogenous items facilitate learning opportunity.

**Proactive Interference and Gf**

**Hypothesis 3:** Learning leads to a build up of proactive interference.

**Hypothesis 4:** Those of higher Gf ability must be better at combating proactive interference (because of Hypothesis 2 and 3).

Proactive interference is not a topic we have mentioned much in this chapter because it is not discussed much in Spearman (1932), Sternberg (1986) nor Holland et al, (1989). However, it is an issue we would like to address because it has been framed as a competing hypothesis to the learning hypothesis (Hypothesis 2), by Unsworth and Engle (2005a). As mentioned in Chapter 2, Unsworth and Engle argue that what is important to performance in Gf tasks is the ability to control attention, especially under conditions of distraction and interference. Specifically, Unsworth and Engle hypothesize that findings where homogenous sequences of items have been shown to be better measures of Gf and share more variance with WMC tasks have not been because of any learning effects. Instead, they argue that it is due to such items creating conditions where more proactive interference builds up from previous solutions.
From this perspective, Gf is responsible for inhibiting what has just been learnt (rather than for learning). Thus, their argument is that under conditions such as homogenous presentations, participants have to control their attention and try to block out or inhibit irrelevant aspects from previous, similar items; and it is for this reason that such items would be better measures of Gf and share more variance with WMC. That is, those individuals who are better at combating proactive interference during the task would perform better.

While Unsworth and Engle (2005a) put forward their proactive interference hypothesis as an alternative to the learning hypothesis, controlled attention may also be required for learning because controlled attention is purportedly needed to maintain information through activation of relevant brain circuitry (Heitz, Unsworth, & Engle, 2005). Thus, the two hypotheses (if true) may not be mutually exclusive.

Furthermore, Holland et al. (1989) predict that learning may result in an entrenchment of certain rules that may lead to proactive interference. Consistent with this, Sweller and Gee (1978) empirically found that the more one learns about a certain rule type (indicated by better performance on a transfer item), the more proactive interference builds up and this leads to poorer performance on subsequent items that rely on different rule types.

Thus, if Hypothesis 2 is correct - that those of higher Gf learn more, then these individuals must also be better at combating proactive interference. Otherwise, the benefits of being better learners would be lost when they encounter other types of items.

Hypotheses 1 and 2 will be empirically addressed in Chapters 4 to 7. Hypotheses 3 and 4 will be empirically addressed in Chapters 5 to 7.
3.7 Summary

Tasks that have the highest Gf loadings are typically inductive reasoning tasks. In order for hypotheses about Gf processes to be formulated and tested, comprehensive, sufficiently general conceptualizations of inductive reasoning are needed. However, such conceptualizations are not common. Three frameworks from different perspectives were identified: those of Spearman (1932), Sternberg (1986) and Holland et al. (1989). Each theory has a different emphasis on what is important in induction. Drawing on these different features has the potential to further our understanding of Gf. Sternberg emphasizes the use of prior conceptual knowledge, Holland et al. emphasize conceptual knowledge and learning, while Spearman emphasizes the eduction of relations and correlates which seems to occur in isolation from previous knowledge and learning. Only Spearman’s account is consistent with the dominant, conventional novelty view which sees Gf as largely uninfluenced by prior experience and knowledge. This is to be somewhat expected, of course, given the historic and empirical links Gf has with g.

A middle-ground position is that Gf tasks are largely novel, but allow learning to occur across the task and what is learnt provides reasoners with the relevant knowledge that Sternberg (1986) and Holland et al. (1989) describe as necessary for induction. Also, individual differences in Gf task performance may be at least partly due to the amount of knowledge learnt across the task, rather than knowledge brought to the task. Specifically, Gf items have traditionally been presented in easy-to-hard order to motivate participants but easier, simpler, more familiar items may (unintentionally) act as necessary learning aids for harder, complex and more novel items. Those of higher Gf ability may be more able to benefit from this learning opportunity than those of lower Gf. However, this learning is likely to create a build-up of proactive interference. Hence, those of higher Gf also need to be better at combating
proactive interference. The learning hypothesis and the proactive interference hypothesis have been framed as competing hypotheses for the link between Gf and WMC (the source of individual differences in Gf tasks); but we argue that they may not be mutually exclusive and may actually go hand-in-hand.

There are four general predictions. First, it is predicted that Complex Gf items contain novel relationships that are very difficult (or near impossible) to induce without guidance from some prior knowledge. This prior knowledge is (gradually) provided by earlier, easier, items which are more familiar. Second, it is predicted that those with higher Gf learn faster than those with lower Gf. Third, it is predicted that learning produces proactive interference. And finally, fourth, it is predicted that those with high Gf are better able to combat proactive interference than those with lower Gf.
4.1 Introduction

In the previous chapter, the theories of Sternberg (1986) and Holland et al. (1989) were put forward as cognitive theories that may be relevant to processes involved in Gf tasks. However, both theories emphasise the importance of conceptual knowledge in reasoning, while Gf is thought to be a factor related to the ability to deal with novel problems.

Two of the four core hypotheses we have proposed (in the previous chapter) are relevant to resolving this apparent contradiction. The first one was that complex Gf items contain novel relationships and are very difficult (or near impossible) to induce without guidance from some prior knowledge. This prior knowledge is (gradually) provided by earlier, easier items within the test which are more familiar. That is, we hypothesise that items become less novel through learning which we predict is a necessary process in Gf tests. In the second hypothesis we predict that this learning is a source of individual differences in Gf test performance over and above reasoning per se. That is, those of higher Gf ability benefit more from within-test learning opportunity than those of lower ability.

These hypotheses will be empirically tested in this chapter using Raven’s Advanced Progressive Matrices (“Raven”) (Raven, 1962). Specifically, we examine whether rule learning occurs, whether it is necessary for it to occur (for induction to take place) and whether it is a source of individual differences in Raven performance.
4.2 Learning and Raven

Although the novelty of Gf tasks may serve to equate pre-existing background knowledge, learning during the task may be necessary to reason effectively. Raaheim (1988) found that there is an optimum level of task novelty that correlates best with independent Gf markers. He used problem solving tasks from the experimental literature which correlated very poorly with Gf measures. However, when subjects became more familiar with the tasks (either through more exposure to such tasks or through being given the rules), the tasks became highly correlated with Gf. Also, as subjects became increasingly familiar with the problem solving tasks, their relationship to the Gf tasks dropped. Hence, Gf tasks may require some amount of learning and may not be devoid of the use of knowledge.

The Raven task is the task that loads the highest and most consistently on Gf factors (Marshalek, Lohman, & Snow, 1983). This suggests that this particular task might tap into more of the processes and abilities central to the Gf construct than any other task. Also, correlations between Raven scores and other measures of intellectual achievement suggests that the underlying processes may be general rather than specific to the task (Carpenter et al., 1990). Hence, many studies have used Raven or Raven-like tasks to investigate the cognitive processes involved in Gf (for a review see Primi, 2002 and for examples see Bors & Vigneau, 2003; DeShon, Chan, & Weissbein, 1995; Embretson, 1998; Meo, Roberts, & Marucci, 2007; Schiano, Cooper, Glaser, & Zhang, 1989; Unsworth & Engle, 2005a; Verguts, De Boeck, & Maris, 2000; Vigneau, Caissie, & Bors, 2006). Through this research, many Gf-related processes have been identified. Indeed, Raven’s is likely to have multidimensional performance determinants (DeShon et al., 1995).

As discussed in the previous chapter, learning is argued to be a process that is essential to inductive reasoning (Holland et al., 1989) and contributes to individual differences in Gf test
performance (Carlstedt et al., 2000). The empirical focus of the current chapter is to examine whether learning takes place in Raven and if so, to what extent it is a source of individual differences in performance on this test.

While some researchers assume that learning does not take place in Gf tasks such as Raven (Sternberg et al., 2002), the evidence regarding this issue is not clear. For instance, Unsworth and Engle (2005b) found that Raven’s correlated significantly with the degree of learning on a serial reaction time task in an intentional learning group. Yet, in Campione, Brown, Ferrara, Jones and Sternberg (1985) when individuals were matched on mental age using the Weschler Intelligence Scale for Children (WISC), intellectually disabled children showed no difference in learning simple Raven’s rules when compared to children of average intelligence or above. Thus, the evidence regarding the contribution of learning processes to individual differences in Raven is not clear. In later sections we will review some of the research aimed at understanding Raven performance and discuss how learning processes may fit in with other processes that have been identified as being involved in Raven; but first, we describe the Raven test in more detail.

### 4.3 Raven’s Advanced Progressive Matrices

Raven’s items are usually considered to involve the induction of rules from an incomplete 3x3 matrix of geometric shapes and figures. An example item is presented in Figure 4.1. The participant has to scan the rows and columns of the matrix and work out the rules that determine how these shapes and figures are laid out. They are then required to select the missing entry from the eight response options which are presented below the matrix, based on the rules that they think determine the organisation or grouping of the shapes/figures. Items differ with regard to the number of rules they contain, the types of rules and the combination of rules. Most items involve the instantiation of multiple rules, either as different rule types or
several instances of the same type of rule. Items are in easy-to-hard order. Generally, earlier items contain fewer rules. Problems also often require the participant to work out which figures, shapes or attributes (such as orientation of the figures/shapes) are governed by the same rule. Carpenter et al. (1990) named this latter requirement “correspondence finding”.

In a series of studies that analysed performance on Raven by examining verbal protocols, eye-fixation patterns and computer simulation models, Carpenter et al. (1990) concluded that the following processes are involved and common to everyone in reaching a solution to a Raven item:

1. Encoding of the figures/shapes.
2. Finding correspondences between figures – “correspondence finding”.
3. Working out the rule relating the figures/shapes.
5. Repeating the above processes on other parts of the item (that is, discovering other rules because a single item often contains multiple rules).

Thus, Carpenter et al. (1990) have identified some processes that occur in Raven that are relevant to all participants. Individual differences may exist for all or only some of these processes. The following section reviews some of the research aimed specifically at identifying processes that contribute to individual differences in performance in Raven.
While Carpenter et al. (1990) did not feature any actual Raven’s items in their paper to protect the security of the Raven Problems, this item has already been featured in Vigneau et al. 2006.

Figure 4.1. An example to illustrate the format of Raven test items. Participants look at the pattern, decide what the missing piece must be like to complete the pattern correctly, both across and down, and then find the correct piece out of the eight options shown. Only one of these options is perfectly correct. The organisation of the shapes and figures in the entries (in the rows and columns) of this item can be described by three rules:

1) Each row/column contains three geometric figures (a circle, a diamond and a square) distributed across its three entries.

2) Each row/column contains dotted lines with different orientations. The orientation of the lines (forward leaning, backward leaning and vertical) is distributed across its three entries.

3) The number of dotted lines is constant within a row but varies down the columns (one, two, three).

The missing entry can be generated from the above rules. Rule 1 specifies that the answer should contain a diamond (because the last column and row already each have a square and circle). Rule 2 specifies that it should contain backward leaning lines and Rule 3 specifies that the number of dotted lines should be three. These rules converge on the correct response option 5. Some of the incorrect response alternatives are designed to be in accordance with an incomplete instantiation of the rules. For example, if a participant knew only of Rule 1, she might choose response option 2 or 8. More difficult problems contain more rules and/or more difficult rules, and/or more shapes/figures per entry. This is the first item from Raven’s Advanced Progressive Matrices, Set II (Raven, 1962). The analysis of the rules in this item is based on the work of Carpenter et al., 1990.

\footnote{While Carpenter et al. (1990) did not feature any actual Raven’s items in their paper to protect the security of the Raven Problems, this item has already been featured in Vigneau et al. 2006.}
4.4 Search for Sources of Individual Differences in Raven Performance

Research exploring processes that contribute to individual differences in Raven performance often manipulate various characteristics of the task (Primi, 2002). Changes in performance due to these manipulations imply that the characteristics manipulated tap into processes that contribute to individual differences on overall Raven performance. Research suggests that the following characteristics of Raven tap into individual differences in performance:

1. The amount of information within an item (the number of shapes, figures, attributes and/or rules).
2. The type of rules involved in an item.
3. The perceptual organisation of the different elements (shapes/figures/attributes) within an item; and interestingly,
4. The amount of learning opportunity provided within the test.

However, it is often not clear what cognitive processes these characteristics implicate. Furthermore, there is often some controversy around the claims that these characteristics contribute to individual differences in Raven performance. We will first review the evidence and then propose how learning processes may be related to the mentioned Raven test characteristics that seem to tap into individual differences.

4.4.1 Amount of Information

The amount of information refers to the number of shapes, figures, attributes or rules involved in an item. It is information that must be processed in working memory and should contribute to working memory load and hence, to the difficulty of an item (Mulholland, Pellegrino, & Glaser, 1980). Gf items often contain large numbers of steps, stimuli and sub-results that must be stored and processed simultaneously (Carpenter et al., 1990); and
individuals with larger working memory capacity (WMC) may be more successful at this because WMC constraints limit one’s performance. Hence, the amount of information in individual Raven items may contribute to individual differences in overall Raven performance because people differ on WMC (Carpenter et al., 1990; Embretson, 1995; Pellegrino & Glaser, 1980).

However, researchers such as Salthouse and Pink (2008), Unsworth and Engle (2005a), and Verguts and De Boeck (2002b) have found that the relationship between WMC tasks and Gf tasks are fairly constant, regardless of the amount of information contained in the Gf items. Thus, Salthouse and Pink (2008) concluded that the relationship between Gf tasks and WMC tasks could not be attributed to individuals performing better on Gf items being capable of processing and storing more within-item information than individuals who perform poorer on Gf items. Hence, the amount of information in individual Raven items may not contribute to individual differences in overall Raven performance after all.

4.4.2 Learning Opportunity

There is some evidence that learning processes contribute to individual differences in Raven. Using a modified version of Raven, Verguts and De Boeck (2002a) showed that participants were able to learn Raven rules and then use them repeatedly in other items. As they describe it, a small set of rules is repeatedly applied over items by subjects and they become more fluent over repeated applications. However, they did not investigate the relationship between their learning variable and individual difference factors such as Gf or WMC. Hence, in a follow-up study Verguts and De Boeck (2002b) examined the same issue again, but this time with reference to WMC. WM load was kept low within each Raven item. They found that Raven performance still correlated with a WMC measure. Also, items with rules presented consecutively were easier than items where rules were presented in alternating sequence.
They concluded that the results together suggest that Raven rules become “primed” (or learnt) and the amount of priming is a factor of individual differences related to WMC. They argued that those with higher WMC were able to store many rules over items and hence perform better than those with lower WMC.

However, Verguts and De Boeck (2002a; Verguts & De Boeck, 2002b) really only showed a) that WMC matters in Raven, even when WM requirements in individual items are low, and b) that it is easier for everyone to solve Raven items when a rule is presented consecutively.

Firstly, it should be noted that understanding what makes items easier for everyone is not necessarily the same thing as understanding sources of individual differences. For example, learning opportunity may benefit those of lower Gf ability just as much as those with higher Gf ability. Secondly, the moderate relationship between WMC and Raven (containing items of low WM requirements) may have been due to some other (perhaps unknown) reason not related to learning. For example, those with higher WMC may be better able to combat proactive interference from solutions to older items, and hence, perform better on current items (Unsworth & Engle, 2005a).

### 4.4.3 Perceptual Organisation

According to Primi (2002) perceptual organisation of shapes and figures in Raven is relevant to the process he calls *abstraction*, which seems similar to the process that Carpenter et al. (1990) call *correspondence finding*. Items often require the participant to work out which elements (shapes/figures/attributes) are governed by the same rule – that is, which elements correspond with one another. It involves working out what is relevant and what is the appropriate representation of the elements. This includes constructing *conceptual* representations that are only loosely tied to perceptual features (i.e., abstractions). Even when people have been told about all the possible rules, they do not always find it easy to see how
the relevant rule(s) apply to the elements, in a particular item (Carpenter et al., 1990). This may be because they are unable to construct the appropriate representation of the elements and hence, would not be able to see which elements are governed by the same rule. This is illustrated in Figure 4.2.

Figure 4.2. An item that illustrates a simple example of the correspondence finding/abstraction requirement and the distribution of two rule. The rule corresponds separately to the diamonds, squares, sets of four dots, circle, horizontal crosses and vertical crosses. Each shape/figure appears twice (and only twice) in a row and/or column – this is the distribution of two rule. However, due to the misleading perceptual cues caused by the overlap of shapes/figures, it is hard to mentally represent them distinctly as diamonds, squares, circles etc. This can make it hard to identify the distribution of two rule.
Primi (2002) argues that the difficulty of this abstraction (or correspondence finding) depends on how the elements are perceptually organised or grouped. Certain ways of organising the elements can result in ambiguity and misleading cues that make the process of finding a correspondence among them difficult. That is, certain groupings of elements can increase the difficulty of working out which figures, shapes or attributes are governed by the same rule. Drawing on the gestalt principle of perceptual grouping of visual perceptions, Primi explains that some groupings are harder to work with than others because they violate our tendency to group things by certain perceptual features such as proximity, similarity, common region and continuity. Organisations that violate our natural grouping tendencies are called “non-harmonic” groupings, and those that do not violate them are called “harmonic”.

Carpenter et al. (1990) found participants struggled most with correspondence finding in items composed of multiple rules, probably because the involvement of several rules also necessitates the presence of several superimposed elements that form perceptually complex figures. In perceptually complex items, the likelihood of the formation of non-harmonic groups of elements based on perceptual features is increased (Primi, 2002).

There is some evidence that correspondence finding contributes to individual differences in Raven. Primi (2002) found that manipulations of perceptual organisation contributed substantially to item difficulty (more so than the amount of information or rule types involved in an item). That is, non-harmonic items were substantially harder than harmonic items and perceptual organisation was a better predictor of item difficulty in a Raven-type item than the amount of information or the rule type involved. Due to this, he concluded that Gf is most strongly associated with the processes of abstraction/correspondence finding.
When considering Primi (2002), it should again be noted that understanding what makes an item more difficult for everyone is not the same thing as understanding individual differences. Increasing a task’s difficulty does not always increase its Gf loading (Elshout, 1985). However, understanding what makes a task difficult is a starting point, to understanding individual differences (Lohman, 2000).

4.4.4 Rule Type

Carpenter et al. (1990) discovered that Raven’s rules could be classified into five main types:

1) **Constant in a row**: the same element (shape/figure/attribute) occurs throughout a row, but changes down a column.

2) **Pairwise progression**: An increment or decrement occurs between adjacent entries in an attribute such as size, position or number.

3) **Figure addition or subtraction**: A element from one column is added to (superimposed or juxtaposed) or subtracted from another figure to produce a third.

4) **Distribution of three values**: Three unique elements are distributed through a row.

5) **Distribution of two values**: Two elements are distributed through a row and the third value is null.

Carpenter et al. (1990) found that Rules 1) and 2) were easier than rules 3) and 4. Rule 5), the “distribution of two values rule” was the hardest. In subsequent discussions we shall call the distribution of two values rule the “D2” rule.

Carpenter et al. (1990) explored the individual differences question by developing two computer simulation models FAIRAVEN and BETTERAVEN. FAIRAVEN mirrored the
performance of a group of average performing university students tested and BETTERAVEN mirrored the performance of the best group of students. BETTERAVEN was programmed to have a larger working memory than FAIRAVEN and had access to all the rules identified in the Raven task. FAIRAVEN on the other hand could not identify the D2 rule (it was not programmed with this rule). Thus, it was unable to solve any item that required the application of this rule. These programming differences were implemented because Carpenter et al.’s (1990) study suggested that the major source of individual differences in Raven performance is working memory capacity (specifically, the ability to manage a large set of problem-solving goals) and the knowledge and induction of the D2 rule.

Mackintosh (1998) questions one aspect of the computer models of Carpenter et al. (1990). BETTERAVEN (and to a lesser extent, FAIRAVEN) were supplied with the necessary rules and therefore only had to recognise which rules were relevant to a given item; thus, the programs did not need to induce the rules from scratch. Mackintosh questions whether this would be the case in human participants, “But where have the rules come from? Do not people differ in their ability to discover them, and are not such differences partly responsible for differences in task performance?” (p. 305).

It could be argued that Carpenter et al.’s (1990) computer models of Raven performance assume that people bring certain declarative rules to the reasoning problem, whereas Mackintosh (1998) seems to suggest that there is a need to discover the rules during the reasoning process.

If this is the case, Carpenter et al.’s (1990) position is consistent with Sternberg (1986) and Holland et al.’s (1989) theories of induction (both outlined in Chapter 3). In these theories, items are not seen as self contained - the participant must come to the task with some
familiarity with the types of relationships and concepts that define the rules in Raven. In Sternberg’s terms, the rules in Raven would be selectively compared with declarative, inferential rules that the participant possesses and brings to the problem. This position has some support from other theorists. For example, Ohlsson and Lehtinen (1997) would argue that before a concept can be recognised, individuals must already possess the concept to be recognised. That is, in order for generalisations to be made (e.g., “three of same type of elements are distributed through a row in this Raven item”), individuals would need to already possess knowledge of those generalisations (e.g., the concept of “things can occur in lots of threes”) before they can see their instances in concrete examples (e.g., the instantiation of the constant in a row rule, in a Raven item).

In contrast, Mackintosh’s (1998) position is consistent with the traditional view of Gf in the differential literature - that Gf items involve reasoning which is not influenced by previous acquisitions of knowledge structures (Gustafsson, 1988; Richardson, 1991). Indeed, Gf tasks have been considered to be context-free tasks that can be used to tap into fluid abilities which are not influenced by a person’s previous experiences (Cattell, 1971). This line of reasoning entails the assumptions:

1) Gf items are self-contained and require participants to make generalisations about stimuli in the item (to find commonalities that form the basis of the rules) (Ohlsson & Lehtinen, 1997),

2) The generalisations the participants make are not dependent on their prior experiences before they arrived at the testing session, nor their experiences on prior items within the task.
In the previous chapter, we noted that in Spearman’s (1932) cognitive theory of “g” – the reduction of relations and correlates reflected such a view.

4.5 Combining the Evidence on Sources of Individual Differences

One way to resolve the tension between the need for knowledge in induction and the conceptualisation of Gf tests as being novel is to consider that learning processes bridge the need for knowledge in induction and the novelty of Gf tests. Gf tests may be largely novel but if they allow learning to occur across the test, this would provide reasoners with the relevant knowledge that theorists such as Sternberg (1986), Holland et al. (1989), and Ohlsson and Lehtinen (1997) argue is required for reasoning. Furthermore, *individual differences* in Gf test performance may be partly due to the amount of knowledge learnt across the test, rather than knowledge brought to the test (Verguts & De Boeck, 2002b).

It is likely that most (if not all) may come to the task with knowledge of the simpler rules used at the beginning of Ravens. As Verguts and De Boeck (2002a) point out, the rules of the first items of the task are usually very easy to work out. However, participants may need to learn through the task to solve more complicated (and hence, arguably, more novel) instantiations of the rules in later items. The assumption here is that complex items are novel because they contain multiple instances of various rules, which by themselves may not be novel, but when combined creates a novel product.

Indeed, as one progresses through the Raven test, the items become more and more difficult, and the correct rule becomes more difficult to elicit (Verguts & De Boeck, 2002b). It has been put forward that this is because harder items in Raven contain instantiations of *multiple rules* and *multiple types of rules*, increasing the likelihood of perceptually complex arrangements (Carpenter et al., 1990). As mentioned previously in this chapter, this also
increases the likelihood of non-harmonic groupings and hence, increases the difficulty of abstraction (i.e., finding correspondence) amongst the elements in the item (Primi, 2002). According to Primi, correspondence finding is the most difficult aspect of Raven. Furthermore, non-harmonic groupings by Primi’s definition are novel and unfamiliar, because they violate our tendency to group things by proximity, similarity, common region and continuity.

However, knowledge of Raven rules should make correspondence finding easier. Correspondence finding is influenced by the knowledge of the relationships between the various shapes/figures/attributes in an item (that is, the rules that govern those relationships) (Mackintosh, 1998) because those rules define the relevant groupings (correspondence) of the shapes/figures/attributes. Referring back to Figure 4.2 for an illustration; the overlapping nature of the shapes and figures makes it hard to mentally represent them distinctly as diamonds, squares, circles etc., and hence, makes it hard to find correspondence among them to identify the D2 rule. However, if one already consciously knows to look for the D2 to start with, the correspondence finding requirement should become easier.

Knowledge of the more complicated, novel instantiations of the rules may only be acquired through learning from easier items which contain simpler, more familiar instantiation of the rules. According to Verguts and De Boeck (2002a, 2002b) participants will try out the rules that they have tried before and the probability of trying out a certain rule depends on the activation of that rule. That is, on its occurrence in previous items and how well participants were able to learn it.

However, it is unclear if learning is a source of individual differences; that is, whether participants differ in how well they learn the rules. For example, rule learning may be a
necessary part of reasoning in Raven, but it may not be a source of individual differences if everyone learns to the same degree and benefits from it equally.

As mentioned previously, Carpenter et al. (1990) found that the best performers on Raven differed from average performers in that they were able to successfully solve items containing the D2 rule. In our search for processes that contribute to individual differences in Raven it may be helpful to empirically examine why the D2 rule has been found to be sensitive to individual differences; that is, whether items with the D2 require any learning processes.

If learning is important to individual differences, it could be that the D2 rule distinguishes those of higher ability from those of lower ability because there is not much opportunity in the task to learn it. It occurs quite late in the task (its first appearance is at item 22 out of a total of 36 items) and slower learners may be less able to learn about it. For example, if it occurred earlier in the task, it may not be able to distinguish high and low performers as well as it does in the current version of Raven.

Alternatively, the D2 rule may be more difficult than other rules because it relates objects in a manner that is not normal in real life (i.e., they are non-harmonic). The D2 rule specifies that two values from a categorical attribute are distributed through a row and the third value is null. This was illustrated in Figure 4.2. It is hard to think of a situation in real life where this is the normal way to group things. Some may learn about this relationship in real life and bring it with them to the task while others may not. According to Holland et al. (1989) individual differences in induction is due to the accumulation of goal directed rules (knowledge). The size and type of the store of rules will differ from individual to individual, depending on their experience. From this perspective, learning within the Raven task may not
be a source of individual differences in Raven task performance (nor even required); individual differences may be more about what the participant brings to the task.

To understand this further, we empirically investigate whether rule learning occurs in Ravens and if so, whether those of low and high Gf ability differ in their ability to learn rules during the test, particularly the D2 rule. We focus on the D2 rule in particular because it has been shown to be associated with individual differences in Raven performance (Carpenter et al., 1990).

This was done by manipulating the number of times people are exposed to the D2 rule. We presented three degrees of exposure: no exposure, limited exposure (the same number of D2 rules used in the Raven’s, 1975), and most exposure, as a between-subjects manipulation. We will call these conditions No, Limited and Most, respectively. The critical test of the learning hypothesis is performance on transfer items that contain multiple instantiations of the D2 rule (i.e., complex, non-harmonic, novel D2 items).

4.6 Competing Hypotheses and Specific Predictions

Based on the various theories and arguments presented so far in this thesis, the following three competing sets of specific hypotheses about the outcome of the D2 (learning opportunity) manipulation can be derived. We shall outline them briefly here first, then elaborate on the details shortly. The first, the “Classical Gf Hypothesis”, is based on the classical view that Gf items are novel. Thus, this hypothesis assumes that reasoning in Gf items is self-contained, based on the eduction of relations and correlates (Spearman, 1932), and is done in isolation from background knowledge and learning (Cattell, 1987). According to this account, learning does not occur across the test and hence does not contribute to individual differences in Raven performance. The second, the “Knowledge Hypothesis”, is
based on the works of Sternberg (1986) and Holland et al. (1989) who think reasoning is guided by prior inferential rules (conceptual knowledge). From this view, learning occurs to some extent but individual differences are largely due to the prior conceptual knowledge brought to the test. A corollary of this view is that too much learning opportunity within Raven should reduce its variance (and thus, its ability to distinguish between those of high and low Gf). This is because the things learnt during the test would serve to equate everyone’s relevant conceptual knowledge and hence, wash away knowledge as a source of individual differences in performance. The last set of hypotheses is based on Verguts and De Boeck (2002a) and Verguts and De Boeck (2002b) who emphasise the importance of rule learning in Raven performance. We shall call this the “Learning Hypothesis”. According to Verguts and De Boeck, rule-learning occurs in Raven and contributes to individual differences in performance. The Learning Hypothesis is in line with our general hypothesis that learning processes act as a bridge between the knowledge required for inductive reasoning and the novelty of Gf tests. Specifically, it is in line with our hypothesis that one must learn the rules from the earlier, more familiar items to be able to solve later items that contain more novel instantiations of the rules, and this learning is a source of individual differences in task performance.

The three sets of competing hypotheses each contain two parts. The first part contains predictions related to whether learning is necessary for reasoning in Raven before individual differences are taken into account. This would show up as differences in group means on the transfer items concomitant with the learning opportunity manipulation. The second part from each set contains predictions related to whether learning contributes to individual differences in Gf test (transfer item) performance. This would show up as interactions between the learning opportunity manipulation and Gf ability (a marker test of Gf). We expand on these hypotheses below:
1.1) **Classical Gf Hypothesis – Group Means:** *If induction is independent of any learning or knowledge, then the D2 manipulation will have no effect on the transfer items’ means in the different conditions (that is, the means will not differ significantly).*

The mean transfer item performance should not differ significantly as a function of exposure to the D2 rule. Also, performance in the No condition should be above chance levels (see Figure 4.3, graph A1) because some would be able to induce the answers (eduction of relations and correlates), based solely on the stimuli in the item. That is, participants discover the rules *during* the reasoning process *within* each of the transfer items.

1.2) **Classical Gf Hypothesis – Individual Differences:** *If induction is independent of any learning or knowledge, the D2 manipulation will have no effect on differences between high and low Gf ability groups on the transfer items.*

The transfer items should have the same correlations with a marker test of Gf, regardless of condition (Figure 4.4, graph B1). This difference could also be conceptualised in terms of mean performance on the transfer items for high Gf ability versus low Gf ability groups (Figure 4.5, Graph C1); there should be no interaction between condition and Gf ability on transfer item performance.

2.1) **Knowledge Hypothesis – Group Means:** *If induction is guided by prior conceptual knowledge, more exposure to the D2 will result in more opportunity to acquire conceptual knowledge and result in better performance on the transfer items.*

Transfer item means should be lowest in the No Condition but above chance levels because according to this line of reasoning, some individuals would already possess the D2 rule while
others will not. Additionally, transfer item means should increase with increased exposure to the D2 rule (as learning is thought to constantly take place in induction). That is, all participants should become more familiar with the D2 rule with more exposure to it (Figure 4.3, Graph A2).

This hypothesis is based on the works of Carpenter et al. (1990), Holland et al. (1989), Ohlsson and Lehtinen (1997) and Sternberg (1986). In these works, it is (at the very least) implied that induction requires individuals to possess relevant conceptual knowledge of the rules before their instances can be detected in concrete examples (that is, before they come to the task). For example, in Sternberg’s terms, the rules in Raven would be selectively compared with declarative, inferential rules that the problem solver already possesses. Also, Holland et al. (1989) consider learning to be an integral part of induction.

2.2) Knowledge Hypothesis– Individual Differences: If induction is guided by prior conceptual knowledge and individual differences is due to possession of the relevant knowledge, more exposure to the D2 rule will provide everyone with the necessary and sufficient knowledge about it. This should decrease the difference between high and low Gf ability groups on transfer item performance.

In terms of individual differences under this account, the transfer items’ correlations with a marker test of Gf should decrease with increased exposure to the D2 rule. This is because individual differences are hypothesized to be due to only some possessing the D2 rule while others do not (Carpenter et al., 1990). With increasing exposure, all will eventually possess the D2 rule by the transfer item stage. Furthermore, with increasing exposure, spotting of the rule may require less control and effort. The greater the degree of automatization, the smaller is the degree of reasoning involved (Sternberg, 1986), and thus, correlations with Gf should
decrease (Figure 4.4, Graph B2). In terms of differences between the means of high and low Gf ability groups, there should be an interaction, with those of lower ability benefitting more from increased exposure and those of higher ability reaching a ceiling level (see Figure 4.5, Graph C2).

3.1) Learning Hypothesis – Group Means: *If induction is dependent on rule-learning from earlier, easier items within the Raven task, more exposure to the D2 rule will result in more opportunity to learn about the rule and hence, better performance on the transfer items. No exposure to the rule will result in chance-level performance.*

Under this account a situation like Figure 4.3, Graph A3 would be predicted. If induction on complex, novel versions of D2 items (such as the transfer items) is dependent on learning from simpler, more familiar versions, then some learning opportunity will be needed before induction can take place. Thus, when participants are not given any prior exposure to the D2 rule, performance on the transfer items will be at chance levels and increase with increased in exposure.

This is based on the work of Verguts and De Boeck (2002a) and Verguts and De Boeck (2002b) who believe that rule learning is important in Ravens. They hypothesise that participant will try out rules they have tried before and the probability of trying out a certain rule depends on the activation of that rule, that is, on its occurrence in previous items and how well it has been learnt. In accordance with this, we hypothesise that easier rules are known to participants and get activated in their minds during early, easy problems. Progressing through the task, the instantiations of the rules in the items become more novel and more difficult, and the correct rule becomes more difficult to elicit. We hypothesize
participants then rely on what they learn from earlier, more familiar, easier items in the task to solve more difficult items that contain more complex, novel versions of the same rules.

3.2) **Learning Hypothesis – Individual Differences:** If induction is dependent on rule-learning from earlier, easier items within the Raven task, more exposure to the D2 rule will result in larger differences between high and low Gf ability groups on the transfer items, because those of higher Gf should be better learners.

Correlations between transfer item performance and a marker test of Gf should increase with increased exposure to the D2 rule. This is because according to Verguts and De Boeck (2002a), those higher in Gf benefit the most from learning opportunity (Figure 4.4, Graph B3). Note that this is different to Hypothesis 2.2 where those of low Gf are expected to benefit most from increased learning opportunity. In terms of differences between means for high and low Gf ability groups, there should be an interaction, with those of higher ability benefitting more from increased exposure (Figure 4.5, Graph C3).

![Figure 4.3. Predicted group means on the transfer items for the three conditions, as predicted by the three competing hypotheses. (A1) is based on the Classical Gf Hypothesis 1.1, (A2) is based on Knowledge Hypothesis 2.1 and (A3) is based on the Learning Hypothesis 3.1.](image)
Figure 4.4. Predicted correlations between the transfer items and a Gf marker, for the three conditions, as predicted by the three competing hypotheses. (B1) is based on the Classical Gf Hypothesis 1.2, (B2) is based on Knowledge Hypothesis 2.2 and (B3) is based on the Learning Hypothesis 3.2.

Figure 4.5. Predicted differences in means on the transfer items for High Gf and Low Gf groups, for the three conditions, as predicted by the three competing hypotheses. (C1) is based on the Classical Gf Hypothesis 1.2, (C2) is based on Knowledge Hypothesis 2.2 and (C3) is based on the Learning Hypothesis 3.2.

Thus, Figure 4.3 illustrates that Classical Gf Hypothesis 1.1 predicts means to be above chance but not differ as a function of condition, whereas Knowledge Hypothesis 2.1 and Learning Hypothesis 3.1 predict increases in mean performance as a function of learning opportunity. However, Hypotheses 2.1 and 3.1 differ in that 2.1 predicts that means should be
lowest in the No Condition but above chance, whereas 3.1 predicts that the means in the No Condition will be lowest and at chance levels.

Figure 4.4 illustrates that Classical Gf Hypothesis 1.2 predicts that transfer items should have the same correlations with the marker test of Gf (regardless of condition), the Knowledge Hypothesis 2.2 predicts that the correlations should decrease with increased exposure to the D2 rule, while the Learning Hypothesis 3.2 predicts that the correlations should increase with increased exposure to the D2 rule.

Lastly, Figure 4.5 illustrates that hypotheses 1.2, 2.2 and 3.2 could be conceptualised in terms of mean performance separately for those of high and low Gf. Classical Hypothesis 1.2 predicts that there will be no interaction between condition and Gf ability. Knowledge Hypothesis 2.2 predicts that there will be an interaction between condition and Gf ability, with those of lower ability benefitting more from increased exposure and those of higher ability reaching a ceiling level. Learning Hypothesis 3.2 also predicts that there will be an interaction between condition and Gf ability, but those of higher Gf should improve more dramatically than those of lower Gf with more exposure to the D2 rule.

4.7 Method

4.7.1 Participants

The participants were students, enrolled in a second year undergraduate subject at the University of Sydney and participated as part of their course work. In total, 284 students participated in the study. Due to a programming error, data was lost for 19 participants from the Limited condition. Due to testing time constraints, demographic information was gathered in class the week following the experiment. Demographic information could only be gathered
for 224 (163 female) out of the original 284 participants. Their mean age = 21.31 years (SD = 4.67).

4.7.2 Cognitive Tasks

1) Modified Raven’s Advanced Progressive Matrices Items

This task was completed on computers. Due to testing time constraints, it had a time limit of 33 minutes (after which the task timed out). However, previous studies in our lab have shown that 33 minutes is usually ample time for most participants to complete a total of 16 items. The task consisted of a “Learning Set” component and a “Transfer Set” component. More information about these items can also be found in Appendix A.1, A.2 and A.3. These two sets will now be explicated.

Raven Learning Sets (13 Items): No, Limited and Most

Items were selected from Raven’s Advanced Progressive Matrices (Raven, 1962) (which usually has 36 items) to form three 13-item versions of the Raven task. These will be referred to as the “Learning Sets”.

*In the Limited condition*, items for the Learning Set were selected to reflect the various types of items and the broad range of difficulties found in the original Raven. Although 13 items per task is rather short, this was necessary due to constraints on the amount of testing time that was available. Also, such short forms of the Raven have been found to retain the psychometric properties of the original version of the task, including its predictive validity (Vigneau et al., 2006). *This version contained four items with the D2 rule*. This condition was included so that any findings regarding the manipulation of the D2 rule could be understood with reference to the original task.
In the No condition, items in the Learning Set were the same as the items in the Limited condition’s Learning Set, with the exception that the four D2 rule items were replaced with items the author constructed that did not contain the D2 rule. Thus, this set contained no D2 rule items. Every attempt was made to ensure that they would be as similar as possible to the D2 rule items they replaced, in terms of the way the items looked, the number of rules in an item and the types of rules included in an item (other than the D2 rule).

In the Most condition, items for the Learning Set were the same as the items in the Limited condition but six additional D2 rule items were constructed and replaced six items that did not contain D2 rules. Thus, this set contained ten D2 rule items. Again, every attempt was made to ensure that they would be as similar as possible to the items they replaced. Also, unlike the Limited condition version (and the original Raven), the first D2 rule item occurs very early in the set.

Raven Transfer Set (3 Items)

Each Learning Set was followed by the same three transfer items (“Transfer Set”), containing multiple instantiations of the D2 rule. The items varied in difficulty. Items 1 and 2 were drawn from the end of the original Raven task and did not appear in any of the Learning Sets. Item 1 was considered to be easier than item 2 according to Raven, Court and Raven (1983). Item 3 was constructed by the author and was aimed at being less difficult than 1 and 2. It was included in case items 1 and 2 turned out to be too difficult and resulted in a floor effect. However, it was placed at the end in case it was too easy and resulted in providing training for items 1 and 2. The aim was for the Transfer Set to be able to capture a range of difficulty (especially in the Limited condition – which was equivalent to no manipulation).
2) Number Series (Gf Marker)

This is a well known Gf test (Marshalek et al., 1983; Quereshi, 2001). It was computer administered, with a time limit of 5 minutes. Items were presented in a random sequence (rather than the traditional easy to hard sequence)\(^3\) and consisted of a series of numbers which are related by various rules. Participants need to determine the next number in the series. Twenty-four items were drawn from an item bank from the Personality and Individual Differences Lab, at the University of Sydney.

An example item is as follows:

\[
12 \ 15 \ 12 \ 18 \ 12 \ 21 \ 12 \ ?
\]

In this case, the first number is a constant repeated at an interval of two, and every second number has 3 added to it to gain the subsequent value.

Thus, the correct answer is: \(24\)

4.7.3 Procedure

Testing was conducted in groups of no more than 23 students. All tasks were computer administrated. An experimenter provided general instructions at the beginning of each session and was present throughout the session to assist participants and ensure that the test protocol was followed. Participants were also provided with the general instructions on paper. A set of computerized instructions and practice phase preceded each task. Participants made their responses for all tasks using a standard keyboard and mouse. Participants were not allowed to use pen and paper for any of the other items.

\(^3\) The author acknowledges that this may result in a methodological effect. This will be addressed in the General Discussion of this chapter and Chapter 5.
The experiment was a between-groups design with the three levels (No, Limited and Most conditions). The dependent variable was Transfer Set performance. Number Series and Learning Set performance were used as covariates as described below.

4.8 Results and Discussion

This section will be organised as follows. First, we will present descriptive statistics and preliminary analyses. We then move on to analyses of the Learning Sets to gain a better understanding of the effects that our D2 rule manipulation had on performance on the Learning Sets – in short, a manipulation check. After that, we present analyses of performance on the Transfer Set as a function of condition and Gf ability. Finally, we present analyses of performance on the Transfer Set as a function of condition, Gf ability and performance on the Learning Set. This last set of analyses was included to examine the possibility that there is a difference in the amount of learning opportunity available when one is merely exposed to the D2 rule, versus when one is able to correctly solve a D2 rule item.

4.8.1 Descriptive Statistics

Available demographic information is presented in Table 4.1. Descriptive statistics for the tasks used in the study are presented separately by condition in Table 4.2. Cronbach’s alpha is in the acceptable range (for research purposes) for all variables. The Transfer Sets have low alphas, but with only three items that is to be expected. The Number Series alpha in the Limited condition is based on 18 items (instead of the full 24) because all participants got item 1 correct and items 20-24 either incorrect or did not reach them in the allocated time. In the Most condition, it is based on 21 items because all participants either got items 22-24 incorrect or did not reach them in time.
Table 4.1
Demographic information.

<table>
<thead>
<tr>
<th>Condition (n)</th>
<th>Age</th>
<th>Gender (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>No (90)</td>
<td>21.44</td>
<td>4.63</td>
</tr>
<tr>
<td>Limited (55)</td>
<td>20.46</td>
<td>3.26</td>
</tr>
<tr>
<td>Mostly (79)</td>
<td>21.96</td>
<td>5.74</td>
</tr>
<tr>
<td>Total (224)</td>
<td>21.96</td>
<td>5.74</td>
</tr>
</tbody>
</table>

*Note*. Based on incomplete data. Study N = 265. Demographic information was collected on a separate occasion to the test session.
Table 4.2
Descriptive statistics for tasks used in the study, for each condition.

<table>
<thead>
<tr>
<th>Task</th>
<th>Items</th>
<th>Mean (%Correct)</th>
<th>SD</th>
<th>n</th>
<th>α</th>
<th>Mean (%Correct)</th>
<th>SD</th>
<th>n</th>
<th>α</th>
<th>Mean (%Correct)</th>
<th>SD</th>
<th>n</th>
<th>α</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number Series (Gf)</td>
<td>24</td>
<td>9.84 (41)</td>
<td>4.49</td>
<td>104</td>
<td>0.91</td>
<td>9.25 (39)</td>
<td>3.78</td>
<td>65</td>
<td>0.87a</td>
<td>9.84 (41)</td>
<td>3.40</td>
<td>96</td>
<td>0.81b</td>
</tr>
<tr>
<td>Raven</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learning Set</td>
<td>13</td>
<td>10.02 (77)</td>
<td>1.99</td>
<td>104</td>
<td>0.58</td>
<td>8.60 (66)</td>
<td>2.71</td>
<td>65</td>
<td>0.73</td>
<td>7.92 (61)</td>
<td>2.83</td>
<td>96</td>
<td>0.71</td>
</tr>
<tr>
<td>Transfer Set</td>
<td>3</td>
<td>1.13 (38)</td>
<td>0.84</td>
<td>104</td>
<td>0.38</td>
<td>1.05 (35)</td>
<td>0.86</td>
<td>65</td>
<td>0.41</td>
<td>1.06 (35)</td>
<td>0.92</td>
<td>96</td>
<td>0.45</td>
</tr>
</tbody>
</table>

*Note. α = Cronbach’s coefficient alpha.*

a. Based on 18 items. Items 1, 20, 21, 22, 23 and 24 were excluded due to zero variance.
b. Based on 21 items. Items 22, 23, 24 were excluded due to zero variance.
4.8.2 Preliminary Analyses

Task correlations are presented in Table 4.3. All correlations were in the acceptable range.

<table>
<thead>
<tr>
<th>Task</th>
<th>Condition</th>
<th>No</th>
<th>Limited</th>
<th>Most</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number Series (Gf)</td>
<td>Learning Set</td>
<td>.34*</td>
<td>.40**</td>
<td>.36**</td>
</tr>
<tr>
<td></td>
<td>Transfer Set</td>
<td>.32*</td>
<td>.40**</td>
<td>.37**</td>
</tr>
<tr>
<td>Learning Set</td>
<td>Number Series (Gf)</td>
<td>.34*</td>
<td>.40**</td>
<td>.36**</td>
</tr>
<tr>
<td></td>
<td>Transfer Set</td>
<td>.34*</td>
<td>.34**</td>
<td>.58**</td>
</tr>
<tr>
<td>Transfer Set</td>
<td>Number Series (Gf)</td>
<td>.32*</td>
<td>.40**</td>
<td>.37**</td>
</tr>
<tr>
<td></td>
<td>Learning Set</td>
<td>.34*</td>
<td>.34**</td>
<td>.58**</td>
</tr>
</tbody>
</table>

*Note: * p < .05, ** p < .01; two-tailed.

Since our aim will be to compare the three condition groups on Gf items following an experimental manipulation, it is important to know that the groups did not differ significantly on Gf at the outset. Due to unequal sample sizes in the three conditions, the Univariate General Linear Model (GLM) analysis of variance procedure was used with Type III sums of squares to see if groups differed on the Gf marker, Number Series. The univariate analysis of variance showed that the groups did not differ significantly on the Gf marker, F (2, 262) = 0.554, p = .58.
4.8.3 Analysis of Learning Sets (Manipulation Check)

There is an expectation that overall Learning Set performance should differ as a function of the D2 learning opportunity manipulation. This is because D2 rules have been shown to be the most difficult Raven rules (Carpenter et al., 1990) and the more difficult items you have in a test, the poorer performance will be. Consistent with this expectation, we test the learning opportunity manipulation in two ways. Firstly, we expect mean performance to differ, as just described, as a function of the number of D2 items in the learning set. Secondly, because item solution demands fluid abilities, those with higher Gf should do better overall and the relationship between learning set performance and Gf should not differ as a function of condition.

A univariate GLM analysis of covariance (ANCOVA) was conducted to see if performance on the Learning Sets differed, with Number Series (Gf) performance entered as a centred covariate but this was mainly included to examine whether those of higher Gf did better on this measure, as they should have. There was a significant main effect of condition $F(2, 259) = 20.84, p > .01, \eta^2 = .139$ and as expected, a significant main effect of Gf ability $F(1, 259) = 40.05, p > .01, \eta^2 = .134$, but there was no significant interaction due to condition and Gf ability $F(2, 259) = 1.99, p = .14, \eta^2 = .015$. That is, the relationship between Gf and Learning Set performance did not differ as a function of condition.

Figure 4.6 illustrates point estimates of mean Learning Set performance in the three conditions for those one standard deviation above and below the mean on Number Series (Gf). As would be expected, those of higher Gf consistently scored higher than those of lower Gf. These point estimates were derived from the regression equation in Appendix A.4.
Figure 4.6. Point estimates of mean Learning Set performance for those 1 standard deviation below and those 1 standard deviation above the mean on Number Series (Gf), in the three conditions.

Since the effect of condition was significant, follow-up pairwise comparisons using a Bonferroni adjustment for multiple comparisons (see Table 4.4) were run. They showed that the No Condition’s Learning Set was significantly easier than the Learning Sets in the other conditions. Although there is a trend in the expected direction, the Most and Limited conditions did not differ significantly. That is, the No Condition’s set was easier than the other conditions’ Learning Sets. This may have been because it did not contain any D2 rule items - which are known to be harder than other rules (Carpenter et al., 1990; Embretson, 1998).
We suspected that the significant differences in difficulty between the No set and other sets would appear largest between items that did not contain the D2 rules and their corresponding items in the other sets that did contain the rule. Figure 4.7 shows a comparison of proportion correct for each Learning Set item in the three conditions and the Raven Manual (Raven et al., 1983). Items 7, 8, 11 and 12 from the No condition appeared to have been much easier than their corresponding items from the other conditions and the means reported in the Raven Manual. These items were constructed by the author and unlike their corresponding items, did not contain the D2 rule – and thus would be expected to have been easier. Conversely, items 3, 4 and 5 from the Most condition appeared to have been much harder than their corresponding items in the other conditions and the Raven Manual. They were created by the author to contain the D2, whereas their corresponding items in the other conditions/Raven Manual did not. Again, this is as expected. Overall, these results are consistent with previous findings (Carpenter et al., 1990) that suggest that the D2 rule is more difficult than other rules. The results also suggest that the difficulty of the Learning Set in the No condition differed from the Learning Sets in the other conditions because of the D2 manipulation.
Interestingly, the last items constructed by the author in the Most condition (items 9 and 13) did not appear to be significantly harder than their corresponding items that did not contain the D2 rule. This may be an indication that participants became more familiar with the D2 rule towards the end of the set.

![Figure 4.7](image)

*Figure 4.7. A comparison of proportion correct for each Learning Set item in the three conditions and the Raven Manual. Circled points indicate item containing the D2 rule.*

We also wanted to check that the Learning Set in the Limited condition was not too different to the original Raven task. The Learning Set in the Limited condition was meant to mirror the original Raven task so that we could relate any findings back to the original task. Item means from the Limited condition should have mirrored the means from the Raven Manual, since identical items were used. Figure 4.7 indicates that this appears to be the case up to item 7. However, from item 8 onwards, those in the Limited Condition outperformed the Raven Manual’s standardisation sample, particularly towards the end. This may have been because
the original Raven was a longer task. Participants from the current study may have been less fatigued on these later items than those whose results were reported in the Raven Manual. Alternatively or additionally, the calculations of item difficulties in the original time-limited Raven task were derived by considering unattempted items as incorrect. Thus, if the standardization sample had had sufficient time to complete all items (as the participants from our study did) perhaps they would have performed as well as our participants.

4.8.4 Analysis of Transfer Set

Another Manipulation Check – Was the Transfer Set Too Difficult?

The Transfer Set was composed of items taken from the more difficult end of the original Raven test. Hence, there is a danger that they may have been too difficult for many participants, who may have resorted to guessing rather than reasoning. Means for Transfer Sets for each condition are around 1 item correct (out of 3) (see Table 4.2). At first glance, this suggests that performance was at, or close to chance levels (see Appendix A.5 for calculations of chance levels). To explore this in more detail, simple t-tests were carried out to see if individual item means differed significantly from chance levels. The probability of choosing the correct option from 8 alternatives by chance is 12.5%. This is presented in Table 4.5.
Only performance on transfer item 2 (in all of the conditions) appears to have been at chance levels. However, this is not unusual for this item. Raven et al. (1983) indicates that it is a difficult item (they reported mean percentage correct of 5% - see Appendix A.3). Furthermore, the Transfer Set items in the Limited Condition captures a range of difficulty as intended (.09 - .65) and the means are similar in the other conditions. Thus, it appears that the Transfer Set was not too difficult and was able to capture individual differences in reasoning and not just random guesses.

**Main Analyses**

**Transfer Set Performance - No Exposure to the D2 Rule**

As just mentioned, Table 4.5 indicates that performance on the Transfer Set items was above chance in the No Condition. Thus, it could be concluded that prior exposure to easier D2 rule items within the task is *not required* for participants to be able to solve D2 items with more complex (arguably, more novel) instantiations of the rules.

### Table 4.5

*Means for individual Transfer Set items and their tests against chance level.*

<table>
<thead>
<tr>
<th>Item</th>
<th>No Mean</th>
<th>SD</th>
<th>t</th>
<th>Limited Mean</th>
<th>SD</th>
<th>t</th>
<th>Most Mean</th>
<th>SD</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transfer_1</td>
<td>0.37</td>
<td>0.48</td>
<td>5.06**</td>
<td>0.31</td>
<td>0.47</td>
<td>3.16**</td>
<td>0.35</td>
<td>0.48</td>
<td>4.67**</td>
</tr>
<tr>
<td>Transfer_2</td>
<td>0.08</td>
<td>0.27</td>
<td>-1.83</td>
<td>0.09</td>
<td>0.29</td>
<td>-0.90</td>
<td>0.13</td>
<td>0.33</td>
<td>0.00</td>
</tr>
<tr>
<td>Transfer_3</td>
<td>0.69</td>
<td>0.46</td>
<td>12.47**</td>
<td>0.65</td>
<td>0.48</td>
<td>8.71**</td>
<td>0.58</td>
<td>0.50</td>
<td>9.061**</td>
</tr>
</tbody>
</table>

*Note. Test value = 0.125; ** p < .01.*

*Significance indicates above chance level performance.*
This is consistent with the Classical Gf Hypothesis and Knowledge Hypothesis but inconsistent with the Learning Hypothesis. The finding is inconsistent with the Learning Hypothesis because it predicts induction is dependent on rule-learning from earlier, easier items within the Raven task and no exposure to such items will result in chance-level performance on items with complex versions of the rule (such as in the Transfer Set). The finding is consistent with the Classical Gf Hypothesis because it predicts that the D2 manipulation will have no effect on performance on the Transfer Set items (because reasoning occurs in isolation from learning). It is also consistent with the Knowledge Hypothesis, which predicts that even with no prior exposure to easier D2 rule items within the task, performance on the transfer items would still be above chance levels because some participants would bring the relevant, prior conceptual knowledge of the complex versions of the D2 rule. Thus, while it is possible to rule out the participants need to learn about the D2 rule in the task, at this point, it is unclear if participants bring conceptual knowledge of the complex versions of the D2 rule to the task or if they are able to induce the answer without recourse to prior knowledge.

Transfer Set Performance – Increasing Exposure to the D2 Rule

While participants may not need learning opportunity about the D2 rule to solve the Transfer Set, a related question that remains is whether more exposure to easier D2 rule items makes participants better at solving more complex D2 rule items. An ANCOVA was conducted to see if performance on the Transfer Set differed across the three conditions with Gf (Number Series) performance entered as a centred covariate. There was no significant main effect of condition F (2, 259) = 0.23, p = .79, eta2 = .002 and a significant main effect of Gf ability F (1, 259) = 37.16, p > .01, eta2 = .120. However, there was no significant interaction due to condition and Gf ability F (2, 259) = 1.01, p = .36, eta2 = .008.
Figure 4.8 illustrates point estimates of mean Transfer Set performance for those one standard deviation above and below the mean on Number Series (Gf) in the three conditions. As would be expected, those of higher Gf consistently scored higher than those of lower Gf. These point estimates were derived from the ANCOVA parameter estimates reported in Appendix A.6.

Figure 4.8. Point estimates of mean Transfer Set performance for those 1 standard deviation below and those 1 standard deviation above the mean on Number Series (Gf), in the three conditions.

The lack of main effect of condition suggests that more exposure to D2 rules does not appear to affect performance on harder D2 rules at the group level. This is consistent with the Classical Gf Hypothesis but contrary to the Knowledge Hypothesis and the Learning Hypothesis. The Classical Gf Hypothesis predicted that the D2 manipulation would not have any effect on the Transfer Set. The Knowledge and Learning Hypotheses predicted that more exposure to the D2 rule would result in more opportunity to learn about the rule and for
participants to acquire the relevant, conceptual knowledge; resulting in better mean group performance.

The lack of interaction between condition and Gf ability suggests that across-item learning processes and knowledge do not contribute to individual differences in performance. Again, this is consistent with the Classical Gf Hypothesis but inconsistent with the Knowledge Hypothesis and the Learning Hypothesis. The two latter hypotheses predicted interactions between condition and Gf ability (albeit in different directions). The Knowledge Hypothesis predicted that individual differences are due to possession of the relevant knowledge hence, more exposure to the D2 rule would provide everyone with knowledge about it, decreasing the gap between high and low Gf ability groups. The Learning Hypothesis predicted that individual differences are due to learning processes (specifically rule learning) and those of higher Gf learn more than those of lower Gf; hence, more exposure to the D2 rule would increase the gap between high and low Gf ability groups. In contrast, the Classical Gf Hypothesis predicted that learning manipulations would have no effect on Transfer Set performance; hence, those of higher Gf ability would consistently and uniformly outperform those of lower Gf ability, regardless of condition – which is what we found.

Summary of Analysis of Transfer Set

Thus far, support is strongest for the Classical Gf Hypothesis. Manipulation of the D2 rule seems to have had no noticeable effect on Transfer Set performance. That is, providing more exposure to the D2 rule does not seem to change performance in any way at the mean group level and when individual differences are taken into account. This suggests that participants do not learn about the rules across the task. Consistent with the Classical Gf Hypothesis, it
appears that induction within an item is done in isolation from previous experience – such as experience from other items in the task.

4.8.5 Analysis of Transfer Set when Actual Learning is taken into Account

Mere exposure to the rules may not be enough for participants to learn about the rules. Learning may require feedback and in Raven (partial) feedback is more evident when an item is answered correctly. That is, as stated by Verguts and De Boeck (2002a):

“...choosing a certain rule to solve the item and finding a response alternative that matches this rule provides the information that the rule chosen is the correct one. The idea is that, if one chooses an incorrect rule, the probability is low that one also finds a matching response alternative below....in [the] case the response is correct, implicit feedback is given about the correctness of the rule (note that the presence of feedback is critical in learning the rules)” (p.536).

With reference to the current experiment, providing more exposure to rules through the Learning Set is providing potential learning opportunity, but that learning opportunity may not be exploited unless the participant can answer the Learning Set items correctly. The more items the participant answers correctly in the Learning Set, the more actual learning opportunity is available to them.

Hence, two additional issues were explored. First, we examined the question of whether or not performance on the Transfer Set differs as a function of actual learning opportunity (performance on the Learning Set). The second issue explored was the question of whether those of high and low Gf ability benefit in the same way (if at all) from this actual learning opportunity.
A hierarchical moderated multiple regression analysis was conducted to determine whether performance on the Learning Set contributed to performance on the Transfer Set over and above Gf ability and whether the effects of actual learning opportunity were moderated by Gf ability. This was done separately for each condition because the items used in the Learning Sets obviously differed from condition to condition. To test the two-way interaction (Number Series performance X Learning Set performance), a multiplicative term was entered into the regression equation after controlling for the component main effects. To prevent multicollinearity this multiplicative term was formed using centred Number Series scores and centred Learning Set scores. The models are presented in Table 4.6 for easy reference and illustrated in Figure 4.9.
Table 4.6
*Summary of hierarchical regression analyses for variables predicting Transfer Set performance in the three conditions.*

<table>
<thead>
<tr>
<th>Condition Variable</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE(B)</td>
<td>β</td>
</tr>
<tr>
<td>No</td>
<td>0.06</td>
<td>0.02</td>
<td>.32**</td>
</tr>
<tr>
<td>Learning Set</td>
<td>0.11</td>
<td>0.04</td>
<td>.26**</td>
</tr>
<tr>
<td>Number Series x Learning Set</td>
<td>0.00</td>
<td>0.01</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$F$ for change in $R^2$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limited</td>
<td>0.09</td>
<td>0.03</td>
<td>.40**</td>
</tr>
<tr>
<td>Number Series</td>
<td>0.07</td>
<td>0.04</td>
<td>.22*</td>
</tr>
<tr>
<td>Number Series x Learning Set</td>
<td>0.01</td>
<td>0.01</td>
<td>.14</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$F$ for change in $R^2$</td>
<td>11.78**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Most</td>
<td>0.10</td>
<td>0.03</td>
<td>.37**</td>
</tr>
<tr>
<td>Learning Set</td>
<td>0.17</td>
<td>0.03</td>
<td>.52**</td>
</tr>
<tr>
<td>Number Series x Learning Set</td>
<td>0.03</td>
<td>0.01</td>
<td>.24**</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$F$ for change in $R^2$</td>
<td>14.78**</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note: Number Series and Learning Sets were centered at their means within each condition.

*p < .05, **p < .01.

a = marginally significant (p = .08)*
Figure 4.9. Three graphs showing point estimates of mean Transfer Set performance for those 1 standard deviation below, above and at the mean on Number Series (Gf) at 1 standard deviation below and above the mean for Learning Set performance (From Model 3, Table 4.6) for each condition, No, Limited and Most.
The effects are illustrated in Figure 4.9. Specifically, it illustrates separately for the three conditions point estimates of mean Transfer Set performance for those one standard deviation below, above and at the mean on Number Series (Gf), and at one standard deviation below and above the mean for Learning Set performance. The Learning Sets’ means and standard deviations referred to in the graphs are the separate ones for each condition. However, for ease of comparison across conditions, the Number Series mean referred to in the graphs is the grand mean and its standard deviation. That is, Low, High and Middle Gf refer to the same numerical value of Number Series performance in each conditions’ graph (-3.95, 0 and 3.95 respectively).

In the No condition, Number Series was a significant predictor of Transfer Set performance, the Learning Set contributed to performance over and above Number Series, but there was no interaction between the two. That is, those of higher Gf ability did better than those of lower Gf ability on the Transfer Set, those who did better on the Learning Set did better on the Transfer Set than those who did less well on the Learning Set; and the relationship between Transfer Set and Learning Set performance was not moderated by Gf ability.

In the Limited Condition, similar results were found. There was a main effect of Number Series (Gf), but the Learning Set’s Beta was only marginally significant (p = .08). Also, the graph hints that this relationship between Learning Set performance and Transfer Set performance is stronger for those of higher Gf than those of lower Gf, but the regression model indicates that this interaction is not significant.

In the Most condition, there were main effects of Number Series, Learning Set performance, and their interaction was significant. Those of higher Gf ability did better than those of lower Gf ability on the Transfer Set, and those who did better on the Learning Set did better on the
Transfer Set than those who did less well on the Learning Set. Also, the relationship between Learning Set and Transfer Set performance was stronger for those of higher Gf ability and this interaction is significant. Follow up analyses of the differences in slopes in the Most condition were explored using methods described in Preacher, Curran and Bauer (2006) and the summary is presented in Table 4.7 The results indicate that the relationship between Transfer Set and Learning Set performance gets stronger with higher Gf ability.

Table 4.7

Estimates for three simple slopes of Transfer Set performance as a function of Learning Set performance, at three different levels of Number Series (Gf) performance in the Most condition.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Low Gf (-1 SD)</th>
<th>Middle Gf (Mean)</th>
<th>High Gf (+1 SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE(B)</td>
<td>t</td>
</tr>
<tr>
<td>Learning Set</td>
<td>0.08</td>
<td>0.04</td>
<td>6.62</td>
</tr>
</tbody>
</table>

Note: Low Gf is at one standard deviation below the mean for Number Series, Middle Gf is equal to the mean and High Gf is at one standard deviation above the mean for Number Series.

Taken together, the results from the three conditions suggest the following. Firstly, Gf ability contributes to Raven performance (as generally expected). Secondly, the more items you are able to solve in the Learning Set, the better you do on the Transfer Set, independent of your Gf ability and regardless of condition. This may indicate that the better you do on Raven’s items, the better you do on other Raven’s items. However, it may also indicate the importance of across-item learning processes to performance on Raven. That is, the more items you are able to solve earlier on (such as those in the Learning Set), the more you are able to learn from them to help you do better on harder items that occur later (such as in the Transfer Set). Furthermore, across-item learning processes appear to be better exploited by those of higher Gf who seem to benefit the most from actual learning opportunity of the D2 rule in the Most Condition. Thus, learning processes may contribute to individual differences in Raven
performance because they do not benefit everyone equally. Those of higher Gf seem to be better at utilising them when there is more opportunity to utilise them (such as in the Most condition).

The evidence for the importance of learning processes to individual differences in Raven performance comes largely from the interaction between Learning Set performance and Gf Ability (Number Series performance) in the Most condition, where there was the most opportunity to learn the D2 rule. Had there not been this significant interaction, one could argue that the main effect of Learning Set in all the conditions could be attributed to “the better you do on Raven’s items, the better you do on other Raven’s items”, and rule out any involvement of learning processes. However, as we increased the opportunity to learn the D2 rule (from the No condition to the Most condition), those of higher Gf seemed to be able to increasingly benefit from it (when actual learning opportunity was taken into account), moreso than those of lower Gf. This would be consistent with the Learning Hypothesis, which predicts that more exposure to the D2 rule will result in larger differences in Transfer Set performance between those of higher Gf and those of lower Gf (since those of higher Gf would be better able to learn the D2 rule). These findings in the Most Condition are inconsistent with the Classical Gf Hypothesis. The hypothesis predicts that the performance (and individual differences in performance) on a given item is not dependent on prior items. The findings are also inconsistent with the Knowledge Hypothesis which predicts that more opportunity to learn would result in smaller differences between those of higher Gf and those of lower Gf (individual differences are thought to be due to prior knowledge rather than to knowledge acquired during the test).
4.9 General Discussion

This study investigated processes responsible for individual differences in Raven performance. One potential process that has been implicated in past theorising but neglected in empirical research is learning. Three competing sets of hypotheses (“Classical Gf”, “Knowledge” and “Learning” Hypotheses) were derived from reasoning theories in the differential and cognitive-differential literature to predict the effects of varying the amount of D2 rule learning opportunity on performance on more complex and (arguably) more novel, D2 items. Each set of hypotheses was divided into predictions about group means and predictions about individual differences. We shall discuss the findings for group means and individual difference separately.

4.9.1 The Effects of the D2 Manipulation on Group Means

It was found that in the no exposure condition, performance on the Transfer Set was above chance levels. Hence, it is likely that rule learning is not a necessary process in solving Raven items. It seems that it is possible to solve items with complex instantiations of D2 without any prior exposure to simpler versions of the rule within the task. This is not consistent with the Learning Hypothesis but consistent with the Classical Gf Hypothesis and the Knowledge Hypothesis. According to Sternberg (1986) and Holland (Holland et al., 1989) (on which the Knowledge Hypothesis is based), induction requires individuals to possess conceptual knowledge, relevant to the rules before their instances can be detected in concrete examples (that is, before they come to the task). Using this account, it was predicted that performance on the transfer items would be above chance levels (even without prior exposure) because some individuals would possess conceptual knowledge, relevant to complex instantiations of the D2 rule, while others will not. However, the finding is also consistent with the Spearman (1932) and Cattell (1987) (on which the Classical Gf
Hypothesis is based). Using this account it was predicted that some would be able to discover the rules (from scratch) during the reasoning process, based on the stimuli in the item – what Spearman calls the eduction of relations and correlates. Thus, it could be concluded that some participants come to the task with prior, conceptual knowledge of the rules or are able to work them out during the reasoning process solely within an item, independent of any prior knowledge. Since we do not have access to participants’ history and their entire store of conceptual knowledge, it is hard to tell which of the two explanations is relevant to our finding.

There is also an alternative explanation for why the Transfer Set performance was above chance levels for the No Condition. It could be that participants do rely on knowledge learnt from earlier items but do so using knowledge learnt from items containing other rules to help them solve the D2 Transfer Set items. In the introduction of this chapter it was put forward that correspondence finding may be influenced by the detection of the relationships between the various shapes/figures/attributes in an item (that is, detection of the rules that govern those relationships) (Mackintosh, 1998). That is, in Sternberg’s (1986) terms, successful selective comparison of declarative inferential rules that the participant possesses may assist the selective encoding of the relevant shapes/figures/attributes (which contributes to correspondence finding). However, the inverse may also be true. Selective encoding may assist the process of selective comparison of declarative inferential rules. That is, knowing that the elements can be parsed and how they can be parsed, may assist in the detection of the rules. Getting earlier items correct (regardless of the rules involved) may assist with this, by letting participants know that it is permissible to think of figures/shapes as overlapping and not just as whole elements. For example, easier rules (such as the “distribution of three” rule) may assist with correspondence finding; knowing about how things correspond may then assist with the detection of harder rules (such as distribution of two). It is not really possible
with the current data to tease this apart from the hypothesis that participants come to the test situation with knowledge of the rules or are able to discover them during the reasoning process within the item.

The study in the next chapter tries to rectify this situation by examining what happens when no learning opportunity at all is provided. That is, when no earlier items are presented before the transfer items.

**Increasing Exposure to the D2 Rule**

It was found that varying the amount of exposure to the D2 rule did not seem to change performance on the Transfer Set in any way at the mean group level. This suggests that as a group, participants do not learn about the rules across the task. This finding is consistent with the Classical Gf Hypothesis that suggests that induction within an item is done in isolation from previous experience, including experience from other items in the task. The finding is inconsistent with the Knowledge and Learning hypotheses which predict that increasing exposure should lead to increasing knowledge of the D2 rule and improved performance.

**4.9.2 The Effects of the D2 Manipulation (Individual Differences taken into Account)**

It was found that varying the amount of exposure to the D2 rule does not seem to change performance on the Transfer Set in any way even when individual differences are taken into account. Those of higher Gf ability consistently and uniformly outperformed those of lower Gf ability on the Transfer Set and there was no interaction with the number of times they were pre-exposed to the D2 rule. Again, this was consistent with the Classical Gf Hypothesis and inconsistent with Knowledge and Learning hypotheses. Thus, according to this finding,
participants do not learn from rules when they are provided with more instances of those rules.

Thus far in this Discussion section, the Classical Gf Hypothesis has received the strongest support. However, providing more exposure to rules through the Learning Set is providing potential learning opportunity but that learning opportunity may not be exploited unless the participant can answer the Learning Set items correctly. The more items the participant answers correctly in the Learning Set, the more actual learning opportunity is available to them.

It was found that the more items participants were able to solve in the Learning Set, the better they did on the Transfer Set, independent of their Gf ability (as indicated by Number Series) and the types of rules presented to them (that is, their condition). This may indicate support for the Learning Hypothesis; that is, the more items you are able to solve in the Learning Set, the more you are able to learn from them, to help you do better on harder items in Transfer Set. Furthermore, in the Most Condition (where there would have been the most opportunity to learn the D2 rule) there was an interaction between Learning Set performance and Gf ability; those of higher Gf ability were more able to benefit from their actual learning opportunity, to do better on the Transfer Set. This supports the Learning Hypothesis and contradicts the Classical Gf and Knowledge hypotheses.

4.9.3 Summary and Conclusion

The empirical aim of this study was to examine whether rule learning occurs in Raven, whether it is necessary for it to occur (before induction can take place) and whether it is a source of individual differences in Raven performance. The particular rule we examined was
the D2 rule, because it has been shown to distinguish good Raven performers from average ones (Carpenter, Just, & Shell, 1990).

It was found that it is not necessary for participants to learn about the D2 from simple items in order for them to solve items containing complex instantiations of the rules later in the task. Thus, we conclude that rule learning is not necessary for induction to take place.

However, we found evidence that rule learning occurs in Raven and contributes to individual differences in Raven performance, with those of higher Gf benefitting more from more exposure to the D2 rule. This was only evident when real learning opportunity was taken into account (as opposed to potential learning opportunity). When participants are exposed to D2 rule items, they only receive potential learning opportunity. It seems they may not be able to learn about the rule unless they get the item correct (and thus receive real learning opportunity).

**4.9.4 Limitations and Further Research**

The overall findings highlight a couple of important issues. Firstly, providing more exposure to rules is not enough to ensure that real learning opportunity is provided. In order to be able to learn from an item (at least in Raven), you need to be able to get it correct - perhaps in order to get implicit feedback that your hypothesis about the nature of the item is correct (Verguts & De Boeck, 2002a). Secondly, this may mean that those of lower ability may not be able to learn as much from items because they are not able to get as many earlier items correct and not because they are less able to learn about rules across the task. That is, across-item learning may be confounded with within-item induction. This is something we address in the next chapter.
Another potential confound that we address in the next chapter is the issue of the presentation format of the marker task. The Number Series task, which was used as the marker of Gf in this study, was presented in random order (rather than the traditional easy to hard presentation format of Gf tasks). However, by presenting it in this way we may have limited the learning opportunity available in the Number Series task. If an important component of individual differences in Gf is about exploiting learning opportunity, then by presenting our marker test in random format may have meant that it was not as good a marker of Gf as we would have preferred. We address this issue in the next study (in the next chapter).

Another shortcoming of our study is that we only examined rule learning - and only learning of the D2 rule. Carpenter et al. (1990) point out that their subjects were able to identify all the other rules and hence, these other rules would not be related to a source of individual differences. However, Mackintosh (1998) suspects that had Carpenter et al. (1990) studied a wider range of ability of subjects (rather than just university students), they would have had to allow for the possibility that some of the other rules may have also been related to individual differences. We suspect that Mackintosh is right.

Lastly, while our study shows learning about the D2 from simpler items is not necessary for induction to take place in more complex versions of D2 items, we were unable to rule out that participants may rely on other forms of learning to be able to solve more complex Raven items. We will also address this in the next study.
CHAPTER 5

EXPERIMENT 2:
DIFFERENT BUT RELATED: WITHIN-ITEM RULE DISCOVERY, ACROSS-ITEM LEARNING, USE OF PRIOR KNOWLEDGE AND PROACTIVE INTERFERENCE
A PILOT STUDY

5.1 Introduction

“To make the novel seem familiar by relating it to prior knowledge, to make the familiar seem strange by viewing it from a new perspective - these are fundamental aspects of human intelligence that depend on the ability to reason by analogy. This ability is used to construct new scientific models, to design experiments, to solve new problems in terms of old ones, to make predictions, to construct arguments, and to interpret literary metaphors” - Gick and Holyoak (1983 p. 2).

This chapter has two aims which are closely related to each other. The first aim is to address issues raised but unresolved by the experiment in the previous chapter (Experiment 1, Chapter 4). Experiment 1 highlighted that inductive processes within Gf tasks have been characterised differently by different researchers:

- Induction as involving learning that may occur across Gf items, within a task (“across-item learning”) (Carlstedt et al., 2000; Holland et al., 1989; Verguts & De Boeck, 2002a, 2002b),
- Induction as involving conceptual knowledge that may be brought to the task (Holland et al., 1989; Sternberg, 1986) and
• Induction as occurring within an item and independent of any prior knowledge or learning that may occur across the items (Cattell, 1987; Gottfredson, 1997; Spearman, 1932).

This chapter further examines the distinction between these views of induction and whether the processes hypothesised in each view occur in Gf tasks, whether it is necessary for them to occur for induction to take place and whether they contribute to individual differences in performance. Also, while the previous chapter focussed on the different predictions that these different views had for changes to performance on Gf tasks when the amount of learning opportunity within a task is varied, this chapter focuses on the similarities between the different views. Particular attention will be given to how the views may be compatible and describe different forms of analogical thinking that occur in Gf tasks.

This chapter’s second aim is to examine learning’s relationship with proactive interference. While across-item learning and combating proactive interference have been put forward as competing explanations for the Gf-WMC link and as competing explanations for individual differences in Gf (Unsworth & Engle, 2005a), it will be put forward that the two cannot be mutually exclusive explanations.

These issues will be explored in this chapter with a pilot study and continue into chapters 6 and 7. Chapters 6 and 7 examine the same issues as those which will be discussed in the current chapter, but will do so using two larger studies and more sophisticated data analyses.

5.2 Issues Raised by Experiment 1, Chapter 4

The results from Experiment 1, Chapter 4 highlighted a number of issues that we aim to explore further in this and subsequent chapters. Firstly, the study highlighted that
examinations of across-item learning may be confounded with within-item induction. The study found that varying the amount of exposure to simpler Raven “distribution of two” (D2) rule items did not have any impact on performance on more complex D2 rule transfer items, at the group level and when individual differences were taken into account. However, providing more exposure to rules is providing potential learning opportunity but that learning opportunity may not be exploited unless the participant can answer the items correctly; and thus, comprehend the rule. The more items the participant answers correctly, the more actual learning opportunity may be available to them. When actual learning opportunity was taken into account (as opposed to potential learning opportunity), those of higher Gf were more able to benefit from more exposure to D2 rules. This may also mean that those of lower Gf ability have been shown to not learn as much as those of higher ability because they are not able to get as many earlier items correct, rather than because they are less able to learn in general. This potential confound is something we address empirically, later in this chapter.

Another issue raised by the previous study that was unsatisfactorily resolved, is the issue of whether learning in Gf tasks is necessary for induction to take place in the more complex items. We had hypothesised that complex Raven items contain novel instantiations of rules, impossible to solve unless participants learn about these rules gradually through simpler items (“the General Learning Hypothesis”). While the study did not support this (specifically, we found that learning about the D2 rule from simpler items is not necessary for induction to take place in more complex versions of D2 items), we were unable to rule out that participants may rely on other forms of learning to be able to solve the more complex D2 items. Although we varied the amount of exposure to the D2 rule based on condition, each participant received the same number of items leading up to the transfer items. It is possible that participants may have been able to learn enough from these items to assist them with the transfer items, regardless of condition. For example, the simpler items in all of the conditions
contained the “distribution of three rule”. This may have helped participants learn about complex instantiations of the distribution of three rule, which may have been sufficiently similar to complex instantiations of the D2 rule. This is only a hypothetical example – we do not know what or even if participants were able to learn from items that did not contain the D2 rule to help them solve complex D2 rule items. We are merely putting forward that it is possible that participants may learn something from simpler items, which they then use to solve more complex items and what they learn may not be restricted to specific rules. One way to check this would be to include a condition where no “learning items” precede the transfer items at all. Only then would we be able to establish whether learning across items is necessary for induction in complex Gf items to take place. That is, if no learning items precede the transfer items and participants are unable to solve the transfer items, then we could conclude that learning within the task is necessary for induction to take place in complex Gf items. Conversely, if participants are able to successfully solve the items, then we could conclude that learning across items is not necessary for induction to take place. This will subsequently be examined further in the current chapter.

Lastly, in Experiment 1 we noted that when the General Learning Hypothesis is not supported it can be hard to empirically distinguish between the alternative explanations. That is, if learning across the task does not occur, what processes would be involved in an item instead? Would they be knowledge application processes (Holland et al., 1989; Sternberg, 1986) or rule discovery processes (induction independent of knowledge) (Cattell, 1987; Mackintosh, 1998; Spearman, 1932)? The next section explores these two competing hypotheses in more detail. The possibility that all three processes (knowledge application processes, learning and induction independent of knowledge) can (and are likely to) occur in Gf tasks will also be explored in further detail in later sections.
5.3 Within-Item Reasoning

In this section, we would like to make explicit what processes might be involved in reasoning in a Gf item, if no learning is involved. Our central hypothesis in this thesis is that learning is involved in Gf tasks and it may contribute to individual differences. Furthermore, we think learning processes mediate the knowledge that is hypothesised by Sternberg (1986) and Holland et al. (1989) to be required for induction and the novelty that is assumed to be characteristic of Gf tasks (Cattell, 1987; Gottfredson, 1997; Spearman, 1932) (discussed in the previous chapters). However, we would also like to explore the alternative possibilities in more detail, because often competing ideas in psychology turn out to be not mutually exclusive but compatible.

In the event that no learning is involved in Gf tasks, it has been put forward that reasoning may involve participants coming to the test situation with conceptual knowledge of the rules (Holland et al., 1989; Sternberg, 1986). Alternatively, they may discover the rules during the reasoning process within the item (Mackintosh, 1998; Spearman, 1932). According to insights from the analogical reasoning literature (a cognitive literature), both can occur in reasoning.

Although much of the research in analogical thinking involves analogical problem solving using stories, the research may also help us to better understand processes involved in Gf tasks. “Analogical thinking” is used to describe thinking that involves the transfer of knowledge from one situation to another by a process of mapping, for the purposes of better understanding the less familiar situation (Gick & Holyoak, 1983; Holland et al., 1989). Mapping is finding a set of one-to-one correspondence (often incomplete) between aspects of one body of information and aspects of another. It involves comparison of two available
concepts ("analogues"), which according to Gick and Holyoak (1983), can vary in their levels of abstraction. They give the following three examples:

- Comparison of two analogues at the same level of abstraction e.g., the heart and a water pump.
- Mapping of a concrete analogue to a more general concept e.g., the heart and the abstract concept of "pump".
- Mapping between two analogues which results in the acquisition of general concepts from the concrete examples e.g., learning the abstract sense of "pump" by comparing hearts and water pumps.

Analogical thinking may operate in Gf tasks in two ways. One way is through the mapping of a concrete analogue to a more general, known concept. This is similar to the process of rule (i.e., knowledge) application described by (Sternberg, 1986). Sternberg considers verbal analogies such as,

\[
\text{LAWYER is to CLIENT, as DOCTOR is to _____ ?}
\]

The solution is "patient". According to Sternberg inferential rules are required and in this case are declarative in nature – knowledge about functional relations. One has to relate the new stimulus information (in this case, the word "lawyer"), to old information already in memory such as the semantic attributes or properties of the word lawyer. In this case, the relevant property is the function that lawyers perform. Lawyers provide professional services to client, in much the same way that a doctor renders a professional service to a patient.

Knowledge about these functional relations is general, conceptual knowledge. In this example, one needs to map the concrete examples of "lawyer and client" to the general
concept of “service provider and service receiver”, and then to map this to “doctor” to arrive at the solution of “patient”. Thus, one arrives at the solution of “patient” by accessing the relevant general concept of “service provider and service receiver”, which is activated by the stimulus “doctor and patient”. However, if one did not possess this general, conceptual knowledge of the functional relation, “service provider and service receiver”, one would not be able to solve the induction problem. Similarly, for Raven items, if one does not possess knowledge of the various rules used in Raven including complex versions of the rules (where multiple instantiations are involved), one may not be able to solve the items.

Another way that analogical thinking may operate in Gf tasks is through *mapping between two concrete analogues which results in the acquisition of general concepts from the concrete examples* (the example we gave earlier was “learning the abstract sense of “pump” by comparing hearts and water pumps”). Gick and Holyoak (1983) found that if two concrete analogues are presented, participants could map them together to learn about a more general concept. They also found that two analogues was the minimum number of analogues required for participants to do this. That is, one example is not enough for a person to learn about a general concept. Things could operate in a very similar way in Gf tasks. For example, in Raven items, the first two rows of the matrix serve as two potential concrete analogues. Potentially, participants could map them together to learn about the general relationships (rules) that govern the shapes and figures in the cells and then map this onto the third row – the row that requires the solution. Thus, if there is an absence of any learning across items, analogical thinking based on the stimulus within items might be involved. This would be consistent with the classical conceptualisation of Gf which assumes participants discover the rules during the reasoning process, within the item (Mackintosh, 1998; Spearman, 1932).

Thus, based on the analogical thinking literature, reasoning in Gf tasks could involve:
• The mapping of a concrete analogue to a more general, known concept or
• Mapping between two concrete analogues which results in the acquisition of general
concepts from the concrete examples.

These too different methods of mapping are equivalent to the knowledge application
processes described by Holland et al. (1989) and Sternberg (1986) and the rule discovery
processes (induction independent of knowledge) described by Cattell (1987), Mackintosh
(1998) and Spearman (1932).

In the case of verbal analogies, it is hard to refute that some use of prior knowledge (known
concepts) is involved. One needs to know about the characteristics of the words (objects)
included. Also, only one full analogue (e.g., “lawyer is to client”) is provided (not enough for
a person to learn about a general concept), thus they would have to refer to general concepts
they already know. Indeed, verbal analogies have been shown to load heavily on both Gf and
Gc. It is less certain whether prior knowledge is necessary for tasks such as Raven. It does
not load as highly on Gc and there are potentially two analogues within an item from which
participants could map together to discover the rules. Given that the spontaneous noticing of
an analogy between two superficially dissimilar situations has been shown to be a rare event
in the laboratory (Gick & Holyoak, 1983)\textsuperscript{4}, it is less certain whether people would be able to
apply general concepts they learnt in real life to tasks such as the Raven.

Nevertheless, if across-item learning is not involved in Gf tasks, theoretically, induction
could involve both the application of prior conceptual knowledge through the mapping of a
concrete analogue to a more general, known concept and the discovery of conceptual

\textsuperscript{4} Although, only mean performance (and not individual differences) were taken into account.
knowledge (e.g., rules) through mapping between two concrete analogues. This is something we will return to shortly.

5.3.1 Within-Item Rule Discovery and Across-Item Learning: Similarities and Differences

At this stage, one may question whether the process of mapping between two concrete analogues (which results in the learning of general concepts) within an item, might be the same as learning across items in the task? After all, they both sound like instances of learning. The distinction is that learning across the task would involve building upon what is already known (and hence, explicit and easy), to understand about something previously unknown and novel (something less explicit and harder). In contrast, the process of mapping between two concrete analogues would involve learning something previously unknown and less explicit, through noticing regularities/similarities between analogues available in an item. However, both are consistent with processes of analogical thinking described by Gick and Holyoak (1983) – both involve the transfer of knowledge from one situation to another by a process of mapping.

This may lead one to entertain the idea that if one is able to learn through analogical thinking within an item, it would be most likely that they would also be able to learn across items using analogical thinking. Nevertheless, this is an empirical question. It is also an empirical question if both or either process is a source of individual difference.

5.4 Application of Prior Knowledge, Across-Item Learning or Within-Item Rule Discovery: It’s Probably All Three?

Returning to the earlier question of whether the application of conceptual knowledge or discovery of rules operates in Gf tasks, we believe that the answer is “both”. Furthermore, we hypothesise that across-item learning also operates in Gf tasks.
Early, simpler items may involve the application of conceptual knowledge. That is, knowledge of rules learnt from previous situations is mapped on to early items in Gf tasks. We shall use the Series Completion task as an example. Here is a simple example item:

18, 27, 36, 45, ___ ?

Participants need to complete the series by figuring out the last number in the series. The correct answer in this series is “54”, because each number is larger than the preceding number by a value of 9.

Sternberg (1986) hypothesizes that one draws upon declarative rules such as the set of possible relations that can be used to solve the items. One possible relation often used in Series Completion is the relation of “ascending sequences”. Holland et al. (1989) concur and according to them,

“... such categorisation derive from our background knowledge about linear orderings, which spans domains as diverse as rooms along a corridor, days of the week, notes in a musical scale etc. The categories people will bring to bear ...will be those that have emerged as useful general categories for dealing with linear structures in a number of experiences. We understand apparently unfamiliar situations so readily because they fit naturally into a pre-existing default hierarchy of relevant [rules and relations]” (p. 304).
However, as one progresses through the task and the task complexity and hence, novelty increases\(^5\), we hypothesized that one must rely on analogical thinking both within and across items to reduce the novelty.

We hypothesise that those with higher Gf are able to more quickly learn novel rules, through *analogical thinking across items*. We shall call this the “Individual Differences Learning Hypothesis”. We also hypothesize that those with higher Gf are able to more quickly learn novel rules *within an item* (through analogical mappings between the concrete stimuli in the item - to acquire new rules). We shall call this the “Classical Gf Hypothesis”.

Analogical thinking within an item should be substantially more difficult than analogical thinking across items. According to Holland et al. (1989) the acquisition of new task-specific knowledge can only be understood in the light of knowledge already possessed by the system. Some concepts will be more quickly acquired. They believe that concepts that are not intrinsically similar to those the learner already possesses are those that are hardest to learn. This would mean that in the absence of learning opportunity across the task, when one comes to a novel item, one would only have recourse to mappings between stimuli in the item to discover unfamiliar rules which should make the item substantially more difficult.\(^6\) How much more difficult is an issue we would like to empirically address later in the chapter.

The argument that in the absence of across-item learning opportunity, unfamiliar rules would be difficult to acquire (through mapping of within-item stimuli), provides us with a potential

---

\(^5\) The assumption here is that *complex* items are also *novel* items and they are novel because they contain multiple instances of various rules, which by themselves may not be novel, but when combined creates a novel product.

\(^6\) In Experiment 1, Chapter 4 we emphasised the possibility that participants would have to rely on across-item learning to reduce the novelty of complex items. The possibility that participants have recourse to mappings between stimuli within items to learn the rules in complex items, was somewhat neglected (although it was an implicit assumptions made by the “Classical Gf Hypothesis”). Thus, we give more attention to it in the current study.
way to tell whether reasoning in complex Gf items involves induction that relies on prior conceptual knowledge or induction that uses mapping between item stimuli to discover rules. Induction that results from prior conceptual knowledge should be much easier than induction that requires rule discovery. Again, this is an empirical issue we shall address later in the chapter.

5.5 The Cost of Learning: The Proactive Interference Hypothesis

As mentioned previously, we hypothesize that those with higher Gf are able to more quickly learn novel rules, through analogical thinking across items. Learning may assist with reasoning, but learning comes at the cost of proactive interference. Items within Gf tasks contain similarities which may assist with later items, but since no two items should be the same, they also contain differences which may result in distraction and interference.

In contrast to the Individual Differences Learning Hypothesis mentioned thus far in this chapter, Unsworth and Engle (2005a) argue that what is important to individual differences in performance in Gf tasks is the control of attention, especially under conditions of distraction and interference, such as when similar items (for example, similar surface features) are presented consecutively and result in proactive interference. We shall call this “Individual Differences Proactive Interference Hypothesis”.

In a review of the literature Heitz, Unsworth and Engle (2005) define attentional control as the “...voluntary, effortful cognitive act that serves to maintain information through activation of relevant brain circuitry, inhibit the irrelevant and distracting information that impinges on us at any one time, and suppress prepotent response tendencies that are task irrelevant” (p.63).
While Unsworth and Engle (2005a) put forward their Individual Differences Proactive Interference Hypothesis as a competitor for the Individual Differences Learning Hypothesis, controlled attention may also be required for learning, since controlled attention is also hypothesized to be needed to maintain information through activation of relevant brain circuitry (Heitz et al., 2005). Evidence that the two hypotheses (if true) may not be mutually exclusive comes from Sweller and Gee (1978) from the cognitive literature, to which we now turn.

5.6 Learning and Proactive Interference: Two Sides of the Same Coin?

Sweller and Gee (1978) empirically showed that “Einstellung” (what we have been referring to as proactive interference) and “the sequence effect” (what we have been referring to as across-item-learning) are two aspects of the same phenomenon. Einstellung was first demonstrated by Luchins (1942) and occurs when subjects given a series of similar problems to solve have trouble finding a solution to a different but simple, subsequent problem. The sequence effect (first demonstrated by Hull, 1920) occurs if subjects are able to solve a series of related problems graded according to difficulty more rapidly when the problems are presented in easy-to-hard sequence, rather than hard-to-easy. This phenomenon only occurs when problems are solved by the application of a rule (Sweller & Gee, 1978).

Within a single task, Sweller and Gee (1978) found results consistent with both the General Learning Hypothesis and a general form of the proactive interference hypothesis (that is, when individual differences aren’t taken into account). Consistent with the General Learning Hypothesis, Sweller and Gee found that solving easier items helps you to solve a harder item that is similar to the easier items (the sequence effect). Consistent with the general form of the proactive interference hypothesis, Sweller and Gee found that solving more items of that certain rule type makes it harder to solve an item of a different rule type later on.
A summary of Sweller and Gee’s explanation is that participants solve problems by testing a series of hypotheses. These hypotheses are sampled from a hypothesis domain (a set of related hypotheses). When subjects solve a series of problems that they perceive as being related, they begin each problem by testing hypotheses as closely related as practicable to their previously correct hypotheses. Einstellung results if a new problem solution requires a different rule – that is, a hypothesis from a different domain.

Sweller and Gee (1978) examined the sequence effect and Einstellung in the one experiment but did not examine the individual differences question. When taken together, the Individual Differences Learning Hypothesis and the Individual Differences Proactive Interference Hypothesis (which are about individual differences) and Sweller and Gee’s (1978) theory (an experimental theory that does not take into account individual differences), result in conflicting predictions. According to the Individual Differences Learning Hypothesis, those with higher Gf learn more solution principles. According to Sweller and Gee (1978)’s findings, the more items you can solve, the more you will suffer from proactive interference when you come to items with different solution principles. Yet, according to Unsworth and Engle’s (2005a) Individual Differences Proactive Interference Hypothesis, those with higher Gf suffer less from proactive interference. That is, these three findings cannot all be correct, without some further qualification. Hence, one of the aims of the current study is to address these issues within a combined individual differences/experimental framework. However, the obvious candidate qualification hypothesis would be that those with higher Gf would be both better at learning across the task and combating proactive interference.
5.7 Summary of Aims and General Hypotheses

One aim of the current chapter is to address issues raised but unresolved by Experiment 1, Chapter 4. Thus, this chapter further examines the distinction and relationship between:

- Induction involving learning that may occur across Gf items, within a task (“across-item learning”),
- Induction involving conceptual knowledge that may be brought to the task and
- Induction within an item that may be independent of any prior knowledge or learning that may occur across the items.

Specifically, this chapter examines whether these processes occur in Gf tasks, whether it is necessary for them to occur for induction to take place, whether they contribute to individual differences in performance and how they might be related.

The analogical thinking literature suggests that in the absence of across-item learning, within-item reasoning may involve conceptual knowledge that may be brought to the task through the mapping of a concrete analogue (i.e., stimuli in the item) to a more general, known concept. In the event that the concept is unknown to the participant, within-item reasoning could occur independent of knowledge through the mapping between the concrete stimuli in the item which results in the acquisition of the relevant general concepts (such as the rules in the item).

The insight gained from referring to the analogical thinking literature is that within-item reasoning (which involves a process of mapping between two concrete analogues and results in the learning of general concepts) in the absence of across-item learning and prior knowledge, is itself, a process of learning which shares similarities with across-item learning.
They are both instances of analogical thinking (involve the transfer of knowledge from one
situation to another by a process of mapping). However, concepts that are not intrinsically
similar to those the participant already possesses are those that are hardest to learn. This
would mean that in the absence of learning opportunity across the task and prior conceptual
knowledge, when one comes to a novel item, the process of mapping between stimuli in the
item to learn the rules would be substantially more difficult.

We suspect that early, simpler items involve the application of conceptual knowledge. However, as one progresses through the task and the task complexity and hence, novelty
increases, we hypothesize that one must rely on analogical thinking both within and across
items to reduce the novelty. In the absence of across-item learning opportunity, induction
would still be theoretically possible, but substantially harder than when across-item learning
opportunity is available. Thus, we make the following general complementary hypotheses:

1) **The Classical Gf Hypothesis** – In the absence of across-item learning and prior
contceptual knowledge, induction on complex Gf items would still be possible through
rule discovery processes (that is, through the mapping between the concrete stimuli in
the item which results in the acquisition of general concepts such as the rules in the
item). Also, those of higher Gf would outperform those of lower Gf (because they
would be better at making the mentioned mappings). However, we suspect
performance would be quite low for all (because according to analogical thinking
theory, acquiring unfamiliar concepts is very difficult). If performance is not low –
this would imply that the use of prior conceptual knowledge is involved instead.

2) **The General Learning Hypothesis** – Simpler items act as across-item learning
opportunity for more complex items with similar rules. When this across-item
learning opportunity is provided, performance will improve greatly for all, compared to when learning opportunity is not provided.

3) **The Individual Differences Learning Hypothesis** – When across-item learning opportunity is provided, those of higher Gf ability will improve more than those of lower Gf ability, due to those of higher Gf ability being better learners.

Another aim of this chapter is to examine learning’s relationship with proactive interference in Gf tasks. Learning has been shown to produce proactive interference in cognitive tasks. We predict that those of higher Gf will learn more than those of lower Gf. Hence, those of higher Gf would also need to be better at combating proactive interference. Thus, we make the following hypotheses:

4) **The General Proactive Interference Hypothesis** – Learning in Gf tasks produces proactive interference for items that contain dissimilar rules.

5) **The Individual Differences Proactive Interference Hypothesis** – those of higher Gf are better at combating proactive interference than those of lower Gf.

5.8 **Teasing Apart Prior Knowledge, Across-Item Learning and Within-Item Rule Discovery: The “Modified Sweller & Gee Task”**

A modified version of a task used in Experiment 2 of Sweller and Gee (1978) is the focus of this study. It has characteristics which make it similar to conventional Gf tasks, but it also has special characteristics which we think allow us to assess within-item rule discovery, across-item learning and to some extents use of prior knowledge, without one confounding the other. The task can also allow us to assess proactive interference.
Like conventional Gf tasks, the “Modified Sweller & Gee” (MSG) task is an inductive reasoning task that contains problems in an easy to hard order, involving related rules but its final item is solvable by a relatively simple rule, unrelated to those previously applicable to the earlier items. It is similar to the Number Series task - participants must work out the next number in the series.

However, unlike conventional Gf tasks, the MSG task is able to assess within-item induction in isolation from any potential influences from across-item learning, because a single item can be administered in isolation from other items. This is possible because each item allows multiple attempts within each item with feedback. The feedback provided is information about the correct answer via extension of the series (adding more analogues), which should have the effect of making the task increasingly easy. This is based on the idea that it gets progressively easier to learn about general concepts from multiple analogues (Gick & Holyoak, 1983; Holland et al., 1989). Normally, a single Gf item should not be used as a measure by itself because it would not be able to reliably estimate a person’s ability. However, allowing participants multiple attempts at the item should circumvent this.

In each item, subjects are presented with a target number. They then have to respond with what they think might be the “answer” to that target, which should be another number. They are then told the correct answer for that target. Participants get to see a cumulative list (a series) of target-correct answer pairs. Based on this, they have to work out the rules that govern what is considered a correct answer. So initially, participants will have to guess the correct answer. But as the list builds up, enough information is provided to induce the correct answer. An example item is presented Figure 5.1.
Figure 5.1. Screen shot of an example item from the instructions that participants get for the task. In this example, participants would have seen the targets: 14, 32, 54, 44, 15, 25 and currently 22. The correct answer to the current target would be "0" because in this example, the correct answer to the target is governed by the rule: answer to target equals the larger digit in the target minus the smaller digit in the target. Participants enter their guess/prediction of the answer in the “Your answer” box and then click the “Check Answer” button with their mouse. The target and its correct answer then appear under the “Record of correct Target-Answer pairs”.

A participant’s ability to solve a complex item without any across-item learning (that is, their ability to discover a rule within an item) could be assessed by providing them with only one single, complex item (we shall call this the “Complex Item”). If a participant is not a particularly strong reasoner, the task will be able to give an indication of this – they would need several attempts to get the correct answer. Furthermore, if prior conceptual knowledge of rules is involved, instead of rule discovery, the item should be solved relatively quickly with few attempts – that is, with few analogues required.
To assess across-item learning, participants could be provided with a series of items in easy-to-hard order. The amount of learning achieved could be assessed by comparing their performance on the harder transfer item (such as the “Complex Item”), with that of a control group with similar Gf-ability participants (defined by a separate Gf marker) that did not receive any easier items at all.

Experiment 1, Chapter 4 highlighted that the examinations of across-item learning may be confounded with within-item induction, since it may not be possible to learn from an item unless you get it right. That is, there is a possibility that those of lower Gf ability have been shown to not learn as much as those of higher ability because they are not able to get as many earlier items correct, rather than because they are less able to learn in general.

In the MSG task, if a participant is not a particularly strong reasoner, the task will be able to give an indication of this – they would need several attempts to get the correct answer. However, they should be able to get the correct answer eventually (because they get multiple attempts at an item and some feedback); and thus, glean some knowledge for future items. That is, even if they are weak at reasoning within items (but not a weak learner across items), they should be able to learn enough from previous items to require fewer attempts on a harder transfer item, compared to if they hadn’t received any learning opportunity at all (indicated by a Gf ability-matched control group who receive only the transfer item). In contrast, if a participant is a poor learner, they would not be able to improve much, relative to their Gf ability-matched control group. Put another way, they may not be able to improve much on the transfer item, even after being given actual learning opportunity. Hence, it is in this way that the MSG task may be able to assess the extent to which learning across items and within-item rule discovery are independent of each other. That is, the complex transfer item should be solved more slowly by the control group, compared to the experimental group, giving an
indication of the amount of learning. Thus, the MSG task should give an idea of participants’ amount of learning across items, while not completely confounding it with whether they are able to solve items in the first attempt – unlike conventional Gf tasks.

The task can also be used to explore the effect of proactive interference (PI). The effect of PI can be measured by the final item in the task which has a very different rule to the other items experienced (we shall call it the “PI Item”). Performance on the PI Item can then be compared to performance on the same item from a control group that only receives the PI item. The PI item should be solved more rapidly by the PI control group than the experimental group – giving an indication of PI.

5.9 A Methodological Issue: The Marker Task and Its Presentation Order

The Series Completion task was used as an independent marker task of Gf. A methodological issue is that Gf tasks are usually presented in easy-to-hard order, which may result in learning opportunity, which may be important to the Gf construct. A relevant question is: if Gf items are not presented in easy-to-hard order, would they still allow for learning opportunity? That is, does the presentation format of Gf tasks matter to the Gf construct?

Leary and Dorans (1985) undertook a literature review of articles that discuss item order effects in the educational testing literature. The effects of individual differences were not touched upon. However, there were some interesting findings. There was some variation in the literature, but overall, the following conclusions were made about aptitude tests. Firstly, Leary and Dorans concluded that randomly presented items were overall, not harder than items presented easy-to-hard. Secondly, hard-to-easy tasks were more difficult than easy-to-hard tasks, but this difference was only found when speeded tasks were used and not in power tasks. This led them to conclude that the apparent efficiency effect was due to
participants not reaching the end of hard-to-easy tasks because they may spend too long on
the hard items and do not have enough time for easier items that they would have got correct
with more time. The effect may also be due to difficult items early in the task causing anxiety
and discouragement, leading to poorer performance on later items. Thus, Leary and Dorans
did not make any speculations about difference in presentation format being due to learning
effects.

Leary and Dorans (1985) did not focus on Gf tasks. It is possible that presentation format of
Gf items does play a role in creating learning opportunity (Carlstedt et al., 2000; Lohman,
2001; Verguts & De Boeck, 2002b). Hence, we use multiple versions of Series Completion
tasks as Gf markers and varied their presentation format. We suspect that random and hard-
to-easy presentations minimise learning opportunity. Hence, we predict that they will have
weaker relationships with any learning effects that we find.

5.10 Experiment 2: A Pilot Study

Experiment 2 was a small pilot study that was conducted to better understand the MSG Task
and to see if it could be used to effectively test the hypotheses of interest, before resources
were put into using it in larger studies. The MSG Task had never been used before (in its
modified form).

In order to simultaneously assess the amount of learning and the PI in the one experiment,
one experimental group and two control groups were needed. The “Experimental Group” was
presented with a series of increasingly difficult items containing similar rules, up to the
penultimate item. In contrast, the final item contained a relatively simple rule, very dissimilar
to those previously applicable. One control group, the “PI Control Group” was presented with
the final item only (the PI Item); while the other control group, the “Complex Control Group”
was presented with the penultimate item only (the Complex Item). The dependent variable for the MSG Task is the number of attempts needed on an item before multiple correct responses are achieved.

5.10.1 Specific Predictions

Based on our general hypotheses in this chapter and the nature of the MSG task, we make the following specific predictions:

1) **The Classical Gf Hypothesis** – It is predicted that in the *Complex Control Group*, those of higher Gf will solve the Complex Item significantly faster (in terms of number of attempts) than those of lower Gf. That is, in the absence of across-item learning, induction would still be possible and those of higher Gf would outperform those of lower Gf. However, we suspect performance would be quite low (solution will require many attempts) for all (relative to the Experimental Group) because the item is assumed to be novel, containing novel concepts that are difficult to acquire. If performance is not low – this would imply that the use of prior conceptual knowledge is involved instead.

2) **The General Learning Hypothesis** – The Experimental Group should solve the Complex Item more rapidly than the Complex Control Group (demonstrating a learning effect). That is, simpler items earlier in the task should act as learning opportunity for more complex items with the same types of rules that appear later in the task. The availability of these simpler items for the Experimental Group (but not the Complex Control Group), should provide the group with the relevant conceptual knowledge that would make inducing the rules in the Complex Item easier (compared to those in the Complex Control Group).

3) **The Individual Differences Learning Hypothesis** – It is predicted that the difference in Complex Item performance between the Experimental Group and the Complex Control
Group, would be larger for those of higher Gf ability, due to them being able to learn more from the Experimental Condition.

4) **The General Proactive Interference Hypothesis** – It is predicted that the PI Item will be solved more rapidly by the PI Control Group than the Experimental Group and thus, demonstrating that learning across a task results in proactive interference for later items when the rules are dissimilar.

5) **The Individual Differences Proactive Interference Hypothesis** – It is predicted the difference between the Experimental Group and the PI Control Group on the PI item should be larger for those of lower Gf ability, due to them suffering more from proactive interference.

6) **A Hypothesis Regarding the Methodological Issue of the Presentation Format of the Gf Marker** - Gf tasks are usually presented in easy-to-hard order. It is possible that this presentation format plays a role in creating learning opportunity (Carlstedt et al., 2000; Lohman, 2001; Verguts & De Boeck, 2002b). Hence, multiple versions of Series Completion tasks were used as Gf markers and differed in presentation format. We suspect that random presentations minimise learning opportunity. Hence, we predict that they will have weaker relationships with any learning effects that we find.

**5.10.2 Method**

**Participants**

The students were enrolled in a first year undergraduate psychology at the University of Sydney and participated as part of their course work. In total, 99 students participated in the pilot study (65% female) with mean age = 19.28 years (SD = 3.41).
5.10.3 Cognitive Tasks

1) Modified Sweller & Gee (MSG) Task

This was an untimed, computerized task. It was based on a task used in Experiment 2 of Sweller and Gee (1978) – which was a cognitive task that was not designed to be a Gf task. Thus, some modifications had to be made to it to suit our purposes. For example, in the original tasks, items were presented to participants verbally and one item required participants to memorise parts of the item. We have changed such aspects to make it more similar to conventional Gf tasks. For example, we presented items visually and no memorisation of any of the items was required. Also, the rules in the original task were relatively simple because they were not intended to capture individual differences. Thus, our version of the task contained more complicated rules.

In each item in the MSG task, subjects are presented with a target number. They then have to respond with what they think might be the “answer” to that target, which should be another number. They are then told the correct answer for that target. Participants get to see a cumulative list (a series) of target-correct answer pairs. Based on this, they have to work out the rules that govern what is considered a correct answer. So initially, participants will have to guess the correct answer. But as the list builds up, enough information is provided to induce the correct answer (or not - depending on their “ability” on this measure). Participants have to enter 12 correct answers before the program moves onto the next item. A relatively large number of required, correct responses were chosen (12) to guard against accidental,
lucky responses that could terminate the item prematurely without providing a reliable estimate of the participant’s ability\textsuperscript{7}.

If participants take more than 50 attempts, the item also terminates. Scores represent total number of attempts (out of an allowable maximum of 50) minus the number of correct responses. Figure 5.1 depicted an example item.

There were 3 conditions. Participants received 1, 2 or 6 items, depending on which condition they were in. Table 5.1 presents the rules of all the items used based on condition. Actual items are presented in Appendix B.1.

Those in the Experimental Condition received all six items. Items 1 to 5 contained similar rules and increased progressively in complexity (with more complex items containing rules with more steps). Item 6 contained a different type of rule to the others and was a relatively simple item. Those in the PI Control Group only received item 6 (the PI item). Those in the Complex Control Group only received item 5 (the most difficult item - the Complex Item) and item 1 (the simplest item). Item 1 was included for this group to narrow down their hypothesis domain. According to Sweller and Gee (1978), there is a positive relationship between the complexity of a hypothesis (or complexity of the rule being tested) and the number of hypotheses or rules that are related to it. Subjects faced with complex problems sample hypotheses from a large/infinite hypothesis domain, resulting in long solution times or no solution. Hence, item 1 was included to narrow down the domain. However, item 1 was developed to be sufficiently different to the Complex Item and should not take too much away from its novelty.

\textsuperscript{7} Also, constraints in the computer programming language meant that the task could not be programmed to terminate after a string of correct responses in a row. Hence, a relatively large number was chosen to take into account the possibility of lucky guesses occurring throughout the item.
Table 5.1
MSG item rules based on the conditions in which they appeared.

<table>
<thead>
<tr>
<th>Condition</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5 (Complex Item)</th>
<th>6 (PI Item)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experimental</strong></td>
<td>* First digit of correct response identical to first digit of target number.</td>
<td>* First digit of correct response identical to first digit of target number.</td>
<td>* First digit of correct response identical to first digit of target number, but increases by 1 when 2nd digit of correct answer is '0'.</td>
<td>* First digit of correct response alternates between '6' and '1'.</td>
<td>* First digit of correct response alternates between '9', '1' and '5'.</td>
<td>* Add 2nd digit of target onto the total 2 digit number.</td>
</tr>
<tr>
<td></td>
<td>* Second digit of correct response always '8'.</td>
<td>* Second digit of correct response always '8'.</td>
<td>* 2nd digit of correct response alternates between '7' and '3'.</td>
<td>* Second digit of correct response identical to second digit of target, but decreases by 1 when first digit of correct response is '1'.</td>
<td>* Second digit of correct response identical to second digit of target, but doubled when first digit of correct response is '5'.</td>
<td></td>
</tr>
<tr>
<td><strong>Complex Control</strong></td>
<td>* First digit of correct response identical to first digit of target number.</td>
<td>* First digit of correct response identical to first digit of target number.</td>
<td>* First digit of correct response alternates between '9', '1' and '5'.</td>
<td>* Second digit of correct response identical to second digit of target, but doubled when first digit of correct response is '5'.</td>
<td>* Add 2nd digit of target onto the total 2 digit number.</td>
<td></td>
</tr>
</tbody>
</table>
2) **Number Series Random (NSR)**

This was basically the same task used in Experiment 1, Chapter 3. This is a well known Gf task (Marshalek et al., 1983; Quereshi, 2001). It was computer administered, with 25 items and a time limit of 5 minutes. Items were presented in a random sequence (rather than the traditional easy to hard sequence) and consisted of a series of numbers which were related by various rules. Participants need to complete the series by figuring out the last number in the series. Items were drawn from an item bank from the Personality and Individual Differences Lab, at the University of Sydney.

An example is:

12 15 12 18 12 21 12 __ ?

In this case, the first number is a constant repeated at an interval of two, and every second number has 3 added to it to gain the subsequent value. Thus, the correct answer is in this case is “24”.

3) **Letter Series Easy-to-Hard (LSEH)**

This is a well known Gf task (*Simon & Kotovsky, 1963*), very much like Number Series, but involving series of letters, related by various rules. Participants needed to complete the series by figuring out the last letter in the series. Items were drawn from an item bank from the Personality and Individual Differences Lab, at the University of Sydney. It was computer administered with 25 items and a time limit of 10 minutes. More time was allowed for this task than the NSR task because items were presented in the traditional easy-to-hard sequence. It was thought that more time would be needed for participants to reach the harder items at
the end of the task. This was to avoid the possibility of having to compare a measure that consisted of mainly hard items (NSR) and a measure that consisted of mainly easy items (LSEH), because participants did not have time to reach the harder items.

An example item is:

\[
\text{B D F H } \_? 
\]

The rule for this item is “add 2 to every letter to gain the subsequent letter in the series”. Hence, the correct answer in this example is “J”.

5.10.4 Procedure

All tasks were computer administered. An experimenter provided general instructions at the beginning of each session and was present throughout the session to assist participants and ensure that the task protocol was followed. A set of computerized (more specific) instructions preceded each task. Participants made their responses for all tasks using a standard keyboard and mouse. Participants were not allowed to use pen and paper for any of the items.

5.11 Results & Discussion

This section will be organised as follows. Firstly, descriptive statistics and some preliminary analyses will be presented. This will be followed by tests of differences between condition groups to examine the General Learning Hypothesis and the General Proactive Interference Hypothesis. We then move on to examinations of individual differences which start with some preliminary analyses of the different Gf markers (LSEH and NSR) and their relationship with the MSG task. This is then followed by significance tests of differences between groups based on condition and Gf ability groups (as defined by the Gf markers), to
examine the Classical Gf Hypothesis, the Individual Differences Learning Hypothesis, and the Individual Differences Proactive Interference Hypothesis.

5.11.1 Descriptive statistics

Demographic information is presented in Table 5.2. Descriptive statistics for marker tasks used in the study are presented separately by condition in Table 5.3. Cronbach’s alpha is in the acceptable range (for research purposes).

Table 5.2
Demographic information.

<table>
<thead>
<tr>
<th>Condition (n)</th>
<th>Age</th>
<th>Gender (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Experimental (46)</td>
<td>19.22</td>
<td>2.91</td>
</tr>
<tr>
<td>Complex Control (26)</td>
<td>19.77</td>
<td>4.98</td>
</tr>
<tr>
<td>PI Control (21)</td>
<td>18.71</td>
<td>1.82</td>
</tr>
<tr>
<td>Total (93)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.3
Descriptive statistics for Gf marker tasks.

<table>
<thead>
<tr>
<th>Task</th>
<th>Items</th>
<th>Complex Control</th>
<th>Experimental</th>
<th>PI Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (%)Correct</td>
<td>SD</td>
<td>n</td>
<td>α</td>
</tr>
<tr>
<td>LSEH</td>
<td>25</td>
<td>11.58 (46)</td>
<td>4.70</td>
<td>26</td>
</tr>
<tr>
<td>NSR</td>
<td>25</td>
<td>8.12 (41)</td>
<td>3.80</td>
<td>25</td>
</tr>
</tbody>
</table>

Note. α = Cronbach’s coefficient alpha.
Descriptive statistics for the MSG task are presented in Table 5.4, separately for each condition. The dependent variable for the MSG task is the number of attempts needed on an item before multiple correct responses are achieved (the “MSG Score”). Thus, lower scores indicate better performance.

Table 5.4
Descriptive statistics for scores on the Modified Sweller & Gee task.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Complex Control (n = 27)</th>
<th>Experimental (n = 46)</th>
<th>PI Control (n = 21)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item</td>
<td>Mean (SD)</td>
<td>Median (Quartiles)</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>1</td>
<td>4.39 (7.97)</td>
<td>2 (1,4)</td>
<td>4.29 (5.93)</td>
</tr>
<tr>
<td>2</td>
<td>7.10 (12.03)</td>
<td>3 (2.4)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>23.17 (19.98)</td>
<td>10 (5.47)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>24.45 (18.35)</td>
<td>12 (7.43)</td>
<td></td>
</tr>
<tr>
<td>Complex</td>
<td>44.17 (10.21)</td>
<td>48 (44.50)</td>
<td>27.46 (19.50)</td>
</tr>
<tr>
<td>PI</td>
<td>6.09 (8.60)</td>
<td>5 (2.6)</td>
<td>4.86 (9.43)</td>
</tr>
</tbody>
</table>

Note. Scores represent total number of attempts (out of an allowable 50) minus number of correct responses.
It is noted that the conditions have vastly unequal sample sizes. This is because participants were randomly assigned to conditions. Tasks were pre-setup in a testing lab with ten computers and one condition set up on each computer. Three computers were set up with the Complex Control Condition, three with the PI Control Condition and four with the Experimental Condition. Testing was conducted in groups of no more than 10 students and the fewest number of participants at any one time was one.

Distributions for many of the MSG items were not normal. This is consistent with the findings of Sweller and Gee (1978). This is not too surprising if one considers the nature of the items. Scores represent performance on a single item not a range of items with varying levels of difficulty. For example, very easy items (such as the PI item) had positively skewed distributions and very hard items (such as the Complex Item, in the Control Condition) had negatively skewed distributions. Some items also had bimodal distributions. Because of this lack of normality, Table 5.4 presents a variety of indicators of central tendency for the MSG task.

5.11.2 Learning and Proactive Interference at the Group Level

Preliminary Analysis

Since our aim will be to compare the three condition groups to make inferences about the nature of Gf after an experimental manipulation, it is important to know that the groups did not differ significantly on Gf at the outset. Due to unequal sample sizes in the three conditions, the Univariate General Linear Model (GLM) analysis of variance procedure was used with Type III sums of squares to see if groups differed on the Gf markers, LSEH and Number Series. The univariate analysis of variance showed that the groups did not differ
significantly on the LSEH, \( F (2, 88) = 1.32, p = .27 \); nor Number Series, \( F (2, 86) = .32, p = .72 \).

Due to the non-normality of the distribution of the MSG task, the non-parametric Mann-Whitney test was used to test for differences between groups. At the group level (before individual differences are taken into account) all indicators are in the expected direction and consistent with Sweller and Gee (1978). We shall elaborate below.

**The General Learning Hypothesis**

A Mann-Whitney U test indicated significantly fewer attempts by the Experimental group (who received all the MSG items) on the Complex item, compared to the Complex Control Group, \( U = 271.5, n_{\text{Control}} = 27, n_{\text{Experimental}} = 46, p > .01, \) two-tailed). This is consistent with the sequence effect found by Sweller and Gee (1978). The result also supports the General Learning Hypothesis. That is, it appears that the simpler items available to the Experimental Group act as learning opportunity for more complex items with the same types of rules that appear later in the task. The availability of these simpler items for the Experimental Group appears to have provided them with relevant conceptual knowledge that has made inducing the rules in the Complex Item easier for them (compared to those in Complex Control Group who were not provided with the simpler items).

**The General Proactive Interference Hypothesis**

The PI Control group solved the PI item significantly more rapidly than the Experimental group \( U = 260, n_{\text{Control}} = 18, n_{\text{Experimental}} = 45, p = .03, \) two-tailed). This is consistent with the Einstellung effect found by Sweller and Gee (1978). The result supports the General
Proactive Interference Hypothesis – that learning across a task results in proactive interference for later items when the rules are dissimilar.

5.11.3 Examinations of Individual Differences

A Methodological Issue Regarding Presentation Format of the Gf Marker

In order to examine individual differences in performance on the MSG task as a function of Gf ability, LSEH (Letter Series, presented easy-to-hard) and NSR (Number Series, presented randomly) were used as markers of Gf ability. However, it is possible that the presentation format of the markers affects the extent to which they tap into learning. We predict that random presentations will have weaker relationships with any learning effects we find in the MSG task. Thus, examinations of individual differences will employ the markers separately.

Preliminary Analysis: Correlations

While the markers correlated with each other highly, they were far from being perfectly correlated \( r (87) = .46, p < .01 \). Table 5.4 presents Spearman’s rho correlations for each of the MSG items with LSEH and NSR, separately for each condition. Spearman’s rho is a non-parametric rank correlation. In the Experimental condition, the increasing trend in correlations between LSEH (but not NSR) with MSG items as they get harder and further into the task, is preliminary evidence that learning occurs in markers presented in easy-to-hard format but not necessarily in markers presented in random format.

Interestingly, the Complex Control Item correlates more highly with LSEH than with NSR. In fact, the control item and NSR have no relationship. This could be viewed as surprising since both the Complex Control Item and NSR were expected to minimise across-item learning opportunity and it could be expected that they would have more in common with
each other. Yet, the Complex Control Item correlates quite highly with LSEH – which does allow for across-item learning opportunity.

<table>
<thead>
<tr>
<th>Gf Marker Task</th>
<th>MSG Item</th>
<th>Complex Control</th>
<th>Experimental</th>
<th>PI Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSEH</td>
<td>1</td>
<td>-.23</td>
<td>-.50**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-.02</td>
<td>.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-.25</td>
<td>.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-.42**</td>
<td>.</td>
<td></td>
</tr>
<tr>
<td>Complex</td>
<td></td>
<td>-.50**</td>
<td>.44**</td>
<td></td>
</tr>
<tr>
<td>PI</td>
<td></td>
<td></td>
<td>-.20</td>
<td>.19</td>
</tr>
<tr>
<td>NSR</td>
<td>1</td>
<td>.21</td>
<td>-.44**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-.04</td>
<td>.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-.00</td>
<td>.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-.20</td>
<td>.</td>
<td></td>
</tr>
<tr>
<td>Complex</td>
<td></td>
<td>.00</td>
<td>.15</td>
<td></td>
</tr>
<tr>
<td>PI</td>
<td></td>
<td></td>
<td>-.18</td>
<td>.011</td>
</tr>
</tbody>
</table>

*Note: *p < .05, **p < .01.
Pearson correlations between LSEH and NSR are: .51*, .54**, and .24 for Complex Control, Experimental and PI Control conditions respectively. Lower correlations in the PI Control condition may be due to small n.

### Creation of High and Low Gf Ability Groups Based on LSEH and NSR

LSEH and NSR were used to divide participants into low and high Gf groups. Those who scored in the bottom 50% (range: 2-11) of LSEH were placed in the “Low LSEH” ability group and those in the top 50% (range: 12-25) were placed in the “High LSEH” ability group.
The same was done for NSR. Those in the bottom 50% (range: 1-7) of NSR were placed in the “Low NSR” ability group and those in the top 50% (range: 8-16) were placed in the “High NSR” ability group. Because LSEH and NSR correlated quite highly, but not perfectly, those in the high (low) group for one marker were not exactly the same as those in the high (low) group in for the other marker. That is, there would not be too much redundancy in looking at the marker groups separately. A cross-tabulation of participants in each ability group is presented in Table 5.6.

<table>
<thead>
<tr>
<th></th>
<th>Low LSEH</th>
<th>High LSEH</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low NSR</td>
<td>27 (30%)</td>
<td>15 (17%)</td>
<td>42 (47%)</td>
</tr>
<tr>
<td>High NSR</td>
<td>19 (21%)</td>
<td>28 (31%)</td>
<td>47 (52%)</td>
</tr>
<tr>
<td>Total</td>
<td>46 (51%)</td>
<td>43 (48%)</td>
<td></td>
</tr>
</tbody>
</table>

*Note: Percentages are out of the total sample.*

We will subsequently be making numerous comparisons between the three conditions within the ability groups. Thus, it is important to know that the condition groups did not differ significantly on Gf at the outset. Due to unequal sample sizes in the three conditions, the Univariate General Linear Model (GLM) analysis of variance procedure was used with Type III sums of squares. Descriptive statistics and F-tests are presented in Table 5.7. No groups differed significantly on their relevant markers.
Table 5.7

Descriptive statistics and F-tests for LSEH (for Low and High LSEH groups) and NSR (for Low and High NSR groups), across the three conditions.

<table>
<thead>
<tr>
<th>Group</th>
<th>Complex Control</th>
<th>Experimental</th>
<th>PI Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (%Correct)</td>
<td>SD</td>
<td>n</td>
</tr>
<tr>
<td>Low LSEH</td>
<td>8.26 (33)</td>
<td>2.4</td>
<td>15</td>
</tr>
<tr>
<td>High LSEH</td>
<td>16.09 (64)</td>
<td>2.94</td>
<td>11</td>
</tr>
<tr>
<td>Low NSR</td>
<td>4.54 (18)</td>
<td>0.82</td>
<td>11</td>
</tr>
<tr>
<td>High NSR</td>
<td>10.92 (44)</td>
<td>2.52</td>
<td>14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>F (df1, df2)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low LSEH</td>
<td>1.92</td>
<td>.44</td>
</tr>
<tr>
<td>High LSEH</td>
<td>2.41</td>
<td>.73</td>
</tr>
<tr>
<td>Low NSR</td>
<td>0.12</td>
<td>.88</td>
</tr>
<tr>
<td>High NSR</td>
<td>0.88</td>
<td>.41</td>
</tr>
</tbody>
</table>
Further preliminary examinations: Across-Item Trends for the MSG task Based on LSEH and NSR Groups

Figure 5.2 illustrates trimean scores across the MSG task for LSEH groups in the Experimental condition. The trimean is a robust measure of central tendency; it is a weighted average of the 25th, 50th, and 75th percentiles. Specifically it is computed as follows: \[ \text{Trimean} = 0.25 \times 25\text{th} + 0.5 \times 50\text{th} + 0.25 \times 75\text{th}. \]

Those in the High LSEH group display what Sweller and Gee (1978) called a “learning-to-learn effect”. That is, despite the items increasing in complexity across the task, trimeans did not increase after item 3, which suggests that this group was learning consistently across the task. The Low LSEH group did not display this effect, suggesting that they did not learn as much as the High LSEH group across the task.

![Figure 5.2. Trimean MSG scores across the task for Experimental groups, by LSEH ability groups.](image)
Figure 5.3 illustrates trimeans across the MSG task for NSR groups in the Experimental condition. In contrast to the differences displayed by the High and Low LSEH groups, the pattern of performance for those in the High and Low NS groups are quite similar, again suggesting that random presentations of Gf markers do not tap into across-item learning as well as easy-to-hard presentations.

This preliminary examination suggests that when participants are classified into high and low Gf ability based on LSEH, there are larger differences between them on the MSG task (in the Experimental condition) than when they are classified into high and low Gf ability based on NSR. This is despite them being exactly the same individuals. The only difference between LSEH and NSR is their presentation format (easy-to-hard and random, respectively) and the type of stimuli involved in the items (letters and numbers respectively). We believe the
greater difference in MSG task performance for the LSEH groups is due to the presentation format of LSEH. That is, the easy-to-hard presentation in the LSEH Gf marker provides across-item learning opportunity. Those who are better at learning across the task do better on the task. Thus, this is reflected in their performance on the MSG task (in the Experimental condition) which was developed to provide learning opportunity across the task. In contrast, NSR does not appear as good at distinguishing between high and low performers on the MSG task, because the random presentation of its items does not seem to effectively tap into learning. In the next sections, we investigate this and other issues further with significance testing.

5.11.4 MSG Task and Significance Testing

In this section we explore the “Classical Gf Hypothesis”, the “Individual Differences Learning Hypothesis” and the “Individual Differences Proactive Interference Hypothesis” (in that order), separately for LSEH and NSR groups. Figure 5.4 illustrates the dependant variables using trimean scores for all LSEH Gf ability groups in all conditions. Figure 5.5 does the same for all NSR Gf ability groups and conditions. Descriptive statistics can be found in Table 5.8.
Figure 5.4. Trimean MSG score on Complex and PI items for Experimental and Control Conditions, by LSEH ability grouping.

Figure 5.5. Trimean MSG score on Complex And PI items for Experimental and Control Conditions, by NSR ability groups.
Table 5.8
Descriptive statistics and Mann-Whitney tests for MSG task scores across conditions, within Gf ability groups.

<table>
<thead>
<tr>
<th>Group</th>
<th>Item</th>
<th>Complex/PI Control Groups</th>
<th>Experimental Group</th>
<th>Mann-Whitney U</th>
<th>n₁</th>
<th>n₂</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean (SD)</td>
<td>Median (Quartiles)</td>
<td>Mean (SD)</td>
<td>Median (Quartiles)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low LSEH</td>
<td>Complex</td>
<td>48.17 (1.79)</td>
<td>48 (47, 49)</td>
<td>35.76 (17.42)</td>
<td>45 (10, 47)</td>
<td>48.00</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>PI</td>
<td>7.94 (13.81)</td>
<td>3 (1, 8)</td>
<td>8 (11.63)</td>
<td>5 (3, 7)</td>
<td>66.00</td>
<td>8</td>
</tr>
<tr>
<td>High LSEH</td>
<td>Complex</td>
<td>39.05 (14.58)</td>
<td>44 (40, 47)</td>
<td>18.62 (18.50)</td>
<td>9 (6, 45)</td>
<td>52.20</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>PI</td>
<td>2.40 (2.05)</td>
<td>2 (1, 4)</td>
<td>4.42 (2.18)</td>
<td>5 (2, 6)</td>
<td>55.50</td>
<td>10</td>
</tr>
<tr>
<td>Low NSR</td>
<td>Complex</td>
<td>44.09 (11.77)</td>
<td>47 (45, 48)</td>
<td>26.39 (19.53)</td>
<td>25 (6, 45)</td>
<td>42.50</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>PI</td>
<td>7.31 (13.94)</td>
<td>3 (1, 5)</td>
<td>6.70 (8.14)</td>
<td>5 (4, 7)</td>
<td>51.00</td>
<td>8</td>
</tr>
<tr>
<td>High NSR</td>
<td>Complex</td>
<td>44.11 (9.94)</td>
<td>47 (43, 49)</td>
<td>28.77 (20.32)</td>
<td>44 (7, 47)</td>
<td>87.00</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>PI</td>
<td>3.00 (2.78)</td>
<td>3 (1, 4)</td>
<td>5.70 (9.32)</td>
<td>3 (2, 5)</td>
<td>69.50</td>
<td>9</td>
</tr>
</tbody>
</table>

Note. n₁ = n for control group, n₂ = n for experimental group.
* p < .05, ** p < .01, two-tailed.
The Classical Gf Hypothesis - LSEH

The High LSEH group solved the Complex item significantly faster than the Low LS group, in the Control Condition (Mann–Whitney $U = 27$, $n_{\text{LowLS}} = 15$, $n_{\text{HighLS}} = 11$, $P > .01$ two-tailed). That is, in the absence of across-item learning opportunity, the High LSEH outperformed the Low LSEH group on the Complex Item. This is consistent with the Classical Gf Hypothesis - in the absence of across-item learning, induction in complex Gf items would still be possible (through the mapping between the concrete stimuli in the item which results in the acquisition of general concepts such as the rules in the item) and those of higher Gf would outperform those of lower Gf because they would be better at these within-item mappings.

Also, Figure 5.4 shows that there is a floor effect for the Complex Control Item for both groups. This suggests that without across-item learning opportunity, complex items are very difficult for all ability groups. According to the analogical thinking literature acquiring unfamiliar concepts (such as through mappings between the numbers in the item) is very difficult. Thus, it is likely that the Complex Control Item contained unfamiliar, novel concepts.

The Classical Gf Hypothesis - NSR

There was no significant difference between High and Low NSR groups on performance on the Complex Item in the Control Condition (Mann–Whitney $U = 71.5$, $n_{\text{LowNS}} = 11$, $n_{\text{HighNS}} = 14$, $p = .77$, two-tailed). That is, there appears to be no difference between the High and Low NSR groups when no across-item learning opportunity is provided. This suggests that random presentations of a Gf marker such as Number Series does not reliably tap individual differences in within-item induction processes.
There is a floor effect for the Complex Control Item, for both the High and Low NSR groups. Despite being able to do well on the NSR (which is a measure of Gf that minimises across-item learning opportunity), the High NSR group were not able to perform very well on the Complex Control Item when no across-item learning is available. Again, this suggests random presentations of Gf markers do not reliably tap within-item induction. This suggests that when across-item learning is minimised, task specific (rather than task general) processes become more important for performance.

**The Individual Differences Learning Hypothesis - LSEH**

Multiple Mann-Whitney tests were used to test for differences between the Experimental and Complex Control condition groups, within each ability group. The tests and their relevant descriptive statistics were presented in Table 5.8.

All LSEH ability groups demonstrated the sequence/learning effect; their respective Experimental groups solved the Complex Item significantly faster than their Complex Control groups (see Table 5.8). That is, both high and low Gf ability groups were able to learn within a task. This is consistent with the General Learning Hypothesis.

While the relevant tests in Table 5.8 indicates that all LSEH ability groups were able to significantly learn, the trend in Figure 5.4 is consistent with the hypothesis that those from the High LSEH group may have been more able to benefit from their learning opportunity than those from the Low LSEH group. There is apparently a larger gap between the High LSEH Control and Experimental groups (compared to the smaller gap between the Low LSEH Control and Experimental groups). Indeed, in the Experimental Condition, the High LSEH group solved the Complex Item significantly faster than those in the Low LSEH group. (Mann–Whitney $U = 142$, $n_{LowLS} = 23$, $n_{HighLS} = 21$, $P = .02$ two-tailed). This is
consistent with Individual Differences Learning Hypothesis - when learning opportunity is provided those of higher Gf ability are more able to benefit from it, resulting in apparently larger differences between high and low Gf ability groups. With such small sample sizes, this notion could not be investigated further with statistical tests in the current pilot study, but it is an issue we return to in the larger studies that follow.

**The Individual Differences Learning Hypothesis - NSR**

All NSR ability groups demonstrated the sequence/learning effect; their respective experimental groups solved the Complex item significantly faster than their Complex control groups (see Table 5.8). That is, both high and low Gf ability groups were able to learn within a task. This is consistent with the General Learning Hypothesis.

While the relevant tests in Table 5.8 indicates that all NSR ability groups were able to significantly learn, Figure 5.5 demonstrates a trend that is consistent with the possibility that those in the Low NSR group benefitted more from the learning opportunity than those in the High NSR group (due to the larger gap between their Experimental and Control Conditions). However, the difference between High and Low NSR on the Complex Experimental Item was not significant (Mann–Whitney $U = 212$, $n_{\text{LowNS}} = 22$, $n_{\text{HighNS}} = 22$, $p = .49$, two-tailed). This suggests that random presentations of a Gf marker such as Number Series do not reliably tap individual differences in across-item learning processes.

**Individual Differences Proactive Interference Hypothesis – LSEH and NSR**

Multiple Mann-Whitney tests were used to test for differences between the Experimental and PI Control condition groups, within each ability group. The tests and their relevant descriptive statistics were included in Table 5.8.
For the High LSEH group, the Control condition solved the PI item significantly faster than the Experimental condition. For the Low LSEH group, this was not significant. That is, those in the High LSEH group suffered from significant proactive interference but the Low LSEH group did not. This is contrary to the Individual Difference Proactive Interference hypothesis that those of lower Gf ability would suffer more proactive interference.

With regard to the NSR groups, those in the Low NSR grouping suffered from (marginally significant) proactive interference ($p = .09$). That is, for the Low NSR groups, the PI Control group solved the PI item with marginally significantly fewer attempts than the PI Experimental group. In contrast, there was no difference between PI Control and Experimental groups for those in the High NSR grouping (again, see Table 5.8).

Thus, only the High LSEH group and the Low NSR group showed evidence of suffering from proactive interference. These were also the groups that demonstrated the most amount of across-item learning. This is consistent with the General Proactive Interference Hypothesis – the more one learns, the more one suffers from proactive interference. However, since the High LSEH group suffered more from proactive interference than the Low LSEH group, the finding is inconsistent with the Individual Differences Hypothesis – that those of higher Gf are better at combating proactive interference.

### 5.12 Preliminary Conclusions

In summary, support was found for the Classical Gf Hypothesis, the General Learning Hypothesis, the Individual Differences Learning Hypothesis and the General Proactive Interference Hypothesis. No support was found for the Individual Differences Proactive Interference Hypothesis. Support was also found for our supplementary hypothesis that random presentations of Gf markers minimise across-item learning opportunity.
Although this is just a pilot study, some preliminary conclusions can be drawn. Firstly, the MSG task replicated the findings of Sweller and Gee (1978). That is, evidence was found to support the General Learning Hypothesis (a.k.a “the sequence effect”) and the General Proactive Interference Hypothesis (a.k.a “Einstellung”). Consistent with the General Learning Hypothesis, simpler items seemed to have acted as across-item learning opportunity for a more complex item with similar rules. When this across-item learning opportunity was provided, performance on the more complex item improved greatly for all, compared to when learning opportunity was not provided. Consistent with the General Proactive Interference Hypothesis the more that participants seem to learn, the more they seemed to suffer from proactive interference on an item of another rule type. Thus, the evidence also suggested that the General Learning Hypothesis and the General Proactive Interference Hypothesis are different facets of a joint phenomenon – which is also consistent with Sweller and Gee (1978).

Secondly, some support was also found for the Classic Gf Hypothesis. In the absence of across-item learning opportunity, those of higher Gf outperformed those of lower Gf on a complex Gf-type item. However, without learning opportunity available across the task, the complex item was extremely difficult (the Complex Control Item resulted in floor effects for all ability groups). There is evidence that in such a context, such items become extremely difficult not necessarily because they are complex, but because they are novel. Evidence that novelty contributes to extreme difficulty comes from the fact that when learning opportunity was provided (to reduce the novelty), the floor effect disappeared (despite the complexity of the item remaining the same).
Thirdly, the Individual Differences Learning Hypothesis received some support. When learning opportunity across the task was provided, all ability groups were able to significantly learn, but there was some evidence that those of higher Gf learn more than those of lower Gf.

A surprising finding was that the High LSEH group was the only group to suffer significant proactive interference. The Low NSR group suffered marginally significant proactive interference. Both groups seem to display the most amount of learning (see Figure 5.4 and Figure 5.5) and thus, is consistent with the General Proactive Interference Hypothesis. The finding that those of higher Gf (i.e., those in the High LSEH group) suffered proactive interference while those of lower Gf (i.e., those in the Low LSEH) did not, is inconsistent with the Individual Differences Proactive Interference hypothesis based on the theories of Unsworth and Engle’s (2005a), which predicts the reverse of what was found.

Thus, it seems that learning across the task is likely to contribute to individual differences in performance on Gf tasks; while combating proactive interference seems less likely to contribute to individual differences in performance. Learning across the task also appears necessary for induction to take place, since without it, the result is floor effects. That is, without some prior conceptual knowledge (provided through across-item learning), induction is unlikely to be able to take place. These preliminary conclusions only apply to Gf markers that have been presented in easy-to-hard format.

While the Number Series Task (presented randomly) correlated significantly with Letter Series Task (presented easy-to-hard), only the easy-to-hard marker was able to distinguish between high and low performers on the Complex Item when across-item learning was presented and when it was not present. Furthermore, the Complex Control Item correlated more highly with LSEH than with NSR. In fact, the Complex Control Item and NSR had no
relationship. This could be viewed as surprising since both the Complex Control Item and NSR were expected to minimise across-item learning opportunity and it could be expected that they would have more in common with each other. Yet, the Complex Control Item correlated quite highly with LSEH - which does allow for across-item learning opportunity. This suggests that when across-item learning is minimised, task specific (rather than task general) processes become more important for performance. Additionally, this may indicate that the within-item induction in the Complex Control Item shares some similarities to the across-item learning that may occur in LSEH, which would be consistent with theorising from the analogical thinking literature. The literature suggests that within-item reasoning (which involves a process of mapping between two concrete analogues and results in the learning of general concepts), in the absence of across-item learning and prior knowledge, is itself, a process of learning which shares similarities with across-item learning. This is because they both involve the transfer of knowledge from one situation to another by a process of mapping.

There were some limitations to this pilot study that will be addressed in the larger studies. Firstly, the difference in results between NSR and LSEH may be due to content effects rather than item presentation order, since presentation order was confounded with task content (numbers or letters). Secondly, random presentations of Gf markers still contain easier items dispersed throughout the task, which may allow for learning opportunity. In order to more effectively minimise learning opportunity, hard-to-easy presentations may be more appropriate.
CHAPTER 6

EXPERIMENT 3:
DIFFERENT BUT RELATED: WITHIN-ITEM RULE DISCOVERY, ACROSS-ITEM LEARNING, USE OF PRIOR KNOWLEDGE AND PROACTIVE INTERFERENCE

NUMBER SERIES

6.1 Experiment 3: The MSG Task and Number Series

Experiment 3 was a larger version of pilot Experiment 2 with only a few modifications. The experimental design was the same as Experiment 2 and we make the same hypotheses and predictions. Once again, the dependant variable was performance on the Complex and PI Items in the MSG task and the independent variables were condition (Experimental, Complex Control and PI Control) and Gf ability. Series Completion was once again used as the marker of Gf, but to control for content effects, only Number Series was used as the Gf marker (Letter Series will be used in Experiment 4).

The Gf marker was presented in two formats to all participants:

- Easy-to-hard (to allow for across-item learning opportunity),
- Hard and moderately hard items only, randomly inter-mixed (to minimise across-item learning opportunity).

Ideally, to minimise learning opportunity, the presentation format should be hard-to-easy. However, due to time constraints, only hard and moderately hard items were included.

A two-level Hierarchical Linear Modelling (HLM) was used to further understand the relationship between performance on the MSG task, condition and Gf ability. This was not possible in the pilot study due to the smaller sample size.
6.2 Method

Participants

In total, 208 participants were involved in the study. Participants were either enrolled in a second year undergraduate psychology course at the University of Sydney (72%) or from the general public and reimbursed for the time (28%). Out of the total, 65% were female. Participants’ mean age was 21.43 years (SD = 4.99).

Cognitive Tasks

1) Modified Sweller & Gee (MSG) Task

This was the same task presented in pilot Experiment 2 with no modifications. Identical to Experiment 2 and 3, there were three versions, one for each of the three conditions: the Experimental Condition, the Complex Control Condition and the PI Control Condition.

2) Number Series (Gf markers)

This was basically the same task as NSR presented in pilot Experiment 2. However, this time it was presented as two separate tasks, in two different formats:

- Easy-to-hard (“NSEH”) – hypothesized to allow learning opportunity,
- Only hard and moderately hard items, randomly intermixed (“NSH”) – hypothesized to minimise learning opportunity.

NSH always preceded NSEH to prevent the easy items from NSEH acting as learning items for NSH.
**Procedure**

Testing was conducted in groups of approximately 10 or 23 participants. All tasks were computer administrated. An experimenter provided general instructions at the beginning of each session and was present throughout the session to assist participants and ensure that the task protocol was followed. A set of computerized (more specific) instructions preceded each task. Participants made their responses for all tasks using a standard keyboard and mouse. Participants were not allowed to use pen and paper for any of the items. Tasks were always presented in the order MSG, NSH, NSEH, to prevent the Gf marker tasks from creating proactive interference for the MSG task.

Those in the Experimental and Complex Control conditions received only MSG, NSH and NSEH. However, those in the PI Control Condition received MSG, followed by a short version of Raven’s Advanced Progressive Matrices (Raven, 1962), followed by NSH then NSEH. The Raven’s task was meant to act only as a filler task. The MSG task in the PI Control Condition was very short, consisting of only one very easy MSG item. Hence, Raven’s was included to control for any fatigue that may have affected those in the Experimental and Complex Control conditions.

**Statistical Model**

HLM was used to model performance on MSG at each attempt on the Complex and PI items; to further our understanding of the relationship between performance on the MSG task, condition and NSEH. The equivalent model was not constructed for NSR due to restriction of range issues that will subsequently be discussed.
6.3 Results & Discussion

This section will be organised as follows. Firstly, we will present descriptive statistics and correlations. We then move on to non-parametric tests that compare performance on the Complex and PI items for control and experimental groups, based on Gf ability groups (created by NSEH). For these tests, dependant variable for the MSG Task is the number of attempts needed on an item before multiple correct responses are achieved. This gives an approximation of the point at which participants start to understand the rule in MSG Complex/PI items and get attempts correct. This was essentially the same type of analyses used in pilot Experiment 2 and the main aim of these analyses here will be to see whether the results from the pilot can be replicated.

We then move on to HLM analyses to gain a more complete picture of participants’ performance (which could not be done in the pilot study because of the small number of participants). The HLM allows us to model the probability that a participant will get an attempt in the Complex/PI items correct (for all attempts), based on the condition they are in and their performance on NSEH.

Non-parametric tests and HLM were not run for NSH due to floor effect and restriction of range issues for this marker. Instead, we visually examine the trimeans on the MSG task for those who scored in the top 5% on the NSH. The rationale for this was to see whether those who performed very well on a task such as NSH (where learning across the task is hypothesized to be minimized) could do equally well on the Complex Control item where learning is also minimized.
6.3.1 Descriptive Statistics

Demographic information is presented in Table 6.1. Descriptive statistics for Gf marker tasks used in the study are presented separately by condition in Table 6.2. Cronbach’s alpha is in the acceptable range (for research purposes). Means for NSH were very low (ranging from 17% - 25% correct across the three conditions), with lower standard deviation than the NSEH set. Hence, any direct comparison between the two sets based on methods which examine variance (such as correlations) will be made with caution.

<table>
<thead>
<tr>
<th>Condition (n)</th>
<th>Mean</th>
<th>SD</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental (70)</td>
<td>20.51</td>
<td>2.27</td>
<td>19 (27)</td>
<td>51 (73)</td>
</tr>
<tr>
<td>Complex Control (68)</td>
<td>22.40</td>
<td>7.46</td>
<td>29 (43)</td>
<td>39 (57)</td>
</tr>
<tr>
<td>PI Control (70)</td>
<td>21.63</td>
<td>4.36</td>
<td>27 (39)</td>
<td>43 (61)</td>
</tr>
<tr>
<td>Total (208)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Descriptive statistics for the MSG task are presented in Table 6.3 separately for each condition. Distributions for many of the MSG items were not normal. This is consistent with the findings of Sweller and Gee (1978) and the findings from Experiment 2, Chapter 5.

<table>
<thead>
<tr>
<th>Task</th>
<th>Items</th>
<th>Mean (%Correct)</th>
<th>SD</th>
<th>n</th>
<th>α</th>
<th>Mean (%Correct)</th>
<th>SD</th>
<th>n</th>
<th>α</th>
<th>Mean (%Correct)</th>
<th>SD</th>
<th>n</th>
<th>α</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSEH</td>
<td>20</td>
<td>14.73 (73)</td>
<td>3.98</td>
<td>67</td>
<td>.84</td>
<td>12.56 (63)</td>
<td>5.3</td>
<td>66</td>
<td>.90</td>
<td>12.31 (62)</td>
<td>4.36</td>
<td>70</td>
<td>.84</td>
</tr>
<tr>
<td>NSH</td>
<td>12</td>
<td>2.98 (25)</td>
<td>2.92</td>
<td>68</td>
<td>.86</td>
<td>2.67 (22)</td>
<td>2.73</td>
<td>70</td>
<td>.83</td>
<td>2.01 (17)</td>
<td>2.20</td>
<td>70</td>
<td>.81</td>
</tr>
</tbody>
</table>

Note: α = Cronbach’s coefficient alpha.
Table 6.3
Descriptive statistics for Modified Swaller & Gee task.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Complex Control (n = 68)</th>
<th>Experimental (n = 70)</th>
<th>PI Control (n = 70)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Median (Quartiles)</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>Item</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>6.10 (9.25)</td>
<td>3 (2, 6)</td>
<td>6.53 (10.99)</td>
</tr>
<tr>
<td>2</td>
<td>9.64 (14.19)</td>
<td>3 (2, 7)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>24.39 (19.79)</td>
<td>10 (5, 45)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>23.60 (18.65)</td>
<td>10 (6, 43)</td>
<td></td>
</tr>
<tr>
<td>Complex</td>
<td>40.03 (13.77)</td>
<td>46 (41, 48)</td>
<td>25.42 (19.38)</td>
</tr>
<tr>
<td>PI</td>
<td>7.57 (11.87)</td>
<td>4 (2, 6)</td>
<td>9.53 (15.22)</td>
</tr>
</tbody>
</table>

*Note.* Scores represent number of attempts (out of an allowable 50) minus number of correct responses.
6.3.2 Correlations

Table 6.4 presents Spearman’s rho correlations for each of the MSG items with NSEH and NSH, separately for each condition. Spearman’s rho is a non-parametric rank correlation.

The increasing trend in correlations between NSEH and MSG items as they increase in difficulty and appear further into the task (seen in Experiment 2) is repeated here for NSEH (-.08 to -.50). Also, as was the case in Experiment 2, the Complex Control Item correlates moderately with NSEH (r = -.35). This may indicate that the within-item induction in the Complex Control Item is similar in process to across-item learning tapped by the NSEH task. There is also a moderate correlation between NSH and the Complex Item in both the Experimental and Control conditions (-.24 and -.38, respectively), despite the floor effect in NSH. Also, NSH and NSEH correlate with each other (.40 to .48). That is, there appears to be some similarity between NSEH and NSH; or in other words, similarity between across-item learning and within-item induction.

The PI Item did not have any significant relationship with NSH in any condition. This may be due to restriction of range issues NSH. The PI Item correlates moderately with NSEH in its Control condition but not in the Experimental condition. This is preliminary evidence against the Individual Differences Proactive Interference Hypothesis - the hypothesis that those of higher Gf ability suffer less from proactive interference. If individual differences in Gf tasks is due to those of higher Gf ability being better at combating proactive interference than those of lower Gf, then the PI Item in the Experimental Condition (where there is more potential for proactive interference) should have correlated more highly with NSEH than the PI Item in the Control Condition.
Table 6.4
Spearman's rho correlations between Gf marker tasks and Modified Sweller & Goo items.

<table>
<thead>
<tr>
<th>Task</th>
<th>Item</th>
<th></th>
<th>Complex Control</th>
<th>Experimental</th>
<th>PI Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>n = 67</td>
<td>n = 66</td>
<td>n = 70</td>
<td></td>
</tr>
<tr>
<td>NSEH</td>
<td>1</td>
<td>-.11</td>
<td>-.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-.23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-.32**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-.10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Complex</td>
<td>-.35**</td>
<td>-.50**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PI</td>
<td></td>
<td>-.05</td>
<td></td>
<td>-.25*</td>
</tr>
<tr>
<td>NSH</td>
<td>1</td>
<td>-.08</td>
<td>-.25*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-.13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-.18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Complex</td>
<td>-.38**</td>
<td>-.24*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PI</td>
<td></td>
<td>.12</td>
<td></td>
<td>-.09</td>
</tr>
</tbody>
</table>

*Note.* *p* < .05, **p** < .01.

Pearson correlations between NSEH and NSH were:

.44**, .48** and .40** for Complex Control, Experimental and PI Control conditions respectively.
6.3.3 Non-parametric Comparisons of Gf Ability Groups (Based on NSEH)

Construction of NSEH Ability Groups

NSEH was used to divide participants into low and high Gf groups. Those who scored in the bottom 50% (score range: 1-13) were placed in the “Low NSEH” ability group and those in the top 50% (score range: 14-20) were placed in the “High NSEH” ability group.

Since our aim will be to compare the conditions within the ability groups, it is important to know that the condition groups did not differ significantly on Gf at the outset. Due to unequal sample sizes in the three conditions, the Univariate General Linear Model (GLM) analysis of variance procedure was used with Type III sums of squares. Descriptive statistics and F-tests are presented in Table 6.5. There were no significant differences across conditions for the High NSEH groups. However, the difference for Low NSEH was significant. Follow-up pair-wise comparisons with Bonferroni adjustments showed that the significant difference was due to those in Complex Control Condition scoring significantly higher on NSEH than those in the Experimental Condition. Thus differences in MSG performance for these groups should be interpreted with this in mind.
Table 6.5
Descriptive statistics and F-tests for Low NSEH and High NSEH groups. on NSEH across conditions.

<table>
<thead>
<tr>
<th>Group</th>
<th>Complex Control</th>
<th>Experimental</th>
<th>PI Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (%Correct)</td>
<td>SD</td>
<td>n</td>
</tr>
<tr>
<td>Low NSEH</td>
<td>10.20 (51)</td>
<td>2.22</td>
<td>24</td>
</tr>
<tr>
<td>High NSEH</td>
<td>17.25 (86)</td>
<td>2.00</td>
<td>43</td>
</tr>
</tbody>
</table>

Note. * p < .05.

Table 6.6
Descriptive statistics and Mann-Whitney tests for MSG task scores across conditions, within Gf ability groups.

<table>
<thead>
<tr>
<th>Group</th>
<th>Item</th>
<th>Control Groups</th>
<th>Experimental</th>
<th>Mann-Whitney U</th>
<th>n₁</th>
<th>n₂</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean (SD) Median (Quartiles)</td>
<td>Mean (SD) Median (Quartiles)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low NSEH</td>
<td>Complex</td>
<td>44.46 (9.36) 47 (43, 49)</td>
<td>30.45 (19.40) 41 (7, 47)</td>
<td>235.00</td>
<td>24</td>
<td>33</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>PI</td>
<td>13.06 (18.28) 3 (1, 23)</td>
<td>9.35 (14.89) 3 (2, 7)</td>
<td>660.00</td>
<td>41</td>
<td>33</td>
<td>.85</td>
</tr>
<tr>
<td>High NSEH</td>
<td>Complex</td>
<td>37.76 (15.23) 44 (40, 46)</td>
<td>13.42 (14.01) 8 (5, 10)</td>
<td>227.00</td>
<td>43</td>
<td>33</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>PI</td>
<td>7.22 (12.90) 3 (2, 5)</td>
<td>3.52 (2.34) 3 (2, 5)</td>
<td>474.00</td>
<td>29</td>
<td>33</td>
<td>.94</td>
</tr>
</tbody>
</table>

Note. n₁ = n for control group, n₂ = n for experimental group.
* p < .05, ** p < .01, two-tailed.
The Classical Gf Hypothesis - NSEH

The non-parametric Mann-Whitney test was used to test for differences between ability groups within the Complex Control condition. Descriptive statistics can be found in Table 6.6. Consistent with Experiment 2, in the Control Condition, the High NSEH group solved the Complex Item significantly faster than the Low NSEH group, ($U = 325$, $n_{\text{Low NSEH}} = 24$, $n_{\text{High NSEH}} = 43$, $p = .01$, two-tailed). This suggests that without across-item learning opportunity, the High NSEH group are better than the Low NSEH group at within-item induction. This is consistent with the Classical Gf Hypothesis - in the absence of across-item learning, induction in complex Gf items would still be possible (through the mapping between the concrete stimuli in the item which results in the acquisition of general concepts such as the rules in the item) and those of higher Gf would outperform those of lower Gf because they would be better at these within-item mappings.

However, there appears to be a floor effect for the Complex Control Item for both High and Low NSEH ability groups. This suggests that without across-item learning opportunity, complex, novel items are very difficult for all ability groups. This is consistent with the analogical thinking literature (Holland et al., 1989) which suggests that acquiring unfamiliar concepts through mappings between concrete analogues (in this case, the numbers in the item) is very difficult. This is also consistent with Experiment 2.

The General Learning Hypothesis and the Individual Differences Learning Hypothesis - NSEH

Multiple Mann-Whitney tests were used to test for differences between the Experimental and Complex Control condition groups, within each ability group. The tasks and their relevant
descriptive statistics are presented in Table 6.6. Figure 6.1 illustrate these effects graphically for the NSEH groups using trimeans.⁹

![Figure 6.1](image)

*Figure 6.1. Trimean MSG scores on Complex and PI items, for Experimental and Control Conditions, by NSEH ability groups.*

All ability groups demonstrated the sequence/learning effect - their respective Experimental groups solved the Complex Item significantly faster than their Complex Control groups. That is, both High and Low NSEH ability groups could learn within a task. This supports the General Learning Hypothesis - simpler items act as across-item learning opportunity for more complex items with similar rules, and when this learning opportunity is provided, performance will improve greatly for all, compared to when learning opportunity is not provided.

---

⁹ The trimean is a robust measure of central tendency; it is a weighted average of the 25th, 50th, and 75th percentiles. Specifically it is computed as follows: Trimean = 0.25 x 25th + 0.5 x 50th + 0.25 x 75th.
While all ability groups were able to learn, Figure 6.1 is consistent with the Individual Differences Learning Hypothesis that predicted that those from the High NSEH group would be more able to benefit from their learning opportunity than those from the Low NSEH group. That is, there is a larger gap between their Complex Control and Experimental groups (compared to the smaller gap between the Low NSEH Complex Control and Experimental groups). Certainly, in the Experimental Condition, the High NSEH group solved the Complex Item with significantly fewer attempts than those in the Low NSEH group, (Mann–Whitney $U = 274$, $n_{\text{LowNSEH}} = 33$, $n_{\text{HighNSEH}} = 33$, $p > .01$. two-tailed). We shall return to exploring this potential interaction effect with HLM, shortly.

Figure 6.2 illustrates trimeans across the MSG task for NSEH groups in the Experimental Condition. Those in the High NSEH group display the “learning-to-learn effect”. That is, despite the items increasing in complexity across the task, trimeans did not increase after item 3. The Low NSEH group also display this effect, but to a lesser extent.

![Figure 6.2. Trimean MSG scores across the task for Experimental groups, by NSEH ability groups.](image-url)
Thus far, there is some support for the Individual Differences Learning Hypothesis. The data are consistent with the idea that without across-item learning opportunity, complex, novel items are very difficult for all ability groups and that those of higher Gf ability are more able to benefit from learning opportunity when it is provided, resulting in larger differences between high and low Gf ability groups.

The General Proactive Interference and the Individual Differences Proactive Interference Hypothesis – NSEH

Multiple Mann-Whitney tests were used to test for differences between the Experimental and PI Control condition groups, within each ability group (see Table 6.6, PI Item). Surprisingly, there were no significant differences - no ability groups seemed to suffer from proactive interference. This is contrary to the Individual Differences Proactive Interference Hypothesis. (those of lower Gf ability would suffer more proactive interference) and contrary to the General Proactive Interference Hypothesis (that predicted that all ability groups would suffer from proactive interference to some extent).

6.3.4 Hierarchical Linear Modelling with NSEH

A two-level Hierarchical Linear Modelling (HLM) was used to further understand the relationship between performance on the MSG task, condition and NSEH. The probability of getting an attempt correct (where “Correct” = 1 if correct; 0 if incorrect)\textsuperscript{10} in the item was modelled with log-odds as the dependent variable,

\[
\text{Prob(Correct} = 1) = \phi_{ij} \quad (1)
\]

\textsuperscript{10} The MSG task was programmed to exit from an item after a total of 12 correct responses for that item. This would have resulted in a lot of missing data for the HLM. Hence, missing responses due to participants successfully exiting an item were assigned a “1” (i.e., scored as “correct”).
\[
\log(\frac{\phi_{ij}/(1-\phi_{ij})} = \eta_{ij}
\]

(2)

Thus, \(\eta\) is the log of the odds (log-odds) of success (in this case, Correct = 1). That is, if the probability of success, \(\phi\), is .5, the odds of success \(\phi/(1-\phi) = .5/.5 = 1.0\) and the log-odds would be \(\log(1) = 0\). That is, when the probability of success is less than .5, the odds are less than 1.0 and the log-odds is negative. When the probability is greater than .5, the odds are greater than 1.0 and the log-odds is positive. While \(\phi\) must be in the interval (0,1), \(\eta\) can be any real value.

The level-1 model is stated by equation (3). \textit{Attempt} is the attempt number in the item, centred at the grand-trimean number of attempts for all participants on the item (attempt number 35 for Complex Item and attempt number 4 for the PI item),

\[
\eta_{ij} = \pi_{0i} + \pi_{1i} \text{Attempt}_{ti}
\]

(3).

Equation (3) predicts \(\eta_{ij}\), the log-odds of success for participant \(i\) at time \(t\); based on \(\text{Attempt}_{ti}\). Thus, \(\pi_{0i}\) represents the log-odds of success of participant \(i\) when \(\text{Attempt}_{ti}\) is equal to 0 (where 0 = grand-trimean number of attempts for the item); and \(\pi_{1i}\) is the growth (improvement) rate for participant \(i\) over the item. Since this model uses Bernoulli sampling, there is no error term in equation (3).

The level-2 model predicts the coefficients in the level-1 model and is stated by equations (4) and (5). \textit{Condition} is the dummy variable for condition (0 = Control; 1 = Experimental) and \textit{NSEH\_Cent} is centred NSEH (0 = grand-mean).

\[
\pi_{0i} = \beta_{00} + \beta_{01} \text{Condition}_{1i} + \beta_{02} \text{NSEH\_Cent}_{2i} + \beta_{03} \text{Condition} \times \text{NSEH\_Cent}_{3i} + r_{0i}
\]

(4)

\[
\pi_{1i} = \beta_{10} + \beta_{11} \text{Condition}_{1i} + \beta_{12} \text{NSEH\_Cent}_{2i} + \beta_{13} \text{Condition} \times \text{NSEH\_Cent}_{3i}
\]

(5)
\( \beta_{00} \) is the value for \( \pi_{0i} \) when all of its predictors are 0 (that is, for someone in the Control condition, with grand-mean NSEH score). \( \beta_{10} \) is the growth rate (that is, the value for \( \pi_{1i} \)) when all of its predictors are 0 (that is, for someone in the Control condition, with grand-mean NSEH score). \( r_{0i} \) is the deviation of participant \( i \) from the trimean performance at attempt number 35 (for the Complex Item) or 4 (for the PI Item). Attempt 35 and 4 are the grand-trimean number of attempts for their respective items.

Thus, the intercept \( \pi_{0i} \) predicts the log-odds of success for participant \( i \), at the grand-trimean number of attempts based on the main effects of Condition, NSEH and their interactions. The Attempt slope \( \pi_{1i} \) predicts the growth rate for participant \( i \), over the course of the item, based on the main effects of Condition, NSEH and their interactions. HLM models were run separately for the Complex Item and the PI Item. We first present the analyses for the Complex Item.

**HLM Analyses of the Complex Item (the Classical Gf Hypothesis and the Learning Hypotheses)**

Table 6.7 presents the estimates of the HLM parameters, which were obtained in HLM 6 (Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2004) for the Complex Item. Intercept \( \pi_{0i} \) represents the log-odds of success at the grand-trimean number of attempts for the Complex Item, that is, attempt number 35. Slope \( \pi_{1i} \) represents the growth rate for participant \( i \) over the item, that is, changes to the log-odds of success as a function of attempt number.

Model 1 includes all predictors. Non-significant and weak predictors were dropped from the model based on two criteria. Predictors with t-ratios of smaller magnitude than +/- 1 were dropped (since the threshold for significance is +/- 2). Predictors with significant coefficients weaker than +/- 0.01 were also dropped (since log-odds of +/- 0.009 corresponds to an odds
The final results are presented in Table 6.7, Model 2.

Table 6.7

*Estimated parameters of the Hierarchical Linear Model for performance on the Complex Item.*

<table>
<thead>
<tr>
<th>Model</th>
<th>Fixed Effect</th>
<th>Coefficient (Log-odds)</th>
<th>Odds Ratio</th>
<th>SE</th>
<th>t-ratio</th>
<th>d.f</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>For Intercept, $\pi_0$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Intercept2, $\beta_{00}$</td>
<td>-2.409</td>
<td>0.09</td>
<td>0.312</td>
<td>-7.721</td>
<td>95</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>Condition, $\beta_{01}$</td>
<td>3.956</td>
<td>34.135</td>
<td>0.745</td>
<td>5.310</td>
<td>95</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>NSEH_Cent, $\beta_{02}$</td>
<td>0.171</td>
<td>1.178</td>
<td>0.083</td>
<td>2.060</td>
<td>95</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>Condition X NSEH_Cent, $\beta_{03}$</td>
<td>0.226</td>
<td>1.219</td>
<td>0.137</td>
<td>1.650</td>
<td>95</td>
<td>0.11</td>
</tr>
<tr>
<td>2</td>
<td>For Intercept, $\pi_0$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Intercept2, $\beta_{00}$</td>
<td>-2.366</td>
<td>0.097</td>
<td>0.307</td>
<td>-7.707</td>
<td>95</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>Condition, $\beta_{01}$</td>
<td>3.543</td>
<td>34.578</td>
<td>0.562</td>
<td>6.304</td>
<td>95</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>NSEH_Cent, $\beta_{02}$</td>
<td>0.149</td>
<td>1.160</td>
<td>0.066</td>
<td>2.258</td>
<td>95</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>Condition X NSEH_Cent, $\beta_{03}$</td>
<td>0.150</td>
<td>1.162</td>
<td>0.093</td>
<td>1.613</td>
<td>95</td>
<td>0.11</td>
</tr>
<tr>
<td>2</td>
<td>For Attempt Slope, $\pi_1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Intercept2, $\beta_{10}$</td>
<td>-0.008</td>
<td>0.991</td>
<td>0.009</td>
<td>-0.889</td>
<td>4973</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>Condition, $\beta_{11}$</td>
<td>0.091</td>
<td>1.095</td>
<td>0.020</td>
<td>4.550</td>
<td>4973</td>
<td>**</td>
</tr>
</tbody>
</table>

*Note. *$p < .05$. **$p < .01$.*

Intercept1, $\pi_0 = $ predicted log-odds of success ($Correct = 1$) on the Complex Item at attempt number 35.

Attempt Slope, $\pi_1 = $ predicted growth rate of success ($Correct = 1$) over the course of the Complex Item.
Performance at the Grand-Trimean Number of Attempts: The coefficients that predict $\pi_0$ indicate that at attempt number 35 (the average number of total number of attempts for all participants), the log-odds of success is significantly predicted by Condition and NSEH, and the interaction between Condition and NSEH is marginally significant. These effects are more clearly illustrated in Figure 6.3.

Thus, by attempt 35, regardless of NSEH ability, those in the Experimental Condition have a significantly higher probability of success than those in the Control Condition. This supports the General Learning Hypothesis.

The marginally significant interaction between Condition and NSEH by Attempt number 35 is illustrated by the larger gap between the broken lines - i.e., between the higher Gf (higher NSEH) Control and Experimental groups, compared to the smaller gap between the solid lines – i.e. the Control and Experimental groups for those of lower Gf. That is, there is some (marginally significant) evidence that those of higher Gf profited more from the opportunity to learn across items than those of lower Gf. This supports the Individual Differences Learning Hypothesis.

There appears to be a floor effect for the Complex Control Item for both groups. This suggests that without across-item learning opportunity, complex, novel items are very difficult for all ability groups. However, the main effect of NSEH (regardless of condition), suggests that those of higher Gf outperform those of lower Gf in terms of within-item induction, when no across-item learning opportunity is provided. On one level, this supports the Classical Gf Hypothesis. However, visual examination of Figure 6.3 shows a floor effect, which suggests that those in the Control Condition (even of higher Gf) were rarely able to reach the correct answer on any attempt – the predicted probability of success for those in the
The 75th percentile of NSEH (Gf ability), in the Complex Control condition never goes above 25%. Thus, the results do not totally support the Classical Gf Hypothesis because it appears that without some exposure to earlier, simpler items to provide participants with the relevant, related knowledge, induction is very difficult. That is, induction within a complex item does not occur in isolation from earlier items – it is not likely. We shall elaborate on this below in the “Growth of Success” result.

**Growth of Success:** The coefficients for \( \pi_1 \) indicate that Condition is the only significant predictor of the rate of growth of success over the course of the Complex Item. It can be seen in Figure 6.3 that only those in the Experimental Condition improved their performance with more attempts. Furthermore, the data in Figure 6.3 suggests that those in the Control Condition were rarely able to reach the correct answer on any attempt. The lack of significance of NSEH in predicting rate of growth suggests that Condition (i.e., exposure to earlier, simpler items) is more important than Gf ability in being able to induce the correct answer. That is, it appears that it is very difficult to make mappings between within-item analogues to induce the general concept (i.e., work out the rules by looking for generalisations among the item’s stimuli) in complex Gf items, unless one is provided with earlier, easier items to provide one with the relevant, prior knowledge. This seems to be the case, regardless of one’s Gf ability. This suggests that the knowledge provided by across-item learning is important to induction, for all levels of Gf ability.
Figure 6.3. Predicted probability of success (Correct = 1) as a function of Condition, NSEH and Attempt, on Complex Item, based on Model 1 from Table 6.7. Black line indicates differences in predicted success between those in the Experimental Condition and those in the Control Condition, separately for those of higher and lower Gf ability at Attempt number 35 (the grand trimean number of attempts).

HLM Analyses of PI Item (the Proactive Interference Hypotheses)

A similar two-level HLM was used to model probability of getting an attempt correct (where “Correct” = 1 if correct; 0 if incorrect) in the PI Item, with its log-odds as the dependent variable and Condition and NSEH as the independent variable. Table 6.8 presents the estimates of the HLM parameters which were obtained in HLM 6 (Raudenbush et al., 2004), for the PI Item. Like for the Complex Item, Intercept $\pi_0$ represents the log-odds of success at the grand-trimean total number of attempts for the PI Item, which in this case is attempt
number 4. Slope $\pi_{i}$ represents the growth rate for participant $i$ over the item, that is, changes to the log-odds of success as a function of attempt number. Model 1 includes all predictors. Non-significant and weak predictors were dropped from the model based on two criteria. Predictors with t-ratios of smaller magnitude than +/- 1 were dropped, as were predictors with significant coefficients weaker than +/- 0.01. The final results are presented in Table 6.8., Model 2.
Table 6.8

*Estimated parameters of the Hierarchical Linear Model for performance on the PI Item.*

<table>
<thead>
<tr>
<th>Model</th>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>Odds Ratio</th>
<th>SE</th>
<th>t-ratio</th>
<th>d.f</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(Log-odds)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>For Intercept1, $\pi_0$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Intercept2, $\beta_{00}$</td>
<td>0.189</td>
<td>1.208</td>
<td>0.224</td>
<td>0.844</td>
<td>93</td>
<td>.40</td>
</tr>
<tr>
<td></td>
<td>Condition, $\beta_{01}$</td>
<td>-0.704</td>
<td>0.494</td>
<td>0.314</td>
<td>-2.242</td>
<td>93</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>NSEH_Cent, $\beta_{02}$</td>
<td>0.019</td>
<td>1.020</td>
<td>0.055</td>
<td>0.345</td>
<td>93</td>
<td>.71</td>
</tr>
<tr>
<td></td>
<td>Condition X NSEH_Cent, $\beta_{03}$</td>
<td>0.040</td>
<td>1.041</td>
<td>0.076</td>
<td>0.526</td>
<td>93</td>
<td>.59</td>
</tr>
<tr>
<td></td>
<td>For Attempt Slope, $\pi_1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Intercept2, $\beta_{10}$</td>
<td>0.128</td>
<td>1.137</td>
<td>0.045</td>
<td>2.853</td>
<td>4824</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>Condition, $\beta_{11}$</td>
<td>0.193</td>
<td>1.213</td>
<td>0.074</td>
<td>2.592</td>
<td>4824</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>NSEH_Cent, $\beta_{12}$</td>
<td>0.017</td>
<td>1.017</td>
<td>0.006</td>
<td>2.615</td>
<td>4824</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>Condition X NSEH_Cent, $\beta_{13}$</td>
<td>0.007</td>
<td>1.007</td>
<td>0.009</td>
<td>0.801</td>
<td>4824</td>
<td>.42</td>
</tr>
</tbody>
</table>

2 For Intercept1, $\pi_0$

<table>
<thead>
<tr>
<th>Model</th>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>Odds Ratio</th>
<th>SE</th>
<th>t-ratio</th>
<th>d.f</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(Log-odds)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Intercept2, $\beta_{00}$</td>
<td>0.129</td>
<td>1.137</td>
<td>0.235</td>
<td>0.549</td>
<td>95</td>
<td>.58</td>
</tr>
<tr>
<td></td>
<td>Condition, $\beta_{01}$</td>
<td>-0.552</td>
<td>0.575</td>
<td>0.332</td>
<td>-1.563</td>
<td>95</td>
<td>.09</td>
</tr>
<tr>
<td></td>
<td>For Attempt Slope, $\pi_1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Intercept2, $\beta_{10}$</td>
<td>0.137</td>
<td>1.147</td>
<td>0.042</td>
<td>3.262</td>
<td>4827</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>Condition, $\beta_{11}$</td>
<td>0.151</td>
<td>1.163</td>
<td>0.057</td>
<td>2.649</td>
<td>4827</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>NSEH_Cent, $\beta_{12}$</td>
<td>0.019</td>
<td>1.019</td>
<td>0.005</td>
<td>3.800</td>
<td>4827</td>
<td>**</td>
</tr>
</tbody>
</table>

Note. *$p < .05$. **$p < .01$.

Intercept1, $\pi_0$ = predicted log-odds of success ($Correct = 1$) on the PI Item at attempt number 4

Attempt Slope, $\pi_1$ = predicted growth rate of success ($Correct = 1$), over the course of the PI Item.
Performance at the Grand-Trimean Number of Attempts: The coefficients for $\pi_0$ indicate that at attempt number 4, the log-odds of success is (marginally) significantly predicted by Condition. Figure 6.4 indicates that at Attempt number 4, regardless of Gf ability, those in the PI Control Condition have a (marginally significant) higher probability of success than those in the Experimental Condition. This is consistent with a proactive interference effect. That is, those who learn about one type of rule, find it harder to solve items of a different rule type later on. However, this effect is general and not predicted by Gf ability. This supports the General Proactive Interference Hypothesis, but not the Individual Differences Hypothesis.

Growth of Success: The coefficients for $\pi_1$ indicate that Condition and Gf ability are significant predictors of the rate of growth (improvement) of success, over the course of the PI Item. It can be seen in Figure 6.4 that on average, the red lines representing the Experimental condition are steeper than the blue lines that represent the PI control condition. That is, the significance of Condition as a predictor of growth of success suggests that while the Control condition did better than the Experimental condition at Attempt 4, the Experimental condition would have quickly caught up. That is, proactive interference appears to be temporary.

Also, the broken lines representing higher Gf ability are steeper than the solid lines representing lower ability. That is, the significance of Gf ability in predicting growth rate of success suggests that on this easy PI Item, those of higher ability improved their performance with more attempts, at a faster rate than those of lower Gf ability, regardless of condition. That is, in this case, they can induce the correct answer with more attempts, at a faster rate; that is, they are better at within-item induction.
Figure 6.4. predicted probability of success (\textit{Correct} = 1) as a function of \textit{Condition}, \textit{NSEH} and \textit{Attempt}, on Complex Item, based on Model 1 from Table 6.8. Black line indicates differences in predicted success between those in the Experimental Condition and those in the Control Condition at Attempt number 4 (the grand trimean number of attempts).

6.3.5 Examination of NSH High Performers

NSH could not be used to split participants into high and low Gf ability groups. The relatively high level of difficulty of the set resulted in a floor effect and restriction of range. Those in the bottom 50% scored in the range of only 0-1, while those in the top 50% scored in the much wider range of 2-12. This restriction of range makes it inappropriate to collapse them into two separate groups and renders HLM less informative.
However, a relevant post-hoc question that arises from this is whether those who performed very well on the NSH task also performed very well on the MSG Complex item, in the Complex Control condition. That is, are there people who are able to reason well inductively, without much across-item learning opportunity? And if so, can they do so on more than one task? Specifically, can those who perform very well on a task such as NSH, where learning across the task is hypothesized to be minimized, do equally well on the Complex Control item where no across-item learning is provided.

The top 5% of participants on NSH were selected to be examined. Their performance on NSH was substantially higher than that for the rest of the sample; their mean score on NSH was 9 (75%, SD = 1.25, n = 18), compared to the entire sample’s mean of 2.55 (21%, SD = 2.65, n = 208). However, their pattern of performance on the MSG task was not very different to that of the rest of the sample (Figure 6.5). They appear to demonstrate a learning effect and proactive interference effect, although, their trimean on the Complex Item (Control Condition) was 34, seems somewhat better than the same trimean score for the rest of the sample which was 45. The sample size for this extreme group was too small for statistical tests of these differences to be meaningful (n ≈ 6, per condition). However, their general pattern of performance suggests that despite their special level of performance on NSH, they did not display the same level of performance on MSG without across-item learning opportunity, and when across-item learning opportunity was provided, they seemed to benefit from it just as much as other participants.
6.4 Summary and General Discussion

The current study was a larger version of pilot Experiment 2 and was intended to test the general hypotheses and predictions outlined in Chapter 5. The results from the current study largely:

- Did not fully support the Classical Gf Hypothesis
- Supported the General Learning Hypothesis
- Gave some support for the Individual Differences Learning Hypothesis.
- Supported the General Proactive Interference Hypothesis

**Figure 6.5.** Trimean MSG scores on Complex and PI items for Experimental and Control Conditions, for those in the top 5% of NSH.
• Did not support the Individual Differences Proactive Interference Hypothesis.

The Classical Gf Hypothesis: Consistent with pilot Experiment 2, it was found that without learning opportunity available across the task, complex, novel items are extremely difficult. The Complex Control Item resulted in floor effects for all Gf ability groups (although those of higher Gf still outperformed those of lower Gf ability). Furthermore, in the absence of across-item learning, participants’ performance on the Complex Item did not improve with more attempts and this applied to all ability groups. Thus, within-item analogues (e.g. the series of numbers conforming to the rule) by themselves are not good facilitators of induction in complex, novel items. This is not totally consistent with the Classical Gf Hypothesis that predicted that in the absence of across-item learning, induction on this complex Gf item would still be possible. However, the hypothesis also predicts that in the absence of across-item learning those of higher Gf would outperform those of lower Gf – which was what the results found. Despite the low level of performance for all ability groups, those of higher Gf did outperform those of lower Gf. Thus, the Classical Gf Hypothesis was partially supported.

The General Learning Hypothesis: When across-item learning opportunity was provided, the floor effect disappeared, despite the theoretical complexity of the item remaining the same. This was true for all Gf ability groups. That is, when learning opportunity across the task was provided, all ability groups were able to significantly learn. This supports the General Learning Hypothesis that states that simpler items earlier in the task should act as learning opportunity for more complex (possibly, novel) items with the same types of rules that appear later in the task. It is likely that the availability of these simpler items for the Experimental Group (but not the Complex Control Group), provided the
group with the relevant conceptual knowledge that made inducing the rules in the Complex Item easier (compared to those in the Complex Control Group).

The Individual Differences Learning Hypothesis: There was some (marginally significant) evidence that those of higher Gf learn more across the task than those of lower Gf. The difference in Complex Item performance between the Experimental Group and the Complex Control Group was marginally significantly larger for those of higher Gf ability in the HLM analysis. Although the results were only marginally significant, there was a consistent trend across the trimeans of all the items in the Experimental Condition that suggested that those of higher Gf indeed do learn more across the task when across-item learning opportunity is provided.

The General Proactive Interference Hypothesis: The significance of condition as a predictor of performance on the PI Item in the HLM (with those in the Control Condition doing better than those in the Experimental Condition), suggests that all participants (regardless of Gf ability) suffer from proactive interference on an item of a different rule type after having solved items of a particular rule-type. This was contrary to the non-parametric Mann-Whitney tests which suggested that no groups suffered from proactive interference. However the HLM provides a more complete picture of what happens in the MSG task, as it is able to model performance at each attempt number and is a more powerful test than the Mann-Whitney test. Thus, we conclude from these results and from the support that the General Learning Hypothesis received, that there is support for the General Proactive Interference Hypothesis, that is, across-item learning leads to proactive interference for all.
**The Individual Differences Proactive Interference Hypothesis:** In the HLM, Gf ability did not predict performance on the PI Item, over and above condition. Thus, the Individual Differences Proactive Interference Hypothesis was not supported. There appears to be no difference in the ability to combat proactive interference for those of different levels of Gf ability. Indeed, the *PI Item correlated moderately with NSEH in its Control Condition, but there was no relationship between the two in the Experimental Condition.* If learning leads to proactive interference and those of higher Gf ability are better at combating it, then the trend should have been the other way around – i.e., the PI Item in the Experimental Condition should have correlated more highly with NSEH than the PI Item in the Control Condition.

**A Hypothesis Regarding the Methodological Issue of the Presentation Format of the Gf Marker:** In terms of presentation format of the Gf markers, there appears to be a lot of similarity between Number Series presented in easy-to-hard format and when only hard and moderately hard items are presented. They correlate moderately with each other. A difference is that when only hard and moderately hard items are presented, the result is floor effect and restriction of range.

Examination of the top 5% on NSH suggested that their pattern of performance on the MSG task was not very different to that of the rest of the sample. One might expect that if NSH minimised across-item learning opportunity, those who scored highly on it might also perform very well on the Complex Control Item (where there is no across-item learning opportunity); that is, there may be those who are generally very capable of within-item schema induction, and can do without across-item learning. However, this was not the case. Like the rest of the sample, they were not able to reason very well without across-item learning opportunity on the MSG task and benefitted from the learning opportunity when it was provided. Thus, their ability to reason well in the absence of much across-item learning
opportunity may have been task specific. It is possible that those who performed very well on NSH may have already been familiar with the rules used in NSH. That is, they may have brought relevant, prior conceptual knowledge of the rules with them to NSH.

A limitation of the current study is that the Gf markers contained the same stimulus type as the MSG task – that is, despite containing different types of items, both involved reasoning with numbers. Thus, in the next study (Experiment 4), we aim to replicate the findings from the current study with the Letter Series task as the Gf marker.
CHAPTER 7

EXPERIMENT 4:

DIFFERENT BUT RELATED: WITHIN-ITEM RULE DISCOVERY, ACROSS-ITEM LEARNING, USE OF PRIOR KNOWLEDGE AND PROACTIVE INTERFERENCE

LETTER SERIES

7.1 Experiment 4: The MSG Task and Letter Series

This experiment was run at the same time as Experiment 3. Its aim was to replicate the findings from Experiment 3 and to see if the relationships between the MSG task and Number Series (the Gf marker) would generalise to the MSG task and the Letter Series task (as the Gf marker). Since both Number Series and the MSG task involve numbers as stimuli we thought an examination of the MSG task’s relationship with a Gf task that did not involve numbers was warranted.

The experimental design was the same as Experiments 2 and 3, and we make the same hypotheses and predictions (refer back to Chapter 5). Once again, the dependant variable was performance on the Complex and PI Items in the MSG task and the independent variables were condition (Experimental and Complex Control or Experimental and PI Control) and Gf ability. The Letter Series Task was used as the marker of Gf. A two-level Hierarchical Linear Model (HLM) was used to understand the relationship between performance on the MSG task, condition and Gf ability (as indicated by the Letter Series task).

The Gf marker was presented in two formats to all participants:
• Easy-to-hard (to allow for across-item learning opportunity),

• Hard and moderately hard items only, randomly inter-mixed (to minimise across-item learning opportunity).

Ideally, to minimise learning opportunity, presentation format should be hard-to-easy. However, due to time constraints, only hard and moderately hard items were included.

7.2 Method

Participants

The students were enrolled in second year undergraduate psychology at the University of Sydney and participated as part of their course work. In total, 169 students participated in the study (70% female) with mean age = 21.41 years (SD = 5.80). Data was collected concurrently with data for Experiment 3.

Cognitive Tasks

1) Modified Sweller & Gee (MSG) Task

This was the same task presented in pilot Experiment 2 and Experiment 3, with no modifications. Identical to Experiment 2 and 3, there were three versions, one for each of the three conditions: the Experimental Condition, the Complex Control Condition and the PI Control Condition.

2) Letter Series (Gf markers)

This was basically the same task as LSEH presented in pilot Experiment 2. However, this time it was presented as two separate tasks, in two different formats:
• Easy-to-hard (‘LSEH’) – hypothesized to allow learning opportunity,
• Only hard and moderately hard items, randomly intermixed (‘LSH’) – hypothesized to minimise learning opportunity.

LSH always preceded LSEH to prevent the easy items from LSEH acting as learning items for LSH.

**Procedure**

Testing was conducted in groups of approximately 10 or 23 participants. All tasks were computer administrated. An experimenter provided general instructions at the beginning of each session and was present throughout the session to assist participants and ensure that the test protocol was followed. A set of computerized (more specific) instructions preceded each task. Participants made their responses for all tasks using a standard keyboard and mouse. Participants were not allowed to use pen and paper for any of the items. Tasks were always presented in the order MSG, LSH, LSEH, to prevent the Gf marker tasks from creating proactive interference for the MSG task.

Those in the Experimental and Complex Control conditions received only MSG, LSH and LSEH. However, those in the PI Control Condition, received MSG, followed by a short version of Raven’s Advanced Progressive Matrices (Raven, 1962), followed by LSH, then LSEH. The Raven’s task was meant to act only as a filler task. The MSG task in the PI Control Condition was very short, consisting of only one very easy MSG item. Hence, Raven’s was included to control for any fatigue that may have affected those in the Experimental and Complex Control conditions.
**Statistical Model**

HLM was used to model performance on MSG at each attempt on the Complex and PI items, with condition and LSEH as predictors. The equivalent model was not constructed for LSH due to restriction of range issues that will subsequently be discussed. An identical model was used for Experiment 3.

**7.3 Results & Discussion**

This section will be organised as follows. Firstly, we will present descriptive statistics and correlations. We then move on to non-parametric tests that compare performance on the Complex and PI items for control and experimental groups, based on Gf ability groups (created by LSEH). For these tests, dependent variable for the MSG Task is the number of attempts needed on an item before multiple correct responses are achieved. This gives *an approximation of the point at which participants start to understand the rule in MSG Complex/PI items and get attempts correct*. This was essentially the same type of analyses used in pilot Experiment 2 and the main aim of these analyses here will be to see whether the results from the pilot can be replicated.

We then move on to HLM analyses to gain a more complete picture of participants’ performance. These analyses will essentially be the same type as those which were the main focus of Experiment 3. The HLM will allow us to model *the probability that a participant will get an attempt in the Complex/PI items correct (for all attempts)*, based on condition and their performance on LSEH. The aim is to see if the findings from Experiment 3 can be replicated using a Gf marker involving different stimuli type to the MSG task (i.e., letters instead of numbers).
Non-parametric tests and HLM were not run for LSH due to floor effect and restriction of range issues for this marker. Instead, we examine the trimeans on the MSG task for those who scored in the top 5% on the LSH. The rationale for this was to see whether those who perform very well on such a task such as LSH, where learning across the task is hypothesized to be minimized, could do equally well on the Complex Control item where learning is also minimized.

### 7.3.1 Descriptive Statistics

Demographic information is presented in Table 7.1. Descriptive statistics for marker tasks used in the study are presented separately by condition in Table 7.2. Cronbach’s alpha is in the acceptable range (for research purposes). Means for the LSH were very low (ranging from 7% - 10% correct across the three conditions), with lower standard deviation than the LSEH set. Hence, any direct comparison between the two sets, based on methods which examine variance (such as correlations) will be done with caution.

<table>
<thead>
<tr>
<th>Condition (n)</th>
<th>Age Mean</th>
<th>SD</th>
<th>Male (%)</th>
<th>Female (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental (61)</td>
<td>22.70</td>
<td>7.75</td>
<td>18 (30)</td>
<td>43 (70)</td>
</tr>
<tr>
<td>Complex Control (49)</td>
<td>20.42</td>
<td>3.27</td>
<td>16 (33)</td>
<td>33 (67)</td>
</tr>
<tr>
<td>PI Control (59)</td>
<td>20.28</td>
<td>4.36</td>
<td>15 (26)</td>
<td>44 (74)</td>
</tr>
</tbody>
</table>

Total (169)
Descriptive statistics for the MSG task are presented in Table 7.3 separately for each condition. Distributions for many of the MSG items were not normal. This is consistent with the findings of Sweller and Gee (1978) and the findings from Experiments 2 and 3.

Table 7.2
Descriptive statistics for Gf marker tasks.

<table>
<thead>
<tr>
<th>Task</th>
<th>Items</th>
<th>Complex Control</th>
<th>Experimental</th>
<th>PI Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean (%Correct)</td>
<td>SD</td>
<td>n</td>
</tr>
<tr>
<td>LSEH</td>
<td>20</td>
<td>9.31 (47)</td>
<td>5.02</td>
<td>47</td>
</tr>
<tr>
<td>LSH</td>
<td>12</td>
<td>1.06 (9)</td>
<td>1.56</td>
<td>49</td>
</tr>
</tbody>
</table>

Note. α = Cronbach’s coefficient alpha.
Table 7.3
Descriptive statistics for Modified Sweller & Gee task.

<table>
<thead>
<tr>
<th>Item</th>
<th>Complex Control (n = 49)</th>
<th>Experimental (n = 61)</th>
<th>PI Control (n = 59)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Median (Quartiles)</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>1</td>
<td>5.62 (8.19)</td>
<td>3 (2, 6)</td>
<td>2.81 (2.17)</td>
</tr>
<tr>
<td>2</td>
<td>8.09 (12.60)</td>
<td>3 (2, 6)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>24.31 (19.93)</td>
<td>10 (5, 47)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>18.33 (16.91)</td>
<td>8 (6, 39)</td>
<td></td>
</tr>
<tr>
<td>Complex</td>
<td>43.21 (10.58)</td>
<td>47 (45, 48)</td>
<td>17.95 (17.69)</td>
</tr>
<tr>
<td>PI</td>
<td>4.56 (7.48)</td>
<td>3 (2, 5)</td>
<td>9.14 (15.41)</td>
</tr>
</tbody>
</table>

Note. Scores represent number of attempts (out of an allowable 50) minus number of correct responses.
7.3.2 Correlations

Table 7.4 presents Spearman’s rho correlations for each of the MSG items with LSEH and LSH, separately for each condition. Spearman’s rho is a non-parametric rank correlation.

The increasing trend in correlations between LSEH and MSG items as they increase in difficulty and appear later in the task (-.06 to -.27), seen in Experiments 2 and 3 is repeated here for LSEH. The process causing these increases in correlations may be learning. That is, learning across the task may be important in NSEH and may become increasingly important in MSG items as they get harder. That is, as the items get harder, participants may have to rely more on what they learn from previous items. Alternatively, the correlations between MSG and LSEH may increase with increases in the complexity of the MSG items.

Also, as was the case in the Experiments 2 and 3, the Complex Control Item correlates moderately with easy-to-hard Gf marker (-.34). This may indicate that the within-item induction in the Complex Control Item is similar in process to across-item learning. Alternatively, the correlations may be due to the complexity of the Complex Control Item and similarities in within-item induction processes, and not related to across-item learning processes.

The PI Item correlates moderately with LSEH in its Control Condition (-.30), but not in the Experimental Condition (-.04). This was also the case for NSEH in experiment 3. This is evidence against the Individual Differences Proactive Interference hypothesis (that those of higher Gf ability suffer less from proactive interference) because if learning leads to proactive interference and those of higher Gf ability are better at combating it, then the PI Item in the Experimental Condition should have correlated more highly with LSEH than the PI Item in the Control Condition.
LSH did not have any significant relationships with any of the MSG items. This may be due to the restriction of range issues with LSH. However, LSH correlated moderately with LSEH (.33 to .65) in each condition.

Table 7.4  
Spearman's rho correlations between Gf marker tasks and Modified Sweller & Goo items.

<table>
<thead>
<tr>
<th>Task</th>
<th>Item</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Complex n = 49</td>
</tr>
<tr>
<td>LSEH</td>
<td>1</td>
<td>-.04</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-.13</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-.26*</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-.31*</td>
</tr>
<tr>
<td></td>
<td>Complex</td>
<td>-.34*</td>
</tr>
<tr>
<td></td>
<td>PI</td>
<td>-.04</td>
</tr>
<tr>
<td>LSH</td>
<td>1</td>
<td>-.09</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-.04</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-.17</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-.25</td>
</tr>
<tr>
<td></td>
<td>Complex</td>
<td>-.10</td>
</tr>
<tr>
<td></td>
<td>PI</td>
<td>.09</td>
</tr>
</tbody>
</table>

*Note.*  *p < .05, **p < .01.
Pearson correlations between LSEH and LSH were: .42*, .65**, and .33* for Complex Control, Experimental and PI Control conditions respectively.
7.3.3 Non-parametric Comparisons of Gf Ability Groups (Based on LSEH)

Construction of LSEH Ability Groups

LSEH was used to divide participants into low and high Gf groups. Those who scored in the bottom 50% (score range: 0 - 9) were placed in the “Low LSEH” ability group and those in the top 50% (score range: 10 - 20) were placed in the “High LSEH” ability group.

Since our aim will be to compare the conditions within the ability groups, it is important to know that the condition groups did not differ significantly on Gf at the outset. Due to unequal sample sizes in the three conditions, the Univariate General Linear Model (GLM) analysis of variance procedure was used with Type III sums of squares. Descriptive statistics and F-tests are presented in Table 7.5. There was no significant difference on Gf across conditions for any group.

The Classical Gf Hypothesis – LSEH

The non-parametric Mann-Whitney test was used to test for differences between ability groups within the Complex Control Condition. Descriptive statistics can be found in Table 7.6. Consistent with Experiments 2 and 3, the High LSEH group solved the Complex item significantly faster than the Low LSEH group, in the Control Condition (Mann–Whitney $U = 177$, $n_{\text{Low LSEH}} = 25$, $n_{\text{High LSEH}} = 22$, $p = .03$, two-tailed). This is consistent with the Classical Gf Hypothesis. Also consistent with Experiments 2 and 3, there appears to be a floor effect for the Complex Control Item for both groups. This suggests that without across-item learning opportunity, complex, novel items are very difficult for all ability groups.
### Table 7.5
*Descriptive statistics and F-tests for Low LSEH and High LSEH groups, on LSEH across conditions.*

<table>
<thead>
<tr>
<th>Group</th>
<th>Complex Control</th>
<th>Experimental</th>
<th>PI Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (%Correct)</td>
<td>Mean (%Correct)</td>
<td>Mean (%Correct)</td>
</tr>
<tr>
<td>Low LSEH</td>
<td>5.60 (28)</td>
<td>6.25 (31)</td>
<td>5.60 (28)</td>
</tr>
<tr>
<td>High LSEH</td>
<td>13.54 (68)</td>
<td>13.66 (68)</td>
<td>13.37 (67)</td>
</tr>
</tbody>
</table>

### Table 7.6
*Descriptive statistics and Mann-Whitney tests for MSG task scores across conditions, within Gf ability groups.*

<table>
<thead>
<tr>
<th>Group</th>
<th>Item</th>
<th>Control Groups</th>
<th>Experimental</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean (SD)</td>
<td>Median (Quartiles)</td>
</tr>
<tr>
<td>Low LSEH</td>
<td>Complex</td>
<td>45.20 (7.58)</td>
<td>47 (44, 49)</td>
</tr>
<tr>
<td></td>
<td>PI</td>
<td>13.74 (18.7)</td>
<td>4 (1, 39)</td>
</tr>
<tr>
<td>High LSEH</td>
<td>Complex</td>
<td>40.43 (13.20)</td>
<td>45 (42, 46)</td>
</tr>
<tr>
<td></td>
<td>PI</td>
<td>2.44 (1.85)</td>
<td>2 (1, 3)</td>
</tr>
</tbody>
</table>

*Note.* n₁ = n for control group, n₂ = n for experimental group.

* p < .05, ** p < .01, two-tailed.
The General Learning Hypothesis and the Individual Differences Learning Hypothesis - LSEH

Multiple Mann-Whitney tests were used to test for differences between the Experimental and Complex Control condition groups, within each ability group. The tests and their relevant descriptive statistics are presented in Table 7.6. All ability groups demonstrated the sequence/learning effect; their respective Experimental groups solved the Complex item significantly faster than their Complex Control groups. That is, both high and low LSEH ability groups could learn within a task. This is consistent with Experiments 2 and 3 and Figure 7.1 illustrates these effects graphically for the LSEH groups using trimeans.

While all ability groups were able to learn, Figure 7.1 suggests that those from the High LSEH group were more able to benefit from their learning opportunity than those from the Low LSEH group, as there is a larger gap between their Complex Control and Experimental groups (compared to the smaller gap between the Low LSEH Complex Control and Experimental groups). This is consistent with Experiment 2 and 3. However, in the Experimental Condition, the High LSEH group did not solve the Complex Item significantly faster than those in the Low LSEH group, (Mann–Whitney $U = 325$, $n_{LowLSEH} = 24$, $n_{HighLSEH} = 36$, $p = .10$ two-tailed), although the difference was marginally significant. We shall return to exploring this potential interaction effect with HLM, shortly.

Figure 7.2 illustrates trimeans across the MSG task for LSEH groups in the Experimental Condition. Those in the High LSEH group display the “learning-to-learn effect”. That is, despite the items increasing in complexity across the task, trimeans did not increase after item 3. The Low LSEH group also display this effect, but to a lesser extent, consistent with Experiment 2 and 3.
Figure 7.1. Trimean MSG scores on Complex and PI items for Experimental and Control Conditions, by LSEH ability groups.

Figure 7.2. Trimean MSG scores across the task for Experimental groups, by LSEH ability groups.
The General Proactive Interference and the Individual Differences Proactive Interference Hypothesis – LSEH

Multiple Mann-Whitney tests were used to test for differences between the Experimental and Complex Control condition groups, within each ability group. The tests and their relevant descriptive statistics were presented in Table 7.6. No groups seemed to suffer from proactive interference; however those in the High LSEH group suffered marginally significant proactive interference. This is similar to the findings from pilot Experiment 2, where the Mann-Whitney test suggested that those in the High LSEH group was the only group to suffer from significant proactive interference. It is also similar in finding to Experiment 3, where according to the Mann-Whitney test, no groups suffered from proactive interference, including the High NSEH group. Overall, this is weak support for the General Proactive Interference Hypothesis (that learning across a task results in proactive interference for later items when the rules are dissimilar and the more you learn, the more you will suffer from proactive interference) and contrary to the Individual Differences Proactive Interference Hypothesis (that those of lower Gf suffer more from proactive interference than those of higher Gf ability).

7.3.4 Hierarchical Linear Modelling with LSEH

A two-level Hierarchical Linear Modelling (HLM) was used to further understand the relationship between performance on the MSG task, condition and LSEH. The probability of getting an attempt correct (where “Correct” = 1 if correct; 0 if incorrect)\(^{11}\) in the item was

\(^{11}\) The MSG task was programmed to exit from an item after a total of 12 correct responses for that item. This would have resulted in a lot of missing data for the HLM. Hence, missing responses due to participants successfully exiting an item were assigned a “1” (i.e., scored as “correct”).
modelled with log-odds as the dependent variable. The two-level model is identical to the one used and described in detail for Experiment 3, Chapter 6.

The level-1 model is stated by equation (1). Attempt is the attempt number in the item, centred at the grand-trimean number of attempts for all participants on the item (attempt number 34 for Complex Item and attempt number 3 for the PI item).

Equation (1) predicts $\eta_{ij}$, the log-odds of success for participant $i$ at time $t$; based on $\text{Attempt}_{ij}$.

Thus, $\pi_{0i}$ represents the log-odds of success of participant $i$ when $\text{Attempt}_{ij}$ is equal to 0 (where 0 = grand-trimean number of attempts for the item). $\pi_{1i}$ is the growth (improvement) rate for participant $i$ over the item. Since this model uses Bernoulli sampling, there is no error term in equation (1):

$$\eta_{ij} = \pi_{0i} + \pi_{1i}\text{Attempt}_{ij} \quad (1)$$

The level-2 model predicts the coefficients in the level-1 model and is stated by equations (2) and (3). Condition is the dummy variable for condition (0 = Control; 1 = Experimental) and LSEH_Cent is centred LSEH (0 = grand-mean):

$$\pi_{0i} = \beta_{00} + \beta_{01}\text{Condition}_{1i} + \beta_{02}\text{LSEH}_\text{Cent}_{2i} + \beta_{03}\text{Condition} \times \text{LSEH}_\text{Cent}_{3i} + r_{0i} \quad (2)$$

$$\pi_{1i} = \beta_{10} + \beta_{11}\text{Condition}_{1i} + \beta_{12}\text{LSEH}_\text{Cent}_{2i} + \beta_{13}\text{Condition} \times \text{LSEH}_\text{Cent}_{3i} \quad (3)$$

$\beta_{00}$ is the value for $\pi_{0i}$ when all of its predictors are 0 (that is, for someone in the Control Condition, with grand-mean LSEH score). $\beta_{10}$ is the growth rate (that is, the value for $\pi_{1i}$) when all of its predictors are 0 (that is, for someone in the Control Condition, with grand-mean LSEH score). $r_{0i}$ is the deviation of participant $i$ from the trimean performance at attempt number 34 (for Complex Item) or 3 (PI Item).
Thus, the intercept $\pi_{0i}$ predicts the log-odds of success for participant $i$, at the grand-trimean number of attempts based on the main effects of Condition, LSEH and their interactions. The Attempt slope $\pi_{1i}$ predicts the growth rate for participant $i$, over the course of the item, based on the main effects of Condition, LSEH and their interactions. HLM models were run separately for the Complex Item and the PI Item. We first present the analyses for the Complex Item.

HLM Analyses of the Complex Item (the Classical Gf Hypothesis and the Learning Hypotheses)

Table 7.7 presents the estimates of the HLM parameters, which were obtained in HLM 6 (Raudenbush et al., 2004) for the Complex Item. All results for the Complex Item are consistent with Experiment 3.

Intercept $\pi_{0i}$ represents the log-odds of success at the grand-trimean number of attempts for the Complex Item, that is, attempt number 34. Slope $\pi_{1i}$ represents the growth (improvement) rate for participant $i$ over the item, that is, changes to the log-odds of success as a function of attempt number.

Model 1 includes all predictors. Non-significant and weak predictors were dropped from the model based on two criteria. Predictors with t-ratios of smaller magnitude than +/- 1 were dropped (since the threshold for significance is +/- 2) as were predictors with significant coefficients weaker than +/- 0.01 (since log-odds of +/- 0.009 corresponds to an odds ratio of 1.009 and would not contribute much to the model). The final results are presented in Table 7.7, Model 2.
Table 7.7
Estimated parameters of the Hierarchical Linear Model for performance on the Complex Item.

<table>
<thead>
<tr>
<th>Model</th>
<th>Fixed Effect</th>
<th>Coefficient (Log-odds)</th>
<th>Odds Ratio</th>
<th>SE</th>
<th>t-ratio</th>
<th>d.f</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>For Intercept, $\pi_0$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Intercept2, $\beta_{00}$</td>
<td>-2.513</td>
<td>0.080</td>
<td>0.271</td>
<td>9.273</td>
<td>101</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>Condition, $\beta_{01}$</td>
<td>3.652</td>
<td>38.588</td>
<td>0.565</td>
<td>6.464</td>
<td>101</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>LSEH_Cent, $\beta_{02}$</td>
<td>0.088</td>
<td>1.092</td>
<td>0.040</td>
<td>2.200</td>
<td>101</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>Condition X LSEH_Cent, $\beta_{03}$</td>
<td>0.212</td>
<td>1.236</td>
<td>0.120</td>
<td>1.767</td>
<td>101</td>
<td>.08</td>
</tr>
<tr>
<td>For Attempt Slope, $\pi_1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Intercept2, $\beta_{10}$</td>
<td>0.008</td>
<td>1.008</td>
<td>0.012</td>
<td>0.667</td>
<td>5242</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>Condition, $\beta_{11}$</td>
<td>0.081</td>
<td>1.084</td>
<td>0.022</td>
<td>3.682</td>
<td>5242</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>LSEH_Cent, $\beta_{12}$</td>
<td>0.000</td>
<td>0.999</td>
<td>0.001</td>
<td>0.000</td>
<td>5242</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>Condition X LSEH_Cent, $\beta_{13}$</td>
<td>0.008</td>
<td>1.008</td>
<td>0.005</td>
<td>1.600</td>
<td>5242</td>
<td>0.11</td>
</tr>
<tr>
<td>Model 2</td>
<td>For Intercept, $\pi_0$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Intercept2, $\beta_{00}$</td>
<td>-2.514</td>
<td>0.080</td>
<td>0.272</td>
<td>9.243</td>
<td>101</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>Condition, $\beta_{01}$</td>
<td>3.527</td>
<td>34.029</td>
<td>0.523</td>
<td>6.744</td>
<td>101</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>LSEH_Cent, $\beta_{02}$</td>
<td>0.088</td>
<td>1.093</td>
<td>0.038</td>
<td>2.316</td>
<td>101</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>Condition X LSEH_Cent, $\beta_{03}$</td>
<td>0.164</td>
<td>1.178</td>
<td>0.085</td>
<td>1.929</td>
<td>101</td>
<td>0.06</td>
</tr>
<tr>
<td>For Attempt Slope, $\pi_1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Intercept2, $\beta_{10}$</td>
<td>0.008</td>
<td>1.008</td>
<td>0.012</td>
<td>0.667</td>
<td>5244</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>Condition, $\beta_{11}$</td>
<td>0.088</td>
<td>1.093</td>
<td>0.022</td>
<td>4.000</td>
<td>5244</td>
<td>**</td>
</tr>
</tbody>
</table>

Note. *$p < .05$, **$p < .01$.

 Intercept1, $n_0$ = predicted log-odds of success ($Correct = 1$) on the Complex Item at attempt number 34.

 Attempt Slope, $\pi_1$ = predicted growth rate of success ($Correct = 1$), over the course of the Complex Item.
Performance at the Grand-Trimean Number of Attempts: The coefficients that predict $\pi_0$ indicate that at attempt number 34, the log-odds of success is significantly predicted by Condition and LSEH and the interaction between Condition and LSEH is marginally significant. These effects are more clearly illustrated in Figure 7.3. At attempt 34, regardless of LSEH, those in the Experimental Condition have a higher probability of success and this supports the General Learning Hypothesis. Regardless of Condition, those with higher scores on LSEH have a higher probability of success. Also, the marginally significant interaction between Condition and LSEH is indicated by the larger gap between the broken lines - i.e. between the higher Gf, Control and Experimental groups (compared to the smaller gap between the solid lines – i.e. the Control and Experimental groups for those of lower Gf). That is, there is some (marginally significant) evidence that those of higher Gf profited more from the opportunity to learn across items than those of lower Gf. Thus, there is some support for the Individual Differences Learning Hypothesis. This result is consistent with the equivalent result from Experiment 3.

Again, there appears to be a floor effect for the Complex Control Item for both groups. This suggests that without across-item learning opportunity, complex, novel items are very difficult for all ability groups. Visual examination of Figure 7.3 suggests that those in the Control Condition (even of higher Gf) were rarely able to reach the correct answer on any attempt – the predicted probability of success for those in the 75th percentile of NSEH (Gf ability) in the Complex Control Condition rarely goes above 10%. Thus, the results are somewhat inconsistent with the Classical Gf Hypothesis because it appears that without some exposure to earlier, simpler items to provide participants with the relevant, related knowledge, induction is practically impossible. This result is consistent with the equivalent result from Experiment 3.
**Growth of Success:** The coefficients for $\pi_1$ in Table 7.7 indicate that Condition is the only significant predictor of the rate of growth of success, over the course of the Complex Item. It can be seen in Figure 7.3 that only those in the Experimental Condition improved their performance with more attempts. Furthermore, Figure 7.3 suggests that those in the Control Condition were rarely able to reach the correct answer on any attempt. The lack of significance of LSEH in predicting rate of growth suggests that Condition (i.e., exposure to earlier, simpler items) is more important than Gf ability in being able to induce the correct answer. That is, it appears that it is very difficult to make mappings between within-item analogues to induce the general concept (i.e., work out the rules by looking for generalisations among the item’s stimuli) in complex Gf items, unless one is provided with earlier, easier items to provide one with the relevant, prior knowledge. This seems to be the case, regardless of one’s Gf ability. This suggests that the knowledge provided by across-item learning is essential to induction, for all levels of Gf ability and is contrary to the Classical Gf Hypothesis but supports the General Learning Hypothesis. This result is consistent with the equivalent result from Experiment 3.
Figure 7.3. Predicted probability of success (Correct = 1) as a function of Condition, LSEH and Attempt, on Complex Item, based on Model 1 from Table 7.7. Black line indicates differences in predicted success between those in the Experimental Condition and those in the Control Condition, separately for those of higher and lower Gf ability at Attempt number 34 (the grand trimean number of attempts).

HLM analyses of PI item (the Proactive Interference Hypotheses)

A similar two-level HLM was used to model probability of getting an attempt correct (where “Correct” = 1 if correct; 0 if incorrect) in the PI Item, with its log-odds as the dependent variable and Condition and LSEH as the independent variable. Table 7.8 presents the
estimates of the HLM parameters, which were obtained in HLM 6 (Raudenbush et al., 2004), for the PI Item. Intercept $\pi_{0i}$ represents the log-odds of success at the grand-trimean number of attempts for the PI Item, that is, attempt number 3. Slope $\pi_{1i}$ represents the growth (improvement) rate for participant $i$ over the item, that is, changes to the log-odds of success as a function of attempt number. Model 1 includes all predictors. Non-significant and weak predictors were dropped from the model based on two criteria. Predictors with t-ratios of smaller magnitude than +/- 1 were dropped as were predictors with significant coefficients weaker than +/- 0.01. The final results are presented in Table 7.8., Model 2.
Table 7.8
Estimated parameters of the Hierarchical Linear Model for performance on the PI Item.

<table>
<thead>
<tr>
<th>Model</th>
<th>Fixed Effect</th>
<th>Coefficient (Log-odds)</th>
<th>Odds Ratio</th>
<th>SE</th>
<th>t-ratio</th>
<th>d.f</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>For Intercept, $\pi_0$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Intercept2, $\beta_{00}$</td>
<td>-0.022</td>
<td>0.977</td>
<td>0.240</td>
<td>-0.093</td>
<td>113</td>
<td>.92</td>
</tr>
<tr>
<td></td>
<td>Condition, $\beta_{01}$</td>
<td>-0.854</td>
<td>0.425</td>
<td>0.311</td>
<td>-2.746</td>
<td>113</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>LSEH_Cent, $\beta_{02}$</td>
<td>0.008</td>
<td>1.008</td>
<td>0.095</td>
<td>0.084</td>
<td>113</td>
<td>.92</td>
</tr>
<tr>
<td></td>
<td>Condition X LSEH_Cent, $\beta_{03}$</td>
<td>0.001</td>
<td>1.001</td>
<td>0.057</td>
<td>0.018</td>
<td>113</td>
<td>.97</td>
</tr>
<tr>
<td>For Attempt Slope, $\pi_1$</td>
<td>Intercept2, $\beta_{10}$</td>
<td>0.203</td>
<td>1.225</td>
<td>0.072</td>
<td>2.819</td>
<td>5842</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>Condition, $\beta_{11}$</td>
<td>0.141</td>
<td>1.151</td>
<td>0.099</td>
<td>1.424</td>
<td>5842</td>
<td>.15</td>
</tr>
<tr>
<td></td>
<td>LSEH_Cent, $\beta_{12}$</td>
<td>-0.016</td>
<td>0.983</td>
<td>0.036</td>
<td>-0.444</td>
<td>5842</td>
<td>.65</td>
</tr>
<tr>
<td></td>
<td>Condition X LSEH_Cent, $\beta_{13}$</td>
<td>0.021</td>
<td>1.021</td>
<td>0.020</td>
<td>1.037</td>
<td>5842</td>
<td>.30</td>
</tr>
<tr>
<td>Model 2</td>
<td>For Intercept, $\pi_0$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Intercept2, $\beta_{00}$</td>
<td>0.079</td>
<td>1.082</td>
<td>0.208</td>
<td>0.380</td>
<td>115</td>
<td>.70</td>
</tr>
<tr>
<td></td>
<td>Condition, $\beta_{01}$</td>
<td>-0.967</td>
<td>0.379</td>
<td>0.273</td>
<td>-3.542</td>
<td>115</td>
<td>**</td>
</tr>
<tr>
<td>For Attempt Slope, $\pi_1$</td>
<td>Intercept2, $\beta_{10}$</td>
<td>0.115</td>
<td>1.122</td>
<td>0.030</td>
<td>3.836</td>
<td>5846</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>Condition, $\beta_{11}$</td>
<td>0.236</td>
<td>1.266</td>
<td>0.075</td>
<td>3.154</td>
<td>5846</td>
<td>**</td>
</tr>
</tbody>
</table>

Note. *$p < .05$. **$p < .01$.

$\pi_0$ = predicted log-odds of success (Correct = 1) on the PI Item at attempt number 3.

$\pi_1$ = predicted growth rate of success (Correct = 1), over the course of the PI Item.

**Performance at the Grand-Trimane Number of Attempts:** The coefficients for $\pi_0$ indicate that at attempt number 3, the log-odds of success is (marginally) significantly predicted by Condition. Figure 7.4 indicates that at Attempt number 3, regardless of Gf ability, those in the PI Control Condition have a (marginally significant) higher probability of
success than those in the Experimental Condition, which is consistent with the General Proactive Interference Hypothesis. That is, those who learn about one type of rule, find it harder to solve items of a different rule type later on. However, this effect is general and not predicted by Gf ability. This supports the General Proactive Interference Hypothesis but not the Individual Differences Proactive Interference Hypothesis.

This is not consistent with the Mann-Whitney test in section 7.2.3 and Experiment 3 where those of higher Gf ability (as indicated by LSEH) suffered from proactive interference. However, these effects were small; the difference between the PI Control Condition and the Experimental Condition for the High LSEH was only marginally significant in section 7.2.3.

*Growth of Success:* The coefficients for $\pi_1$ indicate that only Condition was a significant predictor of the rate of growth (improvement) of success, over the course of the PI Item. It can be seen in Figure 7.4 that on average, the red lines representing the Experimental Condition are steeper than the blue lines that represent the PI Control Condition. That is, while the Control Condition did better than the Experimental Condition at Attempt 3, the Experimental Condition would have quickly caught up. That is, while the result supports the General Proactive Interference Hypothesis, proactive interference is temporary. This is consistent with Experiment 3.
Figure 7.4. Predicted probability of success (Correct = 1) as a function of Condition, LSEH and Attempt, on Complex Item, based on Model 1 from Table 7.8. Black line indicates differences in predicted success between those in the Experimental Condition and those in the Control Condition at Attempt number 3 (the grand trimean number of attempts).

7.3.5 Examination of LSH Higher Performers

LSH could not be used to split participants into high and low Gf ability groups. The relatively high level of difficulty of the set resulted in a floor effect and restriction of range. Those in the bottom 50% scored in the range of only 0-1, while those in the top 50% scored in the much wider range of 2-12. This restriction of range makes it inappropriate to collapse them into two separate groups and renders HLM less informative.

However, a relevant post-hoc question that arises from this is whether those who performed very well on LSH task also performed very well on the MSG Complex Item in the Complex
Control Condition. That is, are there those who are able to reason well inductively, without much across-item learning opportunity? And can they do so on more than one task? Specifically, can those who perform very well on a task such as LSH, where learning across the task is hypothesized to be minimized, do equally well on the Complex Control item where no across-item learning is provided?

*Figure 7.5.* Trimean MSG scores on Complex and PI items for Experimental and Control Conditions, for those in the top 5% of LSH.

The top 5% of participants on LSH were selected to be examined. Their performance on LSH was higher than that for the rest of the sample; their mean score on LSH was 5.27 (44%, SD = 1.49, n = 11), compared to the entire samples mean of 1.02 (8%, SD = 1.45, n = 169). However, their pattern of performance on the MSG task was not very different to that of the
rest of the sample (Figure 7.5). They appear to demonstrate a proactive interference effect and their trimean on the Complex Item (Control Condition) was 46, which is somewhat similar to the rest of the sample’s performance which was 45. The sample size for this extreme group was too small for statistical tests of these differences to be meaningful (n ≈ 4, per condition). However, their general pattern of performance suggests that despite their special level of performance on LSH, they did not display the same level of performance on MSG without across-item learning opportunity, and when across-item learning opportunity was provided, they seemed to benefit from it just as much as other participants. It also appears that they suffered from proactive interference on the PI item.

7.4 General Discussion and Summary of Results for Experiments 2, 3 and 4

Experiments 2, 3 and 4 focused on the distinction between knowledge that may be brought to the task, learning that occurs across multiple items in the task and induction within a single item that may be independent of any prior knowledge. The effect of proactive interference as a consequence of learning and knowledge was also investigated. The experiments examined which of these are relevant to general performance (i.e., common to everyone) and which contribute to individual differences. Finally, an additional question of interest was the effect of item ordering in Gf tasks on the learning opportunity provided in such tasks. Each of these issues will be discussed in detail, in light of the results from these studies.

Results from the current experiment (Experiment 4) were largely consistent with the results from Experiments 2 and 3. Rather than summarising the results for Experiment 4 separately (which would be verging on repeating the summary of results for Experiment 2 and 3), we shall present broad summary of all the important results from all three experiments together and their implications for the relevant hypotheses and predictions.
The Classical Gf Hypothesis

Overall, the Classical Gf Hypothesis, which was based on theorising in the differential literature - such as the works of Cattell (1987), Gottfredson (1997) and Spearman (1932), was partially supported. It predicted that in the absence of across-item learning and prior conceptual knowledge, induction on complex Gf items would still be possible (through the participant’s mappings between the concrete stimuli in the item, which can result in the acquisition of general concepts such as the rules in the item). Also those of higher Gf would outperform those of lower Gf (because they would be better at making the mentioned mappings). However, we suspected performance would be quite low, for all (because according to analogical thinking theory, acquiring unfamiliar concepts is very difficult). If performance is not low – this would imply that the use of prior conceptual knowledge is involved instead.

For the Complex Control Item, when no across-item learning opportunity was provided, performance was indeed generally very low for all Gf ability groups, in all experiments. That is on average, participants required a very large number of attempts (and analogues) before a series of correct responses was achieved. When viewed in isolation, one might conclude that this was due to the item’s difficulty level. However, in the Experimental Condition, when easier items led up to the Complex Item, performance improved dramatically for all groups, despite the complexity of the item staying the same. This suggested that the Complex Item is quite novel and that without easier, possibly more familiar items leading up to it (to allow for across-item learning opportunity), it was difficult for all Gf ability groups. The Classical Gf Hypothesis predicts that in the absence of across-item learning and prior conceptual knowledge, induction on complex Gf items would still be possible. In this case, it was possible – but success rate was very low. Furthermore, participants were allowed many
attempts at the item. Gf items do not conventionally allow as many attempts and analogues as the MSG task. Thus, it is questionable whether induction in complex, novel Gf items would be possible without across-item learning.

However, Gf ability was a significant predictor of performance on the Complex Control Item. That is, even when across-item learning opportunity was not available (and despite the floor effects), those of higher Gf outperformed those of lower Gf on the Complex Item. This is consistent with the component of the Classical Gf Hypothesis which states that regardless of whether there is across-item learning opportunity leading up to an item, those of higher Gf will outperform those of lower Gf.

Also, the Complex Control Item consistently correlated moderately with Gf markers presented easy-to-hard. This may indicate that the within-item induction in the Complex Control Item share processes in common with across-item learning. Indeed, in the introduction, we proposed that within-item induction and across-item learning may be similar in processes because they both seem to be examples of analogical thinking (Gick & Holyoak, 1983) – both involve the transfer of knowledge from one situation to another by a process of mapping. Across item learning may involve building upon what is already known (and hence, explicit and easy), learnt from previous items, to understand something previously unknown and novel (something less explicit and harder) in latter, more complex items. Within-item induction may involve the process of mapping between two concrete analogues, to learn something previously unknown and less explicit, through noticing regularities/similarities between analogues available in an item. The former involves the application of knowledge gained from previous items to a current item and the latter involves the application of knowledge learnt from mappings between analogues, to other analogues within a single item.
The similarity of these processes would explain the Complex Control Item’s correlation with Gf markers presented easy-to-hard.

**The General Learning Hypothesis**

The General Learning Hypothesis, which was based on the work of Sweller and Gee (1978) and Holland et al. (1989) from the cognitive literature, was supported. It predicted that Simpler items act as across-item learning opportunity for more complex items with similar rules. When this across-item learning opportunity is provided, performance on the complex item will improve greatly for all, compared to when learning opportunity is not provided. Mann-Whitney tests showed that in all experiments, all Gf groups benefited from across-item learning opportunity. This was supported by HLM analyses which were conducted in Experiments 3 and 4.

**The Individual Differences Learning Hypothesis**

The Individual Differences Learning Hypothesis, which was based on the works of Carlstedt et al. (2000) and Verguts and De Boeck (2002b) from the cognitive-differential literature, was largely supported. It predicted that when across-item learning opportunity is provided, those of higher Gf ability will improve more than those of lower Gf ability, due to those of higher Gf ability being better learners. The HLM models in Experiments 3 and 4 suggest that when across-item learning opportunity was provided those of higher Gf ability benefitted marginally significantly more from this learning opportunity (compared to their performances when no across-item learning was provided). While the effect was only marginally significant in both experiments, the fact that the effect was found in both experiments (and backed up by trimean trends in Experiment 2), suggests that it is a real effect rather than a type I error; and the marginal significance may be due to lack of power.
The General Proactive Interference Hypothesis and the Individual Differences Proactive Interference Hypothesis

The General Proactive Interference Hypothesis, based on the work of Sweller and Gee (1978), predicted that learning in Gf tasks produces proactive interference for items that contain dissimilar rules. The Individual Differences Proactive Interference Hypothesis, based on Unsworth and Engle (2005a) predicted that those of higher Gf are better at combating proactive interference than those of lower Gf. There was support for the former but not for the latter.

With regard to proactive interference, there was evidence from the HLM in Experiments 3 and 4 that those in the PI Control Condition had a better chance of success on the PI item than those in the Experimental Condition – demonstrating the proactive interference effect. However, there was no evidence that those of lower Gf ability suffer more from proactive interference. Overall, easy-to-hard markers correlated moderately with PI item in the PI Control Condition, but not in the Experimental Condition. One explanation for this is that those of higher Gf suffer more from proactive interference – which would reduce the experimental PI Item’s correlations with Gf.

Certainly, non-parametric Mann-Whitney tests from Experiments 3 and 4 suggest those of higher Gf (but not lower Gf) seem to suffer from proactive interference (at least when the marker was Letter Series presented easy-to-hard). That is, those of higher Gf may suffer more from proactive interference when faced with items of a different rule type because they learn more from items of the same rule type. However, the effect of proactive interference, when present, was always either weak or marginally significant. Nevertheless, there is certainly no evidence that those of lower Gf suffer more from proactive interference than those of higher
Gf. This was a surprising finding, because many studies have shown WMC accounts for much of the variance in Gf tasks and what seems to be important about WMC tasks is that they require the combat of proactive interference (Conway, Kane, & Engle, 2003; Engle et al., 1999).

A possible starting point for an explanation is given by Sternberg (1986). Some rules (concepts) are more entrenched (familiar) in one’s experience than other rules. Indeed, we believe that simpler rules also tend to be more familiar than complex rules. The PI Item was a simple item. It could be that its rule is equally entrenched in the minds of all participants and thus, the item was not very sensitive to differences in proactive interference.

Another potential explanation is that since those of higher Gf learn more across the task, they should display more of the effects of proactive interference than those of lower Gf ability. The fact that there was no significant difference in the amount of proactive interference shown, may suggest that those of higher Gf are indeed better at combating proactive interference.

_A Hypothesis Regarding the Methodological Issue of the Presentation Format of the Gf Marker:_

Gf tasks are usually presented in easy-to-hard order. It is possible that this presentation format plays a role in creating learning opportunity. Gf markers with only hard and moderately hard items were intended to capture performance on these Gf items when learning across the task is minimised. We predicted that they would have weaker relationships with any learning effects that we find. Unfortunately, the manipulation resulted in very low performance for the majority of participants and restriction of range and variance issues for these measures, which limited the type of analyses that could be run with them. It is possible
that this very low level of performance mirrors the very low performance for the Complex Control Item. That is, without much or any across-item learning opportunity, novel items are very difficult for all Gf ability groups.

We examined the top 5% of performers on these types of markers (where only moderately hard and hard items are presented). The rationale behind this was: if these types of markers minimise across-item learning opportunity and the performance on the Complex Control Item is free of any influence from across-item learning opportunity, then those who perform very well on these particular Gf markers may also be able to perform very well on the Complex Control Item. This hypothesis was not supported.

The top 5% on these types of Gf markers’ general pattern of performance suggested that despite their special level of performance on the Gf markers, they did not display the same level of performance on the Complex Control Item without across-item learning opportunity. Furthermore, when across-item learning opportunity was provided, they seemed to benefit from it just as much as other participants. This finding suggests that the ability to reason well without across-item learning opportunity, in traditional Gf items is task specific. That is, it may be due to knowledge of certain concepts that are brought to the task.

### 7.4.1 Implications for Processes Involved in Gf and Individual Differences

Thus, overall, the results from the three experiments:

- Partially supported the Classical Gf Hypothesis
- Supported the General Learning Hypothesis
- Largely supported the Individual Differences Learning Hypothesis.
- Largely supported the General Proactive Interference Hypothesis
Did not support the Individual Differences Proactive Interference Hypothesis.

The partial support of the Classical Gf Hypothesis suggests that induction in a complex Gf item, in isolation from knowledge learnt from previous items in the task or knowledge brought to the task is very difficult but possible. While it may be possible to solve a Gf item in the absence of across-item learning opportunity, the support found for the General Learning Hypothesis suggests that learning does occur in Gf tasks. Furthermore, the support for the Individual Differences Learning Hypothesis suggests that those of higher Gf learn more than those of lower Gf. That is, learning across the task is a source of individual differences in performance. The floor effects that appeared for the Complex Item in the absence of across-item learning opportunity and the restriction of range issues in markers that theoretically limited across-item learning opportunity, suggest that tasks are better able to measure Gf if they allow across-item learning opportunity. There was also some evidence to suggest that when across-item learning opportunity is not provided, performance may then depend on knowledge that participants bring to the task, which may make the task a less “fluid” measure.

While there is a difference between across-item learning and within-item induction that occurs in isolation from previous knowledge, the application of analogical thinking theory (Gick & Holyoak, 1983) suggests that they may be similar processes because both seem to be instances of analogical thinking. Indeed, the positive correlations between conditions/tasks that allowed for across-item learning opportunity and those that did not, suggests that they may indeed be similar processes.

The support for the General Proactive Interference Hypothesis but not the Individual Differences Proactive Interference Hypothesis suggests that proactive interference occurs in
Gf tasks (as a result of across-item learning), but does not contribute to individual differences in performance. That is, those of lower Gf ability do not suffer more from proactive interference.

7.4.2 Significance and Limitations of Experiments 2, 3 and 4

Experiments 2, 3, and 4 were significant for a number of reasons. Firstly, unlike Experiment 1, they were able to assess the amount learnt across-items without confounding this learning with whether participants were able to successfully solve items that were intended to provide learning opportunity. Providing more exposure to rules is providing potential learning opportunity but that learning opportunity may not be exploited unless the participant can answer the items correctly (Verguts & De Boeck, 2002b) – and thus fully comprehend the rule. The more items the participant answers correctly, the more actual learning opportunity may be available to them. It was found in the MSG task when actual learning opportunity was provided to all participants (through allowing participants multiple attempts and allowing the item to get theoretically easier with more attempts, until they are able to successfully solve it) all participants were able to learn but those of higher Gf were able to learn more. What this suggests is that the finding from Experiment 1 – that those of higher Gf benefitted more from across-item learning opportunity – was unlikely due to them simply being better at within-item induction and creating more actual learning opportunity for themselves. Even when the same amount of actual learning opportunity is available to all, as was the case in Experiments 2 through 4, those of higher Gf still benefitted more from this learning opportunity.

In addition to being able to assess across-item learning without confounding it with success of within-item induction, Experiments 2 through 4 were also able to assess within-item
induction isolated from any potential influences from across-item learning. This is because the MSG task allows single items to be administered in isolation from other items.

Furthermore, the use of HLM in experiments 3 and 4 allowed us to model the probability of success at each attempt, thereby providing us with an indication of the variability of performance for those of different levels of Gf within a single item. Hence, the MSG task was also able to allow us to ascertain (to a certain degree) whether induction on complex Gf items (in the absence of across-item learning) involves the use of prior knowledge brought to the test or rule discovery processes that occur in isolation from the use of prior knowledge. The very low levels of success on each attempt on the Complex Item, in the absence of across-item learning opportunity, suggested that complex Gf type items contain rules that are indeed novel to (at least the majority) of participants. Had the rules been more familiar to participants, one would expect a more similar pattern of performance to the Complex Item in the Experimental Condition. In this condition, participants were provided with learning opportunity across the test to familiarise them with the rule and this led to better levels of performance that incrementally improved with each attempt.

A limitation of the current experiments is that although the MSG task has Gf-like characteristics, it is not a conventional Gf task. However, we argued that it is sufficiently similar to conventional Gf tasks and this is supported by the fact that it consistently showed moderate, positive correlations with the Gf markers. Another limitation is that we only used single Gf tests (instead of a Gf factor) and these, being series completion tasks, were rather similar in nature to the MSG task. Future studies could employ multiple tests to more derive a more traditional Gf factor.
In conclusion, within-item induction processes can occur in complex items in isolation from prior knowledge and these processes contribute to individual differences in performance. Learning across items also occurs and it appears to contribute to individual differences in performance. This learning creates proactive interference for items that appear similar but contain different rule types. However, there was not strong enough evidence to suggest that the ability to combat this proactive interference contributes to individual differences in performance.
CHAPTER 8
DISCUSSION AND CONCLUSION

8.1 Summary of Aims and Hypotheses

The aim of this thesis was to gain a better understanding of Gf through investigating why people differ on Gf tasks. Specifically, this thesis examined whether learning processes occur in fluid intelligence (Gf) tests, whether it is essential for them to occur for induction to take place, and whether they contribute to individual differences in performance. Gf has been reliably identified by factor analysis as important in novel induction tasks but its exact nature is still rather poorly understood. Cognitive theories of inductive reasoning that are broad enough to be applicable to all Gf tasks emphasise the importance of prior conceptual knowledge in the inductive process (Holland et al., 1989; Sternberg, 1986). This tension between the view in the differential literature that Gf is a fluid ability to reason when faced with novel problems (Cattell, 1987) and the view in the cognitive literature that possession of relevant prior conceptual knowledge is important to success in induction acts as a potential barrier to understanding Gf.

We proposed that one way to reconcile the seemingly incompatible views is to consider that learning processes may occur across items within Gf tasks, and act as the mediator between the novelty in Gf items and the knowledge needed to solve them. We hypothesised that:

1) Complex Gf items contain novel relationships that are very difficult (or near impossible) to induce without guidance from some prior conceptual knowledge. This prior knowledge is (gradually) provided by earlier, easier, items which are more
familiar. Thus, learning is a necessary process in Gf tests (the “General Learning Hypothesis”).

2) Furthermore, learning is a source of individual differences in Gf test performance. That is, those of higher Gf ability benefit more from across-item learning opportunity than those of lower ability (the “Individual Differences Learning Hypothesis”).

We further hypothesised that

3) Learning leads to a build up of proactive interference.

4) Those of higher Gf ability must be better at combating proactive interference (because of Hypothesis 2 and 3).

The learning hypotheses and the proactive interference hypotheses have been framed as competing hypotheses for the link between Gf and WMC (Unsworth & Engle, 2005a), but we have argued that they may not be mutually exclusive and may actually go hand-in-hand.

8.2 Summary of Findings

In Experiment 1 (Chapter 4) we tested the learning hypotheses by varying the occurrence of the “distribution of two rule” in three versions of the Raven test. Experiment 1 revealed that it is not necessary for participants to learn about the distribution of two rule from simple items, in order for them to solve items containing complex instantiations of the rule later in the test. That is, the results suggested that rule learning is not necessary for induction to take place. However, the results also suggested that rule learning occurs in Raven and contributes to individual differences in Raven performance, with those of higher Gf benefitting more from more exposure to the distribution of two rule.
Experiment 1 also highlighted an unexpected issue that we pursued further in Experiments 2, 3 and 4. The results suggested that in order for participants to be able to learn from an item, they need to be able to get it correct - perhaps in order to get feedback that their hypothesis about the nature of the item is correct (Verguts & De Boeck, 2002a). Secondly, this may mean that those of lower Gf ability may not be able to learn as much from items simply because they are not able to get as many earlier items correct and not because they are less able to learn about rules across the test. That is, across-item learning may be confounded with within-item induction.

Experiments 2, 3 and 4 (Chapters 5 – 7) investigated the learning hypotheses but addressed the potential confound between across-item learning and within-item induction through the MSG task. The proactive interference hypotheses were also examined. It was found that within-item induction processes can occur in complex items in isolation from prior knowledge. However, without across-item learning opportunity, complex items are very difficult for all Gf ability groups. Across-item learning also occurs and appears to contribute to individual differences in performance. This learning creates proactive interference for items that appear similar but contain different rule types. However, there was not strong enough evidence to suggest that the ability to combat this proactive interference contributes to individual differences in performance. There was also some evidence to suggest that the ability to reason well without across-item learning opportunity in traditional Gf items is task specific. That is, it may be due to knowledge of certain concepts that are brought to the task.
8.3 Implications for Processes Involved in Gf tasks

8.3.1 Complex Items: Rule Discovery and Learning Processes

The findings from Experiments 1 to 4 suggest that learning processes are important to induction in Gf tasks as well as to individual differences in Gf tasks. Gf items need to be novel to tap into “fluid” abilities as opposed to “crystallised” abilities. There was evidence in our experiments that suggested that performance on complex, novel items requires some guidance from knowledge provided by the successful solution of earlier, simpler items (which are arguably, more familiar – we shall examine this assumption further, later on). This may be because without some prior relevant knowledge, induction involves a lot of “trial and error” guessing. Thus, for Gf tasks to be both novel and meaningful measures that reliably tap into fluid ability (rather than random guessing), they must involve learning processes that build upon existing knowledge.

According to Holland, Holyoak et al. (1989), induction cannot take place without previous knowledge. Due to the nature of inductive reasoning (reasoning in the absence of complete information), the solution to inductive problems can only be considered plausible (rather than correct) at best. And whether a solution could be characterized as plausible can be determined only with reference to the participant’s current knowledge. Indeed, the stimuli in a Gf item exist with various objective relations and associations between them, with the one considered the “correct relation” by the test-maker often only one of many other potentially correct relations (Passmore, 1935). Sweller and Gee (1978) highlight that prior knowledge may be more important in complex items than in simple ones. They argue that the more complex the rule (relationship) in an item, the larger the number of other potentially correct rules may also be observed in an item. Thus, when items are complex, the correct rule could be hard to detect amongst all the other potentially correct rules unless they are more familiar. Across-
item learning may be necessary to make the correct rule more familiar through a scaffolding process. Indeed, it was found that learning does appear to occur in Gf tests and those of higher Gf learn more than those of lower Gf.

However, it was also found that in the absence of across-item learning opportunity, induction was still possible for some although, a) their success rate was very low, and b) those of higher Gf outperformed those of lower Gf. The low success rates may have been due to the reasons outlined by Passmore (1935) and Sweller and Gee (1978) mentioned previously. However, induction may have still been possible because of the mapping process typical in Gf items. According to (Gick & Holyoak, 1983) a general concept (such as a Gf item rule) can be learnt through the mapping between two analogues (such as the stimuli in a Gf item) -- that is through noticing regularities and similarities between analogues available in an item.

While it is possible to induce the solution to a complex, novel, Gf item in the absence of across-item learning opportunity, conventional Gf tests **do provide across test learning opportunity** through the easy-to-hard presentation format. Thus, it is likely that the solution process in a Gf item involves a combination of within-item mapping of analogues and the application of knowledge learnt across items.

Also, the processes of across-item learning and within-item induction (in the absence of any prior knowledge brought to the item) may be highly related. There was empirical evidence to suggest that those who are good at one are also good at the other. That is, there were positive relationships between measures that involved across-item learning opportunity and those that did not or minimised such opportunities. However, these correlations were not perfect, suggesting that these processes are not identical. Thus, Gf measures that are presented in easy-to-hard order may tap into something slightly different to Gf measures that are presented
in random order. It would be interesting to see in future studies whether tasks with different item orders form different factors.

The findings from this thesis have implications for both classic and modern test theory, which make the assumption that conditional on person ability and item difficulty, items within a test are independent of each other – that is, items can in principle be presented in any order without changing the nature of the construct being measured. However, since learning occurs from one item to another, and if this learning is an ability different and not perfectly correlated with traditional Gf reasoning, the assumption of item independence is contradicted (D. P. Birney & Sternberg, 2006).

8.3.2 What About Early Simple Items?

This thesis has mainly focussed on complex Gf items that occur later in the task. This is because they are more likely to contribute to overall individual differences in performance on the overall test. A neglected topic of discussion is what processes may occur in early, simpler items; most importantly, the very first item that is not preceded by any items. It was put forward in earlier chapters in this thesis that in the absence of across-item-learning, within-item induction may involve the application of inferential rules that the participant brings to the test situation. The alternative was also put forward: within-item induction may involve the discovering of the rules, which may have been totally novel to the participant, through a process of analogical mapping of the analogues provided within an item. We have worked with the assumption that early items contain familiar relationships that we all come across in some form in our lives. This is because some theorists believe that a concept cannot be recognised unless we already possess some prior knowledge of those concepts (Ohlsson & Lehtinen, 1997). For example, insight problems are not difficult to solve once we have been told about how to properly represent the problem (Sternberg & Davidson, 1995). However,
this is only a supposition based on our own induction. It is possible that the earliest Gf items involve induction independent of prior knowledge brought to the test. This is not a question we empirically examined.

However, the earliest items tend to have high success rates - performance on them do not contribute to individual differences. It is possible that knowledge application processes are involved for some participants and rule discovery processes for others. As long as performance on these early items are sufficiently high enough for all participants to learn from them to have a chance at solving subsequent items, which processes are actually involved probably does not really matter.

However, Gf tests may contain biases if early items can only be solved with the application of knowledge brought from outside of the test situation which is possessed by some participants, but not by others. This may be a question worth examining in future studies.

### 8.3.3 What about Proactive Interference?

Experiments 2 – 4 revealed that proactive interference occurs as a consequence of across-item learning but it did not appear to affect those of lower Gf more than those of higher Gf. In fact, there was some evidence to suggest that those of higher Gf may suffer more from proactive interference. Sweller and Gee (1978) explain why this might be so. Proactive interference occurs when the participant encounters a rule that is very dissimilar to all the previous rules that they had previously encountered. According to Sweller and Gee (1978), this is because participants continue testing rules similar to the ones they previously encountered. The more they learn from previous rules, the more entrenched those types of rules become and the more they suffer from proactive interference.
However, our studies did not investigate the proactive interference question very thoroughly. It is possible that the MSG task was not able to replicate the conditions that actually occur in Gf tests that require the combating of proactive interference – in such conditions those of higher Gf may be better at combating proactive interference than those of lower Gf. For example, we investigated proactive interference using a very simple (arguably, familiar) item, whose rule may be well entrenched in the knowledge of all participants. Those of higher Gf may be better at combating proactive interference in more complex, novel items, whose rules are not as well entrenched. This may be a question worthy of future investigation.

### 8.4 Implications for Cognitive Theories

In Chapter 3 we outlined three broad theories of induction which we thought would be helpful to our quest to better understand the cognitive processes involved in Gf tasks. These were the theories of Sternberg (1986), Holland et al. (1989) and Spearman (1932). These theories have different emphases on what is important in induction. Sternberg emphasizes the use of prior conceptual knowledge, Holland emphasizes conceptual knowledge and learning and Spearman emphasizes the eduction of relations and correlates, which seem to occur in isolation from previous knowledge and learning. Only Spearman is consistent with the dominant, conventional novelty view of Gf, which sees it as largely uninfluenced by prior experience and knowledge.

While at first glance they seem to conflict to varying degrees, our conclusion is that these three theories describe different aspects of the Gf induction process. Complex Gf items are indeed novel. However, the possession of relevant prior conceptual knowledge is important to induction success. Learning occurs in Gf tests and provides participants with the relevant, prior knowledge to successfully solve complex Gf items. The amount learnt contributes to
individual differences in performance (with those of higher Gf are able to learn more than those of lower Gf).

We do not believe that these are the only processes involved in Gf tasks. Other theories of reasoning emphasise the importance of other processes such as working memory capacity (Kyllonen & Christal, 1990). It is likely that the processes emphasised in different theories are all important to Gf performance at different levels. For example, learning processes may require working memory capacity. That is, learning processes may be the link between the high correlations between working memory capacity and Gf tasks (Verguts & De Boeck, 2002b).

Many researchers have recommended that differential psychology take note of process theories from cognitive psychology (Cronbach, 1957; Deary, 2001; Lohman, 2000). However, as mentioned in Chapter 2 of this thesis, mapping process theories onto ability constructs is difficult because it is not always obvious how cognitive theories (developed to explain general, average performance) might be applicable to explanations of individual differences in Gf tasks. We hope that we have shown that it is possible to use insights from cognitive psychology to increase our understanding of Gf. However, we would like to emphasise that it required theory synthesis and modification. That is, theories from cognitive psychology may require adjustments when they are used for the purposes of explaining processes involved in differential psychology’s ability constructs.

### 8.5 Implications for Gf

The research from this thesis has two major implications for understanding the Gf construct. Firstly, it has implications for how Gf is conceptualised. In the CHC theory of abilities (McGrew 2005), the most widely accepted organisation of human abilities (Alfonso et al., 2005), “Learning Abilities” are not listed under Gf, but are listed as narrower abilities under
Short-Term Memory (Gsm) and Long-Term Storage and Retrieval (Glr), and are considered poorly defined by existing research (McGrew, 2005). According to Carroll (1993) it has often been proposed that an important aspect of intelligence is the ability to learn, but methodological issues and insufficient data has made it difficult to convincingly demonstrate this relation. Evidence from this thesis suggests that (pending further investigation and evidence) Gf should be reconceptualised to include learning processes.

Secondly, the evidence from our research (potentially) sheds some light on why Gf is predictive of real world performance. Learning processes may be one of the links between Gf test performance and real world performance. That is, Gf measures may be predictive of real world success partly because it taps into learning which is important to performance in real world learning, such as in formal education, vocational training or learning practical knowledge about the world.

8.6 Conclusion

In conclusion, learning processes are important to induction in Gf tasks. Learning across the test occurs, learning needs to occur for induction to take place properly (as opposed to induction which involves some element of random guessing), and learning contributes to individual differences in performance on Gf tasks. It has been commonly assumed that conventional Gf tests do not allow for feedback nor opportunity to learn (Sternberg, Grigorenko et al. 2002), however, we have shown evidence to the contrary. Learning appears to be an important aspect of induction in Gf tasks and hence, fluid intelligence. Thus far, it has been a neglected focus of research in investigations of Gf, but we conclude that it is worthy of further investigation.
CHAPTER 9

REFERENCES


APPENDIX A

A.1 Items Generated for the Raven Learning Sets

Every attempt was made to ensure that they would be as similar as possible to the original APM items they replaced. Things taken into account when generating items included:

- The way the items looked (that is, the type of shapes/figures used)
- Number of rules in each item
- Types of rules included in the item
- Note: Instantiations of d2 rules require a minimum of 3 tokens. Any fewer and one of the entries will appear blank. Hence, some generated d2 rule items contained more rule tokens than the items they replaced.

Item 7, No Exposure condition.
Item 8, No Exposure condition.

Item 11, No Exposure condition.
Item 12, No Exposure condition.

Item 3, Most Exposure condition.
Item 4, Most Exposure condition.

Item 5, Most Exposure condition.
Item 6, Most Exposure condition.

Item 9, Most Exposure condition.
Item 13, Most Exposure condition.

Item 3, Transfer Set.
A.2 Characteristics of Learning Set items used in each condition and their equivalents in the original Raven (Advanced Progressive Matrices)

<table>
<thead>
<tr>
<th>Item No.</th>
<th>Original APM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Proportion</td>
</tr>
<tr>
<td></td>
<td>Correct b</td>
</tr>
<tr>
<td>1</td>
<td>0.98 d3 c</td>
</tr>
<tr>
<td>2</td>
<td>0.91 p c</td>
</tr>
<tr>
<td>3</td>
<td>0.90 p c</td>
</tr>
<tr>
<td>4</td>
<td>0.88 p c</td>
</tr>
<tr>
<td>5</td>
<td>0.79 d3 c</td>
</tr>
<tr>
<td>6</td>
<td>0.69 d3 c</td>
</tr>
<tr>
<td>7</td>
<td>0.50 d2 c</td>
</tr>
<tr>
<td>8</td>
<td>0.49 d2 c</td>
</tr>
<tr>
<td>9</td>
<td>0.38 p d3</td>
</tr>
<tr>
<td>10</td>
<td>0.29 d3 c</td>
</tr>
<tr>
<td>11</td>
<td>0.24 d3 d2</td>
</tr>
<tr>
<td>12</td>
<td>0.17 d3 d2</td>
</tr>
<tr>
<td>13</td>
<td>0.13 d3 c</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Item No.</th>
<th>Original APM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Proportion</td>
</tr>
<tr>
<td></td>
<td>Correct b</td>
</tr>
<tr>
<td>1</td>
<td>0.5 d3 c</td>
</tr>
<tr>
<td>2</td>
<td>0.86 p c</td>
</tr>
<tr>
<td>3</td>
<td>0.84 p c</td>
</tr>
<tr>
<td>4</td>
<td>0.74 d3 c</td>
</tr>
<tr>
<td>5</td>
<td>0.76 d3 c</td>
</tr>
<tr>
<td>6</td>
<td>0.50 d2 c</td>
</tr>
<tr>
<td>7</td>
<td>0.57 d2 c</td>
</tr>
<tr>
<td>8</td>
<td>0.49 p d3</td>
</tr>
<tr>
<td>9</td>
<td>0.42 d3 c</td>
</tr>
<tr>
<td>10</td>
<td>0.50 d3 d2</td>
</tr>
<tr>
<td>11</td>
<td>0.43 d3 d2</td>
</tr>
<tr>
<td>12</td>
<td>0.49 d3 c</td>
</tr>
</tbody>
</table>

A.3 Transfer Set item characteristics

<table>
<thead>
<tr>
<th>Item No.</th>
<th>Original APM a</th>
<th>APM Proportion</th>
<th>Correct b</th>
<th>Rule 1</th>
<th>Rule 2</th>
<th>No. of rule tokens (per row)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>35</td>
<td>0.14</td>
<td>d2</td>
<td>c</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>36</td>
<td>0.05</td>
<td>d2</td>
<td>c</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>*</td>
<td>n/a</td>
<td>d3</td>
<td>d2</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

Notes: d3 = distribution of 3, c = constant, d2 = distribution of 2.
Rule taxonomy based on Carpenter, Just et al. (1990)
(*) = item constructed by the author.
a. = taken from Raven (1962)

A.4 Regression model for variables predicting performance on Learning Set

\[ Y = b_0 + b_1 X_1 + b_2 X_2 + b_2 X_2 + b_4 X_2 X_3 + b_5 X_2 X_3 + r \]

<table>
<thead>
<tr>
<th>Model</th>
<th>B</th>
<th>SE(B)</th>
<th>β</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant (b_0)</td>
<td>9.99</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>Condition_Dummy1 (b_1)</td>
<td>-1.30</td>
<td>0.37</td>
<td>-0.20*</td>
</tr>
<tr>
<td>Condition_Dummy2 (b_2)</td>
<td>-2.12</td>
<td>0.33</td>
<td>-0.38*</td>
</tr>
<tr>
<td>Number_Series Centered (b_3)</td>
<td>0.15</td>
<td>0.04</td>
<td>0.22*</td>
</tr>
<tr>
<td>Condition_Dummy1 X Number_Series Centered (b_4)</td>
<td>0.14</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>Condition_Dummy2 X Number_Series Centered (b_5)</td>
<td>0.15</td>
<td>0.09</td>
<td>0.11</td>
</tr>
</tbody>
</table>

\[ R^2 \] 0.24
\[ F \] 16.42**

Notes: b_0 = est. marginal mean of No condition; b_1 = est. marginal mean of Limited condition - b_0; b_2 = estimated marginal mean of Most condition - b_0.
**p < .01.
A.5 Working out for probabilities of chance level performance for Raven's APM Transfer Set

For each item there are 7 wrong options and 1 correct option. So there is 1/8 (or 12.5%) chance of getting an item correct by chance/guessing alone. Conversely, there is a 7/8 chance of getting an item wrong if one is guessing alone. Therefore, by guessing alone, the chance of getting all three Transfer Set items incorrect is:

$$\frac{7}{8} \times \frac{7}{8} \times \frac{7}{8} = \frac{343}{512}.$$

Therefore, the chance of not getting all three items wrong is:

$$1 - \left(\frac{343}{512}\right) = \frac{169}{512} = 0.33 \text{ or } 33\%$$

This 0.33 indicates that by chance and guessing alone, one is bound to get 1 item correct out of 3.

A.6 Regression model for variables predicting performance on Transfer Set

$$(Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_1X_2 + b_5X_1X_3 + \epsilon)$$

<table>
<thead>
<tr>
<th>Model</th>
<th>B</th>
<th>SE(B)</th>
<th>β</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant ($b_0$)</td>
<td>1.12</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>Condition Dummy1 ($b_1$)</td>
<td>-0.04</td>
<td>0.13</td>
<td>-0.02</td>
</tr>
<tr>
<td>Condition Dummy2 ($b_2$)</td>
<td>-0.07</td>
<td>0.12</td>
<td>-0.04</td>
</tr>
<tr>
<td>Number Series Centered ($b_3$)</td>
<td>0.06</td>
<td>0.01</td>
<td>0.27**</td>
</tr>
<tr>
<td>Condition Dummy1 X Number Series Centered ($b_4$)</td>
<td>0.03</td>
<td>0.03</td>
<td>0.06</td>
</tr>
<tr>
<td>Condition Dummy2 X Number Series Centered ($b_5$)</td>
<td>0.04</td>
<td>0.03</td>
<td>0.09</td>
</tr>
</tbody>
</table>

$R^2 = 0.12$

$F = 7.75**$

Notes: $b_0$ = est. marginal mean of No condition; $b_1 = est. marginal mean$ of Limited condition - $b_0$; $b_2 = estimated marginal mean$ of Most condition - $b_0$. **p < .01.
B.1 **MSG Items with all of their correct target-answer pairs and rules.**

**Item 1:**
- 1st digit of correct response identical to 1st digit of target number.
- 2nd digit of correct response always “8”.
- 1 pair = 1 analogue

<table>
<thead>
<tr>
<th>Record of correct</th>
<th>Target-Answer pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>14 - 18</td>
<td>12 - 18</td>
</tr>
<tr>
<td>32 - 38</td>
<td>13 - 18</td>
</tr>
<tr>
<td>54 - 58</td>
<td>34 - 38</td>
</tr>
<tr>
<td>44 - 48</td>
<td>45 - 48</td>
</tr>
<tr>
<td>15 - 18</td>
<td>31 - 38</td>
</tr>
<tr>
<td>25 - 28</td>
<td>33 - 38</td>
</tr>
<tr>
<td>52 - 58</td>
<td>35 - 38</td>
</tr>
<tr>
<td>24 - 28</td>
<td>43 - 48</td>
</tr>
<tr>
<td>11 - 18</td>
<td>21 - 28</td>
</tr>
<tr>
<td>42 - 48</td>
<td>23 - 28</td>
</tr>
<tr>
<td></td>
<td>34 - 38</td>
</tr>
<tr>
<td></td>
<td>14 - 18</td>
</tr>
</tbody>
</table>
Item 2:
- 1st digit of correct response identical to 1st digit of target number.
- 2nd digit of correct response alternates between “7” and “8”;
- 2 pairs = 1 analogue

<table>
<thead>
<tr>
<th>Record of correct Target-Answer pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>14 - 17</td>
</tr>
<tr>
<td>32 - 38</td>
</tr>
<tr>
<td>54 - 57</td>
</tr>
<tr>
<td>44 - 48</td>
</tr>
<tr>
<td>15 - 17</td>
</tr>
<tr>
<td>25 - 28</td>
</tr>
<tr>
<td>52 - 57</td>
</tr>
<tr>
<td>24 - 28</td>
</tr>
<tr>
<td>11 - 17</td>
</tr>
<tr>
<td>42 - 48</td>
</tr>
</tbody>
</table>

Item 3:
- 1st digit of correct response identical to 1st digit of target number, but increases by 1 when 2nd digit of correct answer is “0”.
- 2nd digit of correct response alternates between “5”, “2” and “0”:
- 3 pairs = 1 analogue.

<table>
<thead>
<tr>
<th>Record of correct Target-Answer pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>14 - 15</td>
</tr>
<tr>
<td>32 - 32</td>
</tr>
<tr>
<td>54 - 60</td>
</tr>
<tr>
<td>44 - 45</td>
</tr>
<tr>
<td>15 - 12</td>
</tr>
<tr>
<td>25 - 30</td>
</tr>
<tr>
<td>52 - 55</td>
</tr>
<tr>
<td>24 - 22</td>
</tr>
<tr>
<td>11 - 20</td>
</tr>
<tr>
<td>42 - 45</td>
</tr>
</tbody>
</table>
### Item 4:
- 1st digit of correct response alternates between “6”, and “1”.
- 2nd digit of correct response identical to 2nd digit of target, but decreases by 1 when 1st digit of correct response is “1”
- 2 pairs = 1 analogue

<table>
<thead>
<tr>
<th>Record of correct Target-Answer pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>14 - 64</td>
</tr>
<tr>
<td>32 - 11</td>
</tr>
<tr>
<td>54 - 64</td>
</tr>
<tr>
<td>44 - 13</td>
</tr>
<tr>
<td>15 - 65</td>
</tr>
<tr>
<td>26 - 14</td>
</tr>
<tr>
<td>52 - 62</td>
</tr>
<tr>
<td>24 - 13</td>
</tr>
<tr>
<td>11 - 61</td>
</tr>
<tr>
<td>42 - 11</td>
</tr>
</tbody>
</table>

### Transfer item:
- 1st digit of correct response alternates between “9”, “1” and “5”.
- 2nd digit of correct response identical to 2nd digit of target, but doubled when 1st digit of correct response is “5”
- 3 pairs = 1 analogue

<table>
<thead>
<tr>
<th>Record of correct Target-Answer pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>14 - 94</td>
</tr>
<tr>
<td>32 - 12</td>
</tr>
<tr>
<td>54 - 58</td>
</tr>
<tr>
<td>44 - 94</td>
</tr>
<tr>
<td>15 - 15</td>
</tr>
<tr>
<td>25 - 510</td>
</tr>
<tr>
<td>52 - 92</td>
</tr>
<tr>
<td>24 - 14</td>
</tr>
<tr>
<td>11 - 52</td>
</tr>
<tr>
<td>42 - 92</td>
</tr>
</tbody>
</table>
**PI item:**
- Add 2nd digit of target onto the total 2 digit number.
- 1 pair = 1 analogue

<table>
<thead>
<tr>
<th>Record of correct Target-Answer pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>14 - 18</td>
</tr>
<tr>
<td>32 - 34</td>
</tr>
<tr>
<td>54 - 58</td>
</tr>
<tr>
<td>44 - 48</td>
</tr>
<tr>
<td>15 - 20</td>
</tr>
<tr>
<td>25 - 30</td>
</tr>
<tr>
<td>52 - 54</td>
</tr>
<tr>
<td>24 - 28</td>
</tr>
<tr>
<td>11 - 12</td>
</tr>
<tr>
<td>42 - 44</td>
</tr>
<tr>
<td>12 - 14</td>
</tr>
<tr>
<td>55 - 60</td>
</tr>
<tr>
<td>41 - 42</td>
</tr>
<tr>
<td>51 - 52</td>
</tr>
<tr>
<td>43 - 46</td>
</tr>
<tr>
<td>33 - 36</td>
</tr>
<tr>
<td>53 - 56</td>
</tr>
<tr>
<td>45 - 50</td>
</tr>
<tr>
<td>31 - 32</td>
</tr>
<tr>
<td>24 - 28</td>
</tr>
<tr>
<td>23 - 26</td>
</tr>
<tr>
<td>34 - 38</td>
</tr>
<tr>
<td>22 - 24</td>
</tr>
<tr>
<td>53 - 56</td>
</tr>
<tr>
<td>42 - 44</td>
</tr>
<tr>
<td>21 - 22</td>
</tr>
<tr>
<td>22 - 24</td>
</tr>
<tr>
<td>15 - 20</td>
</tr>
<tr>
<td>41 - 42</td>
</tr>
<tr>
<td>25 - 30</td>
</tr>
<tr>
<td>35 - 40</td>
</tr>
<tr>
<td>21 - 22</td>
</tr>
<tr>
<td>44 - 48</td>
</tr>
<tr>
<td>43 - 46</td>
</tr>
</tbody>
</table>