

TASK STRUCTURE AND STRATEGY VARIATION IN COMPONENTIAL TASK ANALYSIS

David Alan Conradie

Degree awarded with distinction 11 December 1991

A dissertation submitted to the Faculty of Arts, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements of the Degree of Masters of Arts.

Department of Industrial Psychology
University of the Witwatersrand

ABSTRACT

Technological advancement has changed the nature of interactions within modern man-machine systems, with workload becoming more a load on the mental capacities of the human systems component than on his/her physical capabilities. Moreover, increased system capabilities may have disproportionately reduced the time available to perform these mental operations, thereby adding new and more complex cognitive tasks. A concept crucial to the evaluation of modern man-machine systems is that of the resulting implications of loading and overloading the cognitive capabilities of the human systems component, namely the measurement of mental workload.

For the assessment of mental workload experienced by individuals performing complex cognitive tasks, analytical access is dependent upon an accurate and robust model of task demands. Models of this nature are to be found within the inductive reasoning paradigm, particularly in the domain of analogical reasoning (e.g., Evans, 1968; Sternberg, 1977; Mulholland, Pellegrino and Glaser, 1980). Sternberg's (1977) pioneering attempt to specify the performance components in forced choice geometric analogy solution, in particular the linear componential models he proposed, constituted the basis for the empirical and theoretical issues explored in this research.

The present research design attempted to compensate for some of the flaws inherent in contemporary componential analysis and mental workload research, e.g., the use of small numbers of hypothesis and research sophisticated subjects (e.g., Mulholland, Pellegrino and Glaser, 1980). The design itself is not unusual, but the controls for experimental demand characteristics and the concern to avoid implicit strategy priming and response set creation were important steps to minimise further extraneous influences.

By contrast, the analysis was both novel and explanatory. The application of an appropriate cluster technique offered a significant development in the exploration of the information processing components involved in, and strategies employed for cognitive task solution. Furthermore, the investigation of performance on a stable cognitive phenomenon, i.e., the forced choice geometric analogy task, under conditions of varying complexity and uncertainty, provided valuable insights into the influence of mental workload on cognitive task performance.

A certain amount of time sharing was found to take place in experimental task solution. This revelation was of great significance in the context of the present research, since the notion of time sharing is a rejection of a rigid sequential strategy to cognitive task solution, in favour of a more integrated strategy involving interactive information processing.

Based on the present research evidence it appears that the amount of mental workload imposed by the same actual task demand is a function of cognitive capability and strategic flexibility. Furthermore, the mechanism of strategy choice was identified as being central to the level of cognitive task performance.

The nature of the present experimental design and procedure was such that it produced findings which could have significant implications for future componential analysis and mental workload research, as well as for the evaluation of work involving cognitive task performance.

DECLARATION

I declare that this is my own unaided work, except where due acknowledgement is made to others. It is being submitted for the Degree of Master of Arts to the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.



David Alan Conradie

25th day of MARCH 1991

ACKNOWLEDGEMENTS

I would like to express my appreciation and gratitude to the following people, all of whom made this dissertation possible.

Dr James Fisher, my supervisor, for the guidance and encouragement provided throughout the course of this dissertation.

The staff of the University of the Witwatersrand who prepared the experimental slides and set up the Research Laboratory.

The staff and students of the University of the Witwatersrand, friends and colleagues who participated in the research.

Karen, for her patience, understanding and assistance during the course of this dissertation.

My parents, without whose assistance I would have been unable to carry out this research.

Finally, Sara, Rosel and Denise, for their endurance and care in the typing of this dissertation.

The financial assistance of the Institute for Research Development towards this research is hereby acknowledged. Opinions expressed in this work, or conclusions arrived at, are those of the author and are not to be attributed to the Institute for Research Development.

PART 1

CONCEPTUAL AND THEORETICAL BACKGROUND

CHAPTER 1

INTRODUCTION	1
--------------------	---

CHAPTER 2

THE NATURE OF COGNITIVE WORK	4
------------------------------------	---

The Concept of Mental Workload	5
Determinants of Mental Workload	10
Task Orientated Factors	11
Operator Orientated Factors	19
Areas of Application for Workload Measures	22
Mental Workload Measurement Techniques	23
Method Sensitivity	25
Method Intrusion	25
Primary Task Measures	26
Spare Mental Capacity (Secondary Task) Measures	29
Physiological Measures	33
Subjective Opinion Measures	36
Chapter Summary	40

CHAPTER 3

COGNITIVE TASK STRUCTURE AND STRATEGY	43
---	----

COGNITIVE ABILITIES AND PSYCHOMETRICS	43
---	----

Cognitive Style and Strategy Choice	48
---	----

Assumptions of Psychometrics and Strategy Choice	49
--	----

INVESTIGATING THE NATURE OF MENTAL ABILITIES	51
--	----

THE DIFFERENTIAL APPROACH TO COGNITIVE ABILITIES	52
--	----

Multiple Criteria	52
-------------------------	----

Principles of Factor Analysis	53
-------------------------------------	----

The Purpose of Factor Analysis	54
--------------------------------------	----

Fundamental Unit of Analysis	55
------------------------------------	----

The Objective of Factor Analysis	55
--	----

Major Decisions to be made in Factor Analysis	56
---	----

Method of Factor Analysis	56
---------------------------------	----

Number of Factors	56
-------------------------	----

Type of Rotation	57
------------------------	----

The Decline of the Differential Approach	59
--	----

Misuses of Factor Analysis	59
----------------------------------	----

Limitations of Factor Analysis	61
--------------------------------------	----

THE INFORMATION PROCESSING APPROACH TO COGNITIVE ABILITIES	63
The Fundamental Unit of Analysis	63
 PRINCIPLE ORGANISING CONCEPTS WITHIN THE INFORMATION PROCESSING APPROACH	64
The Structure - Process Definition	64
Structure of Memory	66
Operations and Stages	67
INFORMATION PROCESSING METHODOLOGY	69
Artificial Intelligence versus Simulation	69
Slow Process versus Fast Process Research	70
The Subtraction Method versus The Additive - Factor Method	71
Advantages of the Information Processing Approach	72
Limitations of the Information Processing Approach	74
Computer Simulation	74
Mathematical Models of Response Time	75
The Subtraction Method	75
The Additive Factor Method	79
Common Limitations	80
 CONTEMPORARY APPROACHES TO UNDERSTANDING COGNITIVE ABILITIES	82
THE COGNITIVE CORRELATES APPROACH	82
THE COGNITIVE COMPONENTS APPROACH	84
The Objectives of Componential Analysis	85
Fundamental Unit of Analysis	85
Componential Theories and Models	87
Individual Differences	88
Structure of Componential Analysis	92
Intensive Task Analysis	94
Composite Task or Test Score	94
Interval Scores	94
Component Scores	99
Reference Ability Scores	100
Integration: Structural Regression Model	104
Overview	107
Extensive Task Analysis	107
Advantages of Componential Analysis	109
Problems with Componential Analysis	112
Potential Methodological Problems	112
 STERNBERG'S COMPONENTIAL THEORY OF ANALOGICAL REASONING	115
Nature of Analogy	117
The Componential Theory	117
Overview	117
Components	118
Combination Rule	121

Componential Models of Analogical Reasoning	122
Four Basic Models	122
Two Basic Strategies in Application	130
Sequential Option Scanning	130
Alternating Option Scanning	131
Scope of the Models	132
Evaluation of Sternberg's Componential Theory	134
Further Comments on Sternberg's Theory of Analogy	135
Chapter Summary	138

PART 2

THE PRESENT STUDY

CHAPTER 4

RATIONALE AND PROCEDURE	139
Conceptual and Methodological Issues	139
The Measurement of Mental Workload Associated with forced Choice	
Geometric Analogy Task Performance	143
Subjective Influences on Cognitive Task Performance	147
The Research Design of Sternberg's (1977) Geometric	
Analogy Experiment and it's Relevance to the Present Reser	149
Assessment of Subjective Influences	152
Subjective Influences and Cluster Analysis	152
Subjective Strategy Report	160
Task Complexity Ranking	161
Aims of the Present Research	161
Hypothesis	162
Subjects	162
Design and Apparatus	163
Procedure	166
The Nature of the Data Yielded	173
Chapter Summary	174

CHAPTER 5

DATA ANALYSIS AND RESULTS	175
Overview to Data Analysis	175
The Main Experiment	179
Stimulus Complexity and the Experimental Task	185
Solution Strategy and the Experimental Task	197
Degenerate Forced Choice Geometric Analogy Subjective	
Task Strategy Types	198
Semidegenerate Forced Choice Geometric Analogy (Format A:A::B:B)	
Subjective Task Strategy Types	200
Semidegenerate Forced Choice Geometric Analogy (Format A:B::A:B:)	
Subjective Task Strategy Types	201
Nondegenerate Forced Choice Geometric Analogy Subjective Task	
Strategy Types	202

Degenerate Forced Choice Geometric Analogy Items	204
Semidegenerate (A:A::B:B) Forced Choice Geometric Analogy Items	208
Semidegenerate (A:B::A:B) Forced Choice Geometric Analogy Items	210
Nondegenerate Forced Choice Geometric Analogy Items	213
 The Cluster Analysis Procedures	 218
Degenerate Forced Choice Geometric Analogy Cluster Solutions	223
Semidegenerate (A:A::B:B) Forced Choice Geometric Analogy Cluster Solutions	229
Semidegenerate (A:B::A:B) Forced Choice Geometric Analogy Cluster Solutions	237
Nondegenerate Forced Choice Geometric Analogy Cluster Solutions	244
 Chapter Summary	 255
 CHAPTER 6	
DISCUSSION	257
Overview to Discussion of Results	257
Stimulus Complexity and Task Performance	259
 Stimulus Complexity and Degenerate Geometric Analogy Task Performance	 261
Stimulus Complexity and Semidegenerate (A:A::B:B) Geometric Analogy Task Performance	262
Stimulus Complexity and Semidegenerate (A:B::A:B) Geometric Analogy Task Performance	263
Stimulus Complexity and Nondegenerate Geometric Analogy Task Performance	264
 Task Structure and Solution Strategy	 269
Degenerate Geometric Analogy Subjective Task Strategy Types	271
Semidegenerate (A:A::B:B) Geometric Analogy Subjective Task Strategy Types	278
Semidegenerate (A:B::A:B) Geometric Analogy Subjective Task Strategy Types	285
Nondegenerate Geometric Analogy Subjective Task Strategy Types	292
Degenerate Forced Choice Geometric Analogy Items	309
Semidegenerate (A:A::B:B) Forced Choice Geometric Analogy Items	312
Semidegenerate (A:B::A:B) Forced Choice Geometric Analogy Items	315
Nondegenerate Forced Choice Geometric Analogy Items	317

"Componential Analysis" Revisited	321
1 Element - 1 Transformation Nondegenerate Forced Choice Geometric Analogy Process Models	323
3 Element - 3 Transformation Nondegenerate Forces Choice Geometric Analogy Process Models	325
Clusters and Stimulus Properties	329
The Degenerate Forced Choice Geometric Analogy Cluster Solution	330
The Semidegenerate (A:A::B:B) Forced Choice Geometric Analogy Cluster Solution	333
The Semidegenerate (A:B::A:B) Forced Choice Geometric Analogy Cluster Solution	336
The Nondegenerate Forced Choice Geometric Analogy Cluster Solution	339
Clusters and Individual Differences	343
Degenerate Analogy Clusters and Individual Differences	347
Semidegenerate (A:A::B:B) Analogy Clusters and Individual Differences	348
Semidegenerate (A:B::A:B) Analogy Clusters and Individual Differences	350
Nondegenerate Analogy Clusters and Individual Differences	351
Componential Analysis and Mental Workload	353
Prospects for Componential Analysis and Mental Workload Research	366
Concluding Comments	371
CHAPTER 7	
REFERENCES	373

3.1	ASPECTS OF THEORIES AND MODELS IN COMPONENTIAL ANALYSIS	89
3.2	SOURCES OF INDIVIDUAL DIFFERENCES IN COMPONENTIAL ANALYSIS	91
3.3	COMPONENTIAL TASK ANALYSIS	93
3.4	OVERVIEW OF CONVERGING OPERATIONS IN ANALOGICAL REASONING EXTENSIVE TASK ANALYSIS : EXTENSIVE CONSTRUCT VALIDATION	110
3.5	ADVANTAGES OF INTERVAL, COMPONENT AND REFERENCES ABILITY SCORES	111
3.6	COMPONENTIAL THEORY OF ANALOGICAL REASONING	119
3.7	ATTRIBUTE - VALUE LIST FOR EXAMPLE WASHINGTON : 1 :: LINCOLN : (a.10;b.5)	123
5.1a	CONDITIONS, MEANS AND STANDARD DEVIATIONS FOR THE FACTORIAL RESPONSE TIME DATA SET RELEVANT TO DEGENERATE FORCED CHOICE GEOMETRIC ANALOGY ITEMS	180
5.1b	CONDITIONS, MEANS AND STANDARD DEVIATIONS FOR THE FACTORIAL RESPONSE TIME DATA SET RELEVANT TO SEMIDEGENERATE (A:A::B:B) FORCED CHOICE GEOMETRIC ANALOGY ITEMS	181
5.1c	CONDITIONS, MEANS AND STANDARD DEVIATIONS FOR THE FACTORIAL RESPONSE TIME DATA SET RELEVANT TO SEMIDEGENERATE (A:B::A:B) FORCED CHOICE GEOMETRIC ANALOGY ITEMS	182
5.1d	CONDITIONS, MEANS AND STANDARD DEVIATIONS FOR THE FACTORIAL RESPONSE TIME DATA SET RELEVANT TO NONDEGENERATE FORCED CHOICE GEOMETRIC ANALOGY ITEMS	183
5.2a	ANOVA SUMMARY FOR DEGENERATE GEOMETRIC ANALOGY RESPONSE TIME DATA SET	186
5.2b	ANOVA SUMMARY FOR SEMIDEGENERATE (A:A::B:B) GEOMETRIC ANALOGY RESPONSE TIME DATA SET	187
5.2c	ANOVA SUMMARY FOR SEMIDEGENERATE (A:B::A:B) GEOMETRIC ANALOGY RESPONSE TIME DATA SET	188
5.2d	ANOVA SUMMARY FOR NONDEGENERATE GEOMETRIC ANALOGY RESPONSE TIME DATA SET	189
5.3a	ANOVA CONTRASTS FOR THE DEGENERATE FACTORIAL RESPONSE TIME DATA SET	190
5.3b	ANOVA CONTRASTS FOR THE SEMIDEGENERATE (A:A::B:B) FACTORIAL RESPONSE TIME DATA SET	191
5.3c	ANOVA CONTRASTS FOR THE SEMIDEGENERATE (A:B::A:B) FACTORIAL RESPONSE TIME DATA SET	192
5.3d	ANOVA CONTRASTS FOR THE NONDEGENERATE FACTORIAL RESPONSE TIME DATA SET	193
5.4	SUBJECTIVE ITEM COMPLEXITY CODING	195

5.5a	TASK STRATEGY AND ITEM COMPLEXITY FOR 1 ELEMENT DEGENERATE GEOMETRIC ANALOGY ITEMS	205
5.5b	TASK STRATEGY AND ITEM COMPLEXITY FOR 3 ELEMENT DEGENERATE GEOMETRIC ANALOGY ITEMS	206
5.5c	TASK STRATEGY AND ITEM COMPLEXITY FOR 1 ELEMENT - 1 TRANSFORMATION SEMIDEGENERATE (A:A::B:B) GEOMETRIC ANALOGY ITEMS	209
5.5d	TASK STRATEGY AND ITEM COMPLEXITY FOR 3 ELEMENT - 3 TRANSFORMATION SEMIDEGENERATE (A:A::B:B) GEOMETRIC ANALOGY ITEMS	211
5.5e	TASK STRATEGY AND ITEM COMPLEXITY FOR 1 ELEMENT - 1 TRANSFORMATION SEMIDEGENERATE (A:B::A:B) GEOMETRIC ANALOGY ITEMS	212
5.5f	TASK STRATEGY AND ITEM COMPLEXITY FOR 3 ELEMENT - 3 TRANSFORMATION SEMIDEGENERATE (A:B::A:B) GEOMETRIC ANALOGY ITEMS	214
5.5g	TASK STRATEGY AND ITEM COMPLEXITY FOR 1 ELEMENT - 1 TRANSFORMATION NONDEGENERATE GEOMETRIC ANALOGY ITEMS	215
5.5h	TASK STRATEGY AND ITEM COMPLEXITY FOR 3 ELEMENT - 3 TRANSFORMATION NONDEGENERATE GEOMETRIC ANALOGY ITEMS	217
5.6a	SUMMARY OF CRITICAL CLUSTER SOLUTIONS WITH DELETE= OPTION FOR DEGENERATE GEOMETRIC	224
5.6b	CLUSTER MEANS AND STANDARD DEVIATIONS FOR EACH DEGENERATE FACTORIAL CONDITION	227
5.6c	ANOVA PROFILES BY CLUSTERS FOR DEGENERATE GEOMETRIC ANALOGY RESPONSE TIME DATA SET	230
5.6d	ANOVA CONTRASTS BY CLUSTERS FOR DEGENERATE GEOMETRIC ANALOGY RESPONSE TIME DATA SET	231
5.7a	SUMMARY OF CRITICAL CLUSTER SOLUTIONS WITH DELETE= OPTION FOR SEMIDEGENERATE (A:A::B:B) GEOMETRIC ANALOGY ITEMS	233
5.7b	CLUSTERS MEANS AND STANDARD DEVIATIONS FOR EACH SEMIDEGENERATE (A:A::B:B) FACTORIAL CONDITION	234
5.7c	ANOVA PROFILES BY CLUSTERS FOR SEMIDEGENERATE (A:A::B:B) GEOMETRIC ANALOGY RESPONSE TIME DATA SET	238
5.7d	ANOVA CONTRASTS BY CLUSTERS FOR SEMIDEGENERATE (A:A::B:B) GEOMETRIC ANALOGY RESPONSE TIME DATA SET	239
5.8a	SUMMARY OF CRITICAL CLUSTER SOLUTIONS WITH DELETE= OPTION FOR SEMIDEGENERATE (A:B::A:B) GEOMETRIC ANALOGY ITEMS	240
5.8b	CLUSTER MEANS AND STANDARD DEVIATIONS FOR EACH SEMIDEGENERATE (A:B::A:B) FACTORIAL CONDITION	242
5.8c	ANOVA PROFILES BY CLUSTERS FOR SEMIDEGENERATE (A:B::A:B) GEOMETRIC ANALOGY RESPONSE TIME DATA SET	245
5.8d	ANOVA CONTRASTS BY CLUSTERS FOR SEMIDEGENERATE (A:B::A:B) GEOMETRIC ANALOGY RESPONSE TIME DATA SET	246

5.9a	SUMMARY OF CRITICAL CLUSTER SOLUTIONS WITH DELETE= OPTION FOR NONDEGENERATE GEOMETRIC ANALOGY ITEMS	247
5.9b	CLUSTER MEANS AND STANDARD DEVIATIONS FOR EACH NONDEGENERATE FACTORIAL CONDITION	248
5.9c	ANOVA PROFILES BY CLUSTERS FOR NONDEGENERATE GEOMETRIC ANALOGY RESPONSE TIME DATA SET	253
5.9d	ANOVA CONTRASTS BY CLUSTERS FOR NONDEGENERATE GEOMETRIC ANALOGY RESPONSE TIME DATA SET	254

FIGURES		PAGE NO
3.1	DESIGN OF A STRUCTURAL REGRESSION MODEL	105
3.2	COMPONENTIAL THREE-LEVEL INTENSIVE TASK ANALYSIS	108
3.3	SCHEMATIC REPRESENTATION OF ANALOGICAL REASONING THEORY	120
3.4	SCHEMATIC FLOW CHART FOR MODEL I	124
3.5	SCHEMATIC FLOW CHART FOR MODEL II	127
3.6	SCHEMATIC FLOW CHART FOR MODEL III	128
3.7	SCHEMATIC FLOW CHART FOR MODEL IV	131
3.8	SCHEMATIC REPRESENTATION OF TWO TYPES OF FORCED CHOICE MODELS	133
4.1	FORCED CHOICE GEOMETRIC ANALOGIES SIMILAR TO THOSE INCLUDED IN THE EXPERIMENTAL TRIALS	165
6.1a	SCHEMATIC FLOW CHART OF THE FIRST DOMINANT SUBJECTIVE TASK STRATEGY TYPE APPLICABLE TO DEGENERATE ITEM SOLUTION	273
6.1b	SCHEMATIC FLOW CHART OF THE SECOND DOMINANT SUBJECTIVE TASK STRATEGY TYPE APPLICABLE TO DEGENERATE ITEM SOLUTION	274
6.2a	SCHEMATIC FLOW CHART OF THE FIRST DOMINANT SUBJECTIVE TASK STRATEGY TYPE APPLICABLE TO SEMIDEGENERATE (A:A::B:B) ITEM SOLUTION	279
6.2b	SCHEMATIC FLOW CHART OF THE SECOND DOMINANT SUBJECTIVE TASK STRATEGY TYPE APPLICABLE TO SEMIDEGENERATE (A:A::B:B) ITEM SOLUTION	280
6.3a	SCHEMATIC FLOW CHART OF THE FIRST DOMINANT SUBJECTIVE TASK STRATEGY TYPE APPLICABLE TO SEMIDEGENERATE (A:B::A:B) ITEM SOLUTION	286
6.3b	SCHEMATIC FLOW CHART OF THE SECOND DOMINANT SUBJECTIVE TASK STRATEGY TYPE APPLICABLE TO SEMIDEGENERATE (A:B::A:B) ITEM SOLUTION	287
6.4a	SCHEMATIC FLOW CHART OF THE FIRST DOMINANT SUBJECTIVE TASK STRATEGY TYPE APPLICABLE TO 1 ELEMENT - 1 TRANSFORMATION NONDEGENERATE ITEM SOLUTION	296
6.4b	SCHEMATIC FLOW CHART OF THE SECOND DOMINANT SUBJECTIVE TASK STRATEGY TYPE APPLICABLE TO 1 ELEMENT - 1 TRANSFORMATION NONDEGENERATE ITEM SOLUTION	297

6.4c	SCHEMATIC FLOW CHART OF THE THIRD DOMINANT SUBJECTIVE TASK STRATEGY TYPE APPLICABLE TO 1 ELEMENT - 1 TRANSFORMATION NONDEGENERATE ITEM SOLUTION	298
6.4d	SCHEMATIC FLOW CHART OF THE FOURTH DOMINANT SUBJECTIVE TASK STRATEGY TYPE APPLICABLE TO 1 ELEMENT - 1 TRANSFORMATION NONDEGENERATE ITEM SOLUTION	299
6.5a	SCHEMATIC FLOW CHART OF THE FIRST DOMINANT SUBJECTIVE TASK STRATEGY TYPE APPLICABLE TO 3 ELEMENT - 3 TRANSFORMATION NONDEGENERATE ITEM SOLUTION	303
6.5b	SCHEMATIC FLOW CHART OF THE SECOND DOMINANT SUBJECTIVE TASK STRATEGY TYPE APPLICABLE TO 3 ELEMENT - 3 TRANSFORMATION NONDEGENERATE ITEM SOLUTION	304
6.5c	SCHEMATIC FLOW CHART OF THE THIRD DOMINANT SUBJECTIVE TASK STRATEGY TYPE APPLICABLE TO 3 ELEMENT - 3 TRANSFORMATION NONDEGENERATE ITEM SOLUTION	305
6.5d	SCHEMATIC FLOW CHART OF THE FOURTH DOMINANT SUBJECTIVE TASK STRATEGY TYPE APPLICABLE TO 3 ELEMENT - 3 TRANSFORMATION NONDEGENERATE ITEM SOLUTION	306

PART 1

CONCEPTUAL AND THEORETICAL BACKGROUND

CHAPTER 1

INTRODUCTION

A concept crucial to the evaluation of modern man-machine systems is that of the resulting implications of loading and overloading the cognitive capability of the human component, namely, the measurement of mental workload.

In the past, mental workload research has been reliant upon implicit models of cognitive task demands, however, it is a contention of this thesis that for assessment of mental workload experienced by individuals performing complex cognitive tasks, success is dependent upon an accurate and robust model of task demands. Models of this nature are to be found in the inductive reasoning paradigm, particularly in the domain of analogical reasoning (e.g., Evans, 1968; Sternberg, 1977; Mulholland, Pellegrino and Glaser, 1980). Inductive reasoning ability has been prominent in theories of human intelligence, playing a central role in its assessment. A variety of tasks have been used to assess this ability (e.g., classification problems, series completion problems, analogy problems and matrix problems), of which Pellegrino and Glaser (1980) regard the analogy problems as the most pervasive.

Until recently, questions concerning the nature of individual differences in cognitive capacity have been addressed primarily by the psychometric branch of psychology. This field has been successful in developing tests that measure significant aspects of human behaviour that are related to the processing of verbal and nonverbal information and that are useful in discriminating among individuals who are more or less likely to succeed in

academic or related endeavours. What has been lacking in the field is a clear understanding of the nature of the processes underlying performance on the tests, how individuals differ in terms of these processes, and the task characteristics that influence these performance differences.

Despite the long history of inclusion of analogy problems on tests of aptitude and intelligence, a combined theoretical and empirical analysis of the component cognitive processes involved in analogical problem solving has only recently become an active research area (e.g., Pellegrino and Ingram, 1977; Sternberg, 1977, 1978; Sternberg and Rifkin, 1979; Pellegrino and Glaser, 1979, 1982; Sternberg and Nigro, 1980; Mulholland, Pellegrino and Glaser, 1980). Sternberg's (1977) attempt to specify the performance components in this problem type, constitutes the basis for the empirical and theoretical issues explored in this thesis.

Based upon his pioneering investigation of the information processing components involved in analogy task solution Sternberg (1977) developed a componential theory of analogical reasoning in which he identified a set of information processing components which we believed were sufficient for describing the flow of information processing from the beginning to the end of analogy task solution, and proposed several models, each representing different assumptions concerning the order and mode of component execution. Sternberg's research falls into the category of cognitive research referred to as componential analysis. Although Sternberg's theory and subsequent research provides valuable insights into the nature of mental abilities, it has until now been excluded from mental workload research.

The present study shall bring together the previously unrelated research paradigms of mental workload and componential analysis in an attempt to investigate the extent to which Sternberg's (1977) theory of analogical reasoning, in particular the linear componential models he proposed, provides information regarding the nature of task structure, information which could serve to enhance the present understanding of the concept of mental workload and its measurement.

This research shall attempt to investigate the extent to which findings from an apparently stable cognitive phenomenon, forced choice geometric analogy task solution, performed under conditions of varying complexity, can offer insights into the strategic influences of mental workload variation on cognitive task performance.

By facilitating the possible identification of inappropriate work routines at the cognitive level of task performance, the present research may contribute to the development of a more sophisticated notion of task structure, which together with an enhanced understanding of the concept of mental workload, could stand to benefit future research into the development of strategies for reducing the mental workload experienced by the human component in modern man-machine systems, and have wide ranging implications for the design of jobs involving the performance of complex cognitive tasks.

The main body of this thesis falls into two distinct parts. Chapters 2 and 3 set the theoretical and conceptual framework for the research, whilst Chapters 4, 5, and 6 are concerned with the execution, analysis and interpretation of the research itself.

CHAPTER 2

THE NATURE OF COGNITIVE WORK

In the past, the relatively limited capacity of technical devices allowed for the accurate description of the required human machine inputs, their interactions and subsequent system outputs of man-machine systems. Furthermore, the overt nature of the inputs, interactions and consequences facilitated system performance evaluation for the purpose of enhancing system efficiency and protecting user wellbeing.

Technological advancement, however, has changed the nature of interactions within the man-machine system, with modern equipment taking over many of the complex and integrated tasks previously assigned to the human systems component. Consequently, the individual's role in automated systems has changed from that of a generator and transmitter of physical energy to that of a decision maker, monitor and supervisory controller (Leplat, 1978; Linton et al., 1979; Mulder and Mulder, 1983).

The fact that the individual's need for physical exertion has been reduced, should however, not be taken as the implication that his/her workload has been reduced simultaneously. Indeed, workload has become more a load on the mental capacities of the individual than on his/her physical capabilities. Moreover, Linton et al. (1979) argue that increased system capabilities may have disproportionately reduced the time available to perform these mental operations, thereby adding new and more complex cognitive tasks.

The view presently held by cognitive ergonomists, is that the introduction and rapid expansion of automation within the workplace has served to transform, rather than reduce the involvement of the human in systems operations (Linton et al., 1979; Moray, 1982). Such an orientation has led to the realisation that traditional descriptive and evaluative techniques are no longer adequate for present day sophisticated man-machine systems (Tsang and Johnson, 1987). The need has therefore arisen for system design to describe human performance in terms of cognitive function, and to evaluate it accordingly. A concept crucial to the evaluation of modern man-machine systems is that of the resulting implications of loading and overloading the cognitive capabilities of the human component, namely, the measurement of mental workload.

The Concept of Mental Workload

Albeit that the understanding and assessment of mental workload is considered a topic of critical importance in the field of cognitive ergonomics (Moray, 1982; Casali and Wierwille, 1983; Wickens et al., 1986; Barber, 1988), and given the vast amount of work that has been conducted in this area, there is as yet no universally accepted definition of mental workload and consequently no agreement on how to measure it. Moreover, much of the research conducted in the domain of cognitive ergonomics, e.g., Weiner (1982); Casali and Wierwille (1983); Wickens et al., (1986), while undoubtedly based upon valid possibilities, remains essentially predictive in nature, thus failing to contribute significantly to a theoretical understanding of workload.

Characteristic to all empirical investigation, the rationale for researching mental workload is inextricably linked to the definition of the issues to be investigated. In recent times, one of the more popular approaches to

workload analysis was to consider mental workload as analogous to physical workload. Theorists adopting this approach, e.g., Welford (1978); Weiner (1982); Mulder and Mulder (1983), have attempted to define mental workload in terms which apply to its physical equivalent:- force, work, time stress, strain etc. It is, however, important to take cognisance of the fact that the majority of such theories were proposed several years ago, before research conducted in the area of mental workload had reached present levels of sophistication. As a result, caution should be adopted in considering such theories as definitive literature upon which to base current research. In fact, a major criticism levelled against workload research, is the continued reliance by other workload theorists on the physical approach as providing the basis for their investigation. Of greater concern, however, is the nature of the conclusions drawn from such experimentation. Essentially, research based upon a physical concept of workload constantly measures physiological activity in a stimulus-response type paradigm. Since it is possible to measure physiological change in both an accurate and objective fashion, it is very often taken for granted that the results yielded by such experimentation are both valid and reliable. Such conclusions should, however, be regarded as highly questionable, since despite the attractiveness of the physical analogy, due to its familiarity and ease of quantification it provides an oversimplified view of the complex cognitive processes of mental workload and their effects on human performance on problem solving tasks.

Most significant here is the fact that physical approaches are limited to a consideration of only those system inputs and responses which are overt and measurable. It would therefore appear that physical theories lack the scope necessary to consider internal psychological processes. As a result, the physical stimulus-response type approach will

either fail to detect important internal processes and consequences altogether, or will be incapable of explaining certain overt responses due to their internal psychological origins. The end result is a theoretical base with negligible predictive power, the value of which, for the purpose of assessing mental workload on cognitive tasks, is highly questionable.

In recent years, a realisation of the limitations of the physical approach led to the development of an alternative paradigm concerning mental workload. Within this paradigm mental workload is regarded in information, processing terms e.g., Bainbridge (1978), Hacker et al., (1978), Moray (1979, 1979a).

The information processing paradigm was developed as an alternative to the stimulus-response concepts of behaviourism (Sternberg, 1977). The major argument against stimulus-response concepts was that they were a gross oversimplification:-

"... correlations between stimulation and response must be mediated by an organised representation of the environment, a system of concepts and relations within which the organism is located. A human being builds up an internal representation, a model of the universe, a schema, a cognitive map ..." (Miller et al., 1960, pp.6-7).

The common new emphasis was:-

"... the exploration of complex processes and the acceptance of the need to be explicit about internal symbolic mechanisms" (Newell and Simon, 1972, p.4).

Human information processing theory assumes that behaviour is a consequence of the internal flow of information within

the organism, which is believed to occur at a number of stages (McCormick and Sanders, 1982). Certain cognitive handling facilities are present at each stage, their purpose being to process information into some understandable and usable form (Norman and Bobrow, 1975). Such facilities include:- some form of information reception mechanism, known as the Perception Mechanism (Card et al. 1983), which is responsible for the differentiated recognition of stimuli (Navon and Gopher, 1979) and their transformation into a state which can be interpreted by later processing mechanisms. McCormick and Sanders (1982) emphasise that the source of stimuli can arise either externally to the organism, or internally from some other processing mechanism and be represented for perception through feedback loops and integrated linkages between stages. After perception, stimuli enter what Card et al., (1983) refer to as 'Processing Stages' which are thought to include:- some form of information storage (long and short term memory) (Smith and Langolf, 1981); operations designed to enter information into memory by coding (Shiffrin, 1975); location and retrieval mechanisms, decision making mechanisms based on information and task requirements (Kantowitz and Roediger, 1980); and finally, response operations selected from a range of response resources (Navon and Gopher, 1979).

Information processing theorists describe mental workload in terms of mental task demands which impose a load on cognitive information processing mechanisms (Bainbridge, 1978). The efficiency with which each mental processing mechanism deals with mental task demands is a function of two factors. Firstly, the task characteristics (Navon and Gopher, 1979), which include interalia - stimulus quality; distribution of demand across processing resources (Wickens, 1984); time available for processing of each discrete stimulus (Hacker et al., 1978), and secondly, the expertise or sophistication of the processing mechanisms

themselves (Bainbridge, 1978; Hacker et al., 1978; Leplat, 1978; Teiger, 1978).

It would appear realistic, at this stage of the present discussion, to contend that the lack of consensus upon the definition and measurement of mental workload (Williges and Wierwille, 1979; Moray, 1982) may be attributable to the attempts by theorists to define workload as a unitary concept, having simple characteristics and effects.

A unitary approach to mental workload and its measurement leads to methodological problems since the level of definition is often inappropriate for the level of research conducted. Thus, as a result of the uncertainty surrounding the concept, certain theorists have adhered to a number of untested and often invalid assumptions from which they propose methods of defining and evaluating mental workload (Teiger, 1978; Welford, 1978).

Since it is a theoretical construct, mental workload might best be defined operationally, as this would facilitate dealing with the concept at various levels of detail so that the details of the experimental design are adequately justified, thereby yielding logical and interrelated results and conclusions. For present purposes then, an information processing approach will be used as a basis for the construction of an operational definition of mental workload and its consequences for task performance.

It was highlighted earlier in this chapter that the developing sophistication of man-machine systems has altered the nature of task demands encountered by the human component, from mainly psychomotor demands to demands on his/her cognitive functions. Consequently, the workload experienced by operators of sophisticated automated systems can be more accurately described as 'Mental Workload' and redefined as 'the extent to which an individual is able to

fulfill the mental task demands imposed by the task in which he/she is engaged.'

Phase two of an operational definition involves the consideration of certain aspects in more detail. In human information processing terms, performance and workload appear to depend upon the interaction of four factors:- the mental task demands (Bainbridge, 1978); the information processing capacities of the performer (Hacker et al., 1978); the strategies used to relate demands to cognitive capacities; and when a range of cognitive strategies is available, skill in choosing the most efficient (Norman and Bobrow, 1975). Rather than considering the factors individually, for present discussion purposes, they will be dichotomised into:- those factors which are task determined, i.e., the first of the four factors listed above; and those factors which are operator determined, i.e., the latter three factors listed above. Rolfe et al., (1974) and Leplat, (1978) refer to these two factors/groups of factors as 'Task Orientated' and 'Operator Orientated' determinants of mental workload.

Determinants of Mental Workload

The advent of complex, integrated man-machine systems necessitated the creation of a procedure that could specify human performance characteristics in a system. The specification of those requirements should facilitate improvements in system design and overall system effectiveness. "For practical purposes, the acquisition of that type of information constitutes a task analysis" (Companion and Corso, 1982, p.459). Drury (1983) describes task analysis as providing a formal comparison between the demands which the task imposes on an individual and the capabilities he/she possesses to deal with these demands.

Task Orientated Factors

Essentially, any given task imposes an actual demand upon the cognitive processing mechanisms required for its interpretation and resolution. The nature of the task itself will determine the extent of the load placed upon each mechanism involved in task performance.

Research conducted by Wickens et al. (1986) to assess the actual workload demands imposed by complex cognitive tasks served to highlight two important characteristics of workload. Firstly, that human information processing resources are limited. Irrespective of an individual's level of skill/experience, his/her ability to process information will be subject to an upper limit (Hicks and Wierwille, 1979). The most significant implication being that overload is possible. It has been noted (Norman and Bobrow, 1975) that as workload levels near the upper limits of an individual's capability, a gradual deterioration occurs in his/her information processing performance.

The gradual nature of performance deterioration may be attributed to the second characteristic highlighted by Wickens et al (1986), namely, interference between the demands of individual tasks and the processing stages themselves. Kantowitz and Roediger (1980) theorised that human information processing stages can be arranged in different patterns, with the simplest pattern, 'serial processing' occurring when there is a chain of stages with the output of one stage being the input of the next. Due to the complex nature of certain cognitive tasks, e.g., process control, serial processing is highly unlikely. More likely, information will be subject to parallel processing - when several stages simultaneously have access to the same inputs and outputs, or hybrid processing - where serial and parallel processing occur in the same activity. Increasing task demands on human information

processing mechanisms results in a corresponding increase in the complexity of processing patterns in order to maintain simultaneous performance across a number of sub-tasks. The limited capacity of the human information processing system prevents an individual from performing more and more operations in order to satisfy task demands. Instead, 'time sharing' - the performance of two or more activities in rapid alternation (McCormick and Sanders, 1982) will occur. A consequence of time sharing is that interference occurs between processing activities (Wickens, 1987), resulting in deterioration in the performance of one or all of the tasks between which resources are shared. Dual task interference is greater when the tasks share the same inputs or response modality or make equivalent demands on internal processing mechanisms (McLeod, 1976).

These characteristics, namely, that the demands imposed by two or more related tasks may compete for the resources of a limited capacity system, have important implications for the techniques of mental workload measurement. These implications will be discussed in finer detail in the section Spare Mental Capacity (Secondary Task) Measures later in this chapter.

Numerous ways exist in which to describe and analyse work and human involvement with it (Companion and Corso, 1982; Peterson and Bowas, 1982). What is frequently overlooked is that these different forms of work and task analysis have evolved to serve different analytic objectives. Commenting on task analysis, Drury (1983) maintains:-

"This analysis is identical whether it is to be used to design the equipment, the procedure, the training scheme or the selection requirements. How it is implemented will differ greatly in each case ..."
(p.19).

Selecting the 'correct' form of analysis is therefore dependent upon the objectives of the analyst. Having discussed the use of task analysis in the context of mental workload research, attention will now be directed at its role within the domain of ergonomics.

The success of any attempt to design interventions into man-machine systems will be largely dependent upon the extent of the designer's understanding of the fundamental functions to be performed by the system under investigation (Meister, 1971; 1976). It is important to emphasise that at the initial stages of consideration, the ergonomist is primarily concerned with the identification of function rather than with organisational roles. Within this context, Drury (1983) defines functions as the major logical units of behaviour (in terms of either human or machine activity) which are needed to accomplish the objectives of the system. Closely associated with this concept of abstract function is the notion of the 'task' as the molecular unit of human-machine system activity.

Tasks are similarly defined by Drury (1983) as being the smallest units of behaviour which need to be differentiated in order to solve the problem at hand.

"The principal aim of the ergonomist in analysing work-supporting structures must thus be to provide a detailed description of functions and tasks in terms of their interaction with the overall system's purpose and its implications for the human-machine workspace-environment configurations which must support task performance" (Fisher, 1986, p.453).

As task demands upon the human component in modern man-machine systems become more cognitive, involving sophisticated human information processing, traditional techniques of task description and analysis, which are

largely observational with the emphasis being on standardisation, which focus on psychomotor abilities and physical proficiencies, and which have their foundations in organisational psychology and psychometrics, for example, Fine's Functional Job Analysis (Fine, 1963) and McCormick's Position Analysis Questionnaire (McCormick, Jeanneret, and Mecham, 1969), become less useful and increasingly difficult to perform. Rather, an approach focusing on human cognitive abilities, emphasising decision-making and information handling aspects of work is required.

Harker and Eason (1980) proposed a highly versatile approach to task analysis based upon the application of socio-technical systems theory to the structure of the task, with the emphasis on systems ergonomics and the cognitive psychology of task demands. Commenting on forms of task analysis such as those proposed by Fine (1963) and McCormick, Jeanneret, and Mecham (1969), as well as more current forms, Harker and Eason (1980) emphasise that task conditions tend to be confused with task performance and that, moreover, by concentrating on the conditions and performance factors for the usual fulfilment of task demands the analyst will fail to conceive the variety of task conditions with which the task performing structure must cope. This will have significant implications for the range of skills needed by the human operator as well as for the flexibility of the fundamental task structure needed in order to cope with varying demands.

Harker and Eason (1980) suggest that by concentrating on the transformations that the task structure must undertake, a more comprehensive understanding of these varying demand characteristics will be obtained. This will facilitate the reconceptualisation of the task in line with the theory of socio-technical systems (traditionally applied to organisations), where the task is regarded as a self-organising, self-sufficient system displaying

properties and components similar to those of any other organisation unit.

The analysis itself comprises six phases of investigation, focusing on these components and their relationships. For present purposes a detailed discussion of these phases would seem appropriate.

The first phase requires the establishment of the boundaries of the task system. This enables the analyst to prescribe the extent of his/her immediate interest, to identify the level of organisational action at which the analysis will take place, and to "preserve (as input and outputs, etc.), the critical relationships with other organisational levels without involving the (often synthetic) partitioning of task components into assumed component levels" (Fisher, 1986, p.454).

The second phase involves determining the goals of the task system. The extent to which these can be specified will indicate a great deal about the flexibility of the system's structural requirements. However, it is possible to identify for any task structure a range of goals, many being sought simultaneously. Where there is a multiplicity of goals there must be a trade off between them, and an analysis of goals requires the analyst to establish a priority of goals. Furthermore, it is necessary to differentiate between those goals set as the product of work production as against those which are concerned with ensuring the continued viability of the task structure itself (e.g., human safety, health, and care and maintenance of equipment, etc.).

Phase three comprises identification of the type and degree of 'openness' to the task inputs. Of importance here are the range and predictability of the inputs themselves as well as the degree of control over them that

the task system is able to exert. The more 'open' the condition, the greater the demand for the task structure to be flexible and adaptive. The essential requirement in this phase of analysis is to establish the task parameters which are both variable and critical to task performance, and to focus attention upon the performance strategies necessary within the task performing structure.

The fourth phase in the analysis concerns the detail of the transformation processes undertaken by the task system. The main objective being to identify the invariant transformational characteristics, time constraints, sequential, reciprocal and parallel relationships, etc., and their relationships with goal and input variability.

The fifth phase involves extracting from the description of the system which has emerged from the first four phases, a minimum system structure which would achieve the transformations necessary, given goal and input conditions. This enables the analyst to remove any preconditions of system structure which may be induced by overfamiliarity with existing or traditional systems and use the minimum structure as the basis for a variety of task performing configurations.

Having reached an understanding of the various factors influencing the task there is still the requirement of specifying the characteristics of those parameters impinging upon task performance which may be externally determined. The sixth and final phase in the analysis is concerned with the process of making such specifications. Based upon the minimum systems structure emerging from phase five, together with his/her understanding thereof, the analyst is required to consider a further six aspects in order to begin constructing an alternative optimal task supporting entity, viz, what skills and capabilities need to be contained within the task system, what should be the

role of the human; what the implications are for interface design; what changes if any should be made to the task structure in order to best satisfy the minimum structure revealed at the fifth phase; and lastly what provision should be made for selection and training (Harker and Eason, 1980).

Despite the theoretical appearance of the analytical procedures proposed by Harker and Eason (1980), when considered from the perspective of socio-technical systems theory, it becomes evident that they are the re-assertion of certain fundamental concepts drawn from organisational theory, rearranged to explore task structuring. In practice, the nature of these concepts enables the analyst to use the same tool across a wide range of task structure. Fisher (1986) considers the major advantage of this approach to task analysis as being "the ability to incorporate full expression of task components and their relationship to the broader organisational structures which support them, without dependence upon any detailed analytic involvement with the entire organisational structure" (p.456).

The form of task analysis proposed by Harker and Eason (1980) would appear to surpass most current forms of analysis by allowing the analyst to take into consideration the broad range of conditions with which the task performing structure must cope, thereby enhancing the flexibility of the fundamental task structure to cope with varying demands. Furthermore, the proposed procedures provide information concerning the nature of the task in terms of the necessary human, technological and environmental sub-systems at a level of description facilitating detailed consideration of the interacting elements but without necessitating a predisposition to a solution strategy (Fisher, 1986).

Given the issues highlighted in this section concerning the use of task analysis in the context of both mental workload research and ergonomics, it is contended that for the assessment of mental workload experienced by individuals performing complex cognitive tasks, analytical success is dependent upon an accurate and robust model of task demands. Models of this nature are to be found in the inductive reasoning paradigm, particularly in the domain of analogical reasoning (e.g., Sternberg, 1977; Mulholland, Pellegrino and Glaser, 1980).

Based upon his pioneering investigation of the information processing components involved in analogy task solution Sternberg (1977) developed a componential theory of analogical reasoning in which he identified a set of information processing components which he believed were sufficient for describing the flow of information processing from the beginning to the end of analogy task solution, and proposed several models, each representing different assumptions concerning the order and mode of component execution (see the sections Sternberg's Componential Theory of Analogical Reasoning and Componential Models of Analogical Reasoning in Chapter 3 for detailed discussions of the theory and models respectively). Sternberg's research falls into the category of cognitive research referred to as componential analysis. Although Sternberg's theory and subsequent research provide valuable insights into the nature of mental abilities, it has until now been excluded from mental workload research. However, based upon the findings of their componential investigation into the relationship between item content and processing operations in geometric analogy task solution, Mulholland, Pellegrino and Glaser (1980) suggested that working memory limitations associated with high levels of item complexity could have a moderating influence on overall solution strategy. No research has since been undertaken to investigate this assumption, as it would apply to geometric analogy task performance.

It is a major contention of this thesis that Sternberg's theory, and componential models in particular, could well provide valuable information regarding the nature of cognitive task structure, and the influence thereof upon the mechanisms of strategy choice. This information could serve to enhance the present understanding of mental workload and its strategic influences on cognitive task performance.

Operator Orientated Factors

Operator-orientated factors include the information processing capacity of the human systems component, and the proficiency with which he/she recognises the use for, and make use of these capacities in task performance.

A common source of difficulty in relating processing capacity to task demand is that the precise ways in which tasks are performed differ widely between individuals and within any one individual on different occasions. There are usually several ways in which task demands can be related to or "mapped-onto" capacities. In other words, there are usually different methods or 'strategies' for achieving adequate performance, some being more efficient than others. Skill, in terms of cognitive task performance, would therefore appear to lie in selecting the most efficient out of any range of available information processing strategies. In fact, numerous theorists, e.g., Bainbridge (1978); Hacker et al. (1978); Leplat (1978), consider processing expertise to be a function of skill and practice.

Bainbridge (1978) suggests that increasing experience may have several effects on both the number and type of decisions which must be made to choose an appropriate strategy. The experienced individual may use a particular strategy automatically, because the alternatives have been

assessed against decision criteria on previous occasions, so a decision would be redundant. The individual has learned the properties of the processing strategy and of his/her own behaviour in relation to it, thus facilitating the selection of an appropriate strategy without exploration, by fore-knowledge of results. Consequently, the experienced individual does not have to carry out the alternative strategies to discover whether they are appropriate, therefore time and effort are not wasted on inadequate methods. The processing strategy is controlled with low error and low uncertainty, allowing the individual to have better control of the amount of mental effort he/she expends.

For an inexperienced individual all processing strategies are more complex and uncertain, and therefore slower. A major portion of the individual's strategy is concerned with how to complete the task, with decisions about his/her own performance such as learning and exploration, rather than dealing directly with the task demands. The experienced individual does not have to make decisions at these learning levels, so not only is the amount of work reduced, but a higher proportion is directly task related. An evaluation of the task performance of an experienced individual would suggest that he/she has a larger information processing capacity than an inexperienced individual, but this may occur only because more of the mental work he/she performs in a given time is directly task related.

Thus, it would appear that increased experience serves to enhance an individual's level of proficiency on a specific task or group of tasks, thereby reducing the mental effort he/she expends in order to meet task demands. Consequently, the same actual demand will impose different levels of operational workload dependent on the cognitive capabilities of an individual.

An increase in the demands imposed by a cognitive task should affect strategy choice (Bainbridge, 1978). When task demands increase, it is important to establish whether an individual can use a different working method which will yield increased task performance for the same amount of mental work (Sperandio, 1972).

In well practised behaviour, changes in efficiency have occurred at all levels of behaviour organisation, with complex interrelations between levels and lower levels being independent of 'conscious' monitoring (Bainbridge, 1978). Should the individual revert to devising new methods and an active choice of processing strategy, this could disrupt the parts of task performance which are organised autonomously. These 'lower' parts of task performance will be integrated, in that they anticipate certain cues at certain times. Failure of these cues to arrive at the anticipated times can have a rapidly destructive effect on task performance, thereby exacerbating the individual's uncertainty about his properties and adequacy of his/her own performance.

An individual might select several methods of simplifying the task should he/she be unable to perform all of the necessary sub-tasks. For example, he/she may attempt to continue as normal, find this impossible and so omit whole sections of the task at random, alternatively the individual might concentrate on high priority aspects and forget others. Given a fuller understanding of the strategy choice mechanisms, it may be possible to predict what types of breakdown reaction could occur; assuming that each individual gives main weighting to some particular type of parameter in the task performance decision.

The preceding discussion of operator orientated determinants of mental workload would appear to suggest that a breakdown in task performance under high demands

could occur as a natural function of normal decision-making mechanisms, rather than through an actual deterioration in information processing capacities.

Areas of Application For Workload Measures

To a large degree, the measurement techniques chosen for workload assessment will be determined by the intended usage of evaluation results. However, an overwhelming majority of workload researchers tend to overlook this fundamental point, adopting the attitude that once developed, an assessment technique would be valid for all purposes for which mental workload information is required.

Chiles and Alluisi (1979), recognising the somewhat optimistic nature of this assumption, have identified three major application areas for workload research. Firstly, identifying individuals with appropriate information processing expertise and attitudes constitutes a fundamental element of Personnel Selection. Secondly, system design is an area where measurement would be primarily concerned with the prediction of workload to be imposed by new systems, system components or procedures. Finally, with the case of modern high tech systems, where new technology is continually being introduced, expanding the capabilities of the man-machine system and changing the nature of man-machine interface, it is of particular importance that workload assessment techniques have the ability to distinguish between 'task orientated' and 'operator orientated' sources of workload. This will enable the researchers to obtain more information about the specific locus of unacceptable workload (i.e., design, training, emotional etc.), by using a method best suited to that area of investigation. It is appropriate at this stage of the discussion to conduct an assessment of mental workload measurement techniques. The ensuing discussion will hopefully demonstrate how an inadequate theoretical

framework affects evaluation methodologies and results in incomplete and often invalid conclusions concerning the nature and effects of mental workload.

Mental Workload Measurement Techniques

Developing a theoretical understanding of mental workload constitutes only part of any attempt towards the alleviation of unacceptable levels of mental load. Any attempt at practical intervention relies for its success upon the translation of theoretical concepts into practical methods for the measurement and evaluation of workload and its consequences on cognitive task performance. Attempts at this critical process led to the proposal of numerous measures, test instruments and analytical procedures, collectively termed 'workload estimation techniques' (Casali and Wierwille, 1983). However, the lack of consensus in theoretical understanding of mental workload (Moray, 1982) has resulted in a severely limited empirical base for assisting workload researchers in the systematic selection and subsequent application of appropriate workload estimation techniques (Casali and Wierwille, 1984).

It is of particular importance to make certain that both the selection and application of workload measurement techniques are conducted under the strictest, controlled experimental conditions, as this will ensure that the design, procedural or operator-related decisions made on the basis of theoretical assumptions and incorporated into the workstation, are in fact valid and will have the effect of enhancing overall task performance.

Essentially, the high degree of complexity associated with the validation of mental workload measurement techniques may be attributed to the transient nature of workload itself, its locus within an individual's cognitive system,

the (critical) consequences of individual overload and traditional problems such as experimental control and reliability (Casali and Wierwille, 1983).

"Mental workload itself cannot be directly observed and therefore its presence and its severity must be inferred from changes in overt behaviour and/or measurable physiological and psychophysiological functioning" (Casali and Wierwille, 1984, p.1033)."

It is significant to note that many of the measurement techniques proposed by mental workload theorists are based on the assumption of a connection between mental workload and some specific behavioural, physiological and psychophysiological change. Such changes are considered to be a consequence of load. As a result it is assumed that by measuring the extent and changes in consequences, a direct indication of the extent and changes in mental workload will be obtained.

In view of the theoretical understanding of human information processing and its outcomes, it would seem valid to assume that an individual would experience some reaction or consequence as a result of mental workload. The fact that human information processing mechanisms are activated, implies some eventual task related output (performance, decision making, etc.). However, problems are encountered when it becomes necessary to decide upon which of the behavioural, physiological or psychophysiological dimensions will change as a result of mental workload, how these changes will be manifested, and how they should be assessed. The type of measurement technique selected will determine the criteria upon which decisions of this nature are based. Before embarking upon an assessment of each type of technique, it is important to consider two requirements, highlighted by Casali and Wierwille (1983), that must be satisfied by evaluation

techniques, i.e., Method Sensitivity and Low Method Intrusion.

Method Sensitivity

Casali and Wierwille (1983) have defined method sensitivity as the ability of a measurement technique to "reliably discriminate between levels of loading" (p.625). A measurement technique should be able to isolate the effects of the task being considered, without reflecting extraneous variation from other task demands or irrelevant tasks (Casali and Wierwille, 1984) and, having achieved this be able to provide an account of the variation of these effects with changes in load.

The assessment of method sensitivity is largely dependent on the specification of the loading levels themselves. Should a researcher find it necessary to discriminate only between high and low workload levels, this could be achieved by a relatively insensitive technique. However, of greater concern would be a situation requiring fine distinctions between levels of loading. Under such circumstances the degree of method sensitivity would be a function of its ability to determine 'statistically significant differences in workload' (Hicks and Wierwille, 1979). It is important to emphasise that the determination of loading levels should not be dependent on measurement technique capability, but rather, a measurement technique should be selected on the basis of its sensitivity.

Method Intrusion

Method intrusiveness is defined as an "undesirable, artificial change in instructed task performance, solely due to the concurrent use of a workload estimation technique, related procedure, or associated apparatus" (Casali and Wierwille, 1983, p.625).

Ideally, workload measurement techniques should in no way intrude upon task performance, as this will have the undesirable effects of altering performance, thereby resulting in unrepresentative task behaviour and inaccurate indices of load. To assess the intrusion effects of a workload measurement technique would require that a researcher apply the technique during an instructed task sequence, and thereafter obtain measures of task set performance.

Based on the outcomes of a survey they conducted, which focused on over 400 references dealing with the definition and/or measurement of mental workload, Wierwille and Williges (1978) concluded that workload measurement techniques can be broadly classified into four main categories on the basis of their assumptions about the nature of mental workload and its consequences for systems performance. These four categories are:-

- Primary Task Measures;
- Spare Mental Capacity Measures;
- Physiological Measures;
- Subjective Opinion Scales.

Attention will now be focused upon the theoretical basis, procedural applications, sensitivity and intrusiveness of each of these categories.

Primary Task Measures

Mental workload theorists (e.g., Leplat, 1978; Hicks and Wierwille, 1979; Casali and Wierwille, 1983; 1984) refer to the task set about which workload information is required as the Primary Task.

Primary task measurement techniques assume that underlying changes in the workload imposed by the task will be

reflected by measured changes in primary task performance. According to the primary task hypothesis, workload increases will produce a corresponding deterioration in task performance. A reduction in workload levels will, however, allow the level of task performance to return to equilibrium. A common feature of primary task measures is that in no way do they attempt to explain or evaluate the cognitive components of load, they merely regard performance variation as an index of variation in load (Leplat, 1978). Task performance variation is quantified by comparison of actual primary task performance to preassigned task objectives. Typically, task performance is broken down into a number of task dimensions for which it is possible to obtain quantitative performance data, in terms of deviations from baseline performance levels.

The number of primary task dimensions assessed is largely determined by the complexity of the primary task itself, e.g., process control will involve the performance of more sub-tasks than a simple motor task. Consequently there will be more dimensions along which variation can be assessed. It is important to note that despite the complexity thereof, the breakdown of the primary task into task dimensions facilitates the selective analysis of primary task dimensions. Certain workload theorists (e.g., Hicks and Wierwille, 1979; Casali and Wierwille, 1983, 1984) have combined the measured variations in a number of dimensions into a single index of primary task variation. Wierwille and Gutman (1978) demonstrated multivariate measures to be superior to a single dimension measure for the purposes of post hoc statistical analysis. Other theorists, however, disagree. For example, Bortolussi et al. (1986) argued that a single sensitive dimension will provide a more accurate indication of load variations than will a multivariate index, which may comprise certain dimensions insensitive to variations in workload. To date, however, the choice of primary task dimensions has been largely at the discretion of the individual researcher.

Casali and Wierwille (1984) have proposed an alternative method for the selection of primary task dimensions. They suggest that task dimensions be chosen on the basis of the type of processing activities required for their performance. Casali and Wierwille (1984) based their selection of primary task dimensions on the Classification of Universal Operator Dimensions proposed by Berliner et al. (1964). The advantage of this approach is that there is an attempt to link primary task measures to a theory of workload. Consequently, the dimensions measured are likely to have more validity (in so far as the theory itself is valid) than those selected purely on the basis of overt task responses. However, as with most mental workload measurement techniques, this approach experiences the recurrent problem of having to measure the behavioural component which due to the predominantly cognitive nature of primary task dimensions, becomes increasingly complex and less apparent. To evaluate dimensions other than task parameter deviations in this way, would require more research into the structure of the primary task.

Given the fact that primary task measurement techniques have the ability to measure only overt task performance, the main disadvantage of the method is thus its overall low sensitivity (Gartner and Murphy, 1976). As a result, many of the components of the total workload may be overlooked, e.g., two different task configurations may reflect equal primary task ratings, but in reality the effort expended by the subject on one task may greatly exceed efforts on the other. Furthermore, in terms of the alleviation of high workload levels, primary task indices fail to indicate whether the source of workload is either 'task orientated' or 'operator orientated' or whether the problem is one of perceptual, mediational or response resource loading, i.e., the measure has low diagnosticity (Wickens et al., 1986).

The advantages of primary task measurement techniques are

twofold. Firstly, since performance is monitored either by a third party or, as is more common, by a recording facility in the equipment itself, such methods are potentially non-intrusive. Secondly, the subject is not required to perform any additional behaviour for measurement purposes (Hicks and Wierwille, 1979) since behaviour assessed by primary task measures is already present as part of the normal task repertoire.

Spare Mental Capacity (Secondary Task) Measures

In an attempt to create more novel techniques for workload measurement, researchers began to move away from the measurement of how much mental capacity is expended, to the consideration of how much is not. The question of reserve mental capacity led to the development of measurement techniques through the Spare Capacity theory and Secondary Task Method. The spare mental capacity techniques assume an upper limit on the human operator's processing abilities.

As discussed in the section Task Orientated Factors earlier in this chapter, the informational demands of two or more mental tasks compete for the resources of a limited capacity human information processing system, any overloading of which will cause performance decrements in one or more of the tasks (Wetherall, 1981). Alternatively, should no overloading occur, a certain amount of reserve mental capacity will remain dormant. By measuring the reserve capacity, an indication is gained of the mental load involved in task performance (Kalsbeek, 1968). Albeit that primary task measures assume performance variations to be representative of changes in task imposed workload, Kalsbeek (1981) points out that an individual may keep task performance constant by summoning further effort from a reserve store kept for the purpose of dealing with short periods of overload.

The secondary task method attempts to evaluate the extent of reserve or spare mental capacity and use this as an indication of the extent to which the human information processing system is engaged in meeting task demands. This is achieved by the introduction of a secondary loading task to be performed simultaneously with the primary task. Successful secondary task performance would suggest that the processing capacity of the individual exceeds the demands of the primary task (Knowles, 1963) and that there is adequate spare mental capacity for secondary task performance. However, as the demands of the primary task increase, reserve capacity will decrease and secondary task performance will deteriorate. It is necessary to take into consideration the implications of secondary task measures by making judgements regarding their sensitivity and intrusiveness. The first, and most central of these issues is that of Dual Task Interference.

As discussed in the section, The Concept of Mental Workload earlier in this chapter, processing occurs in different stages and at each stage there are a number of processing mechanisms or resources. Should two tasks/task sets distribute demands across these resources, these will be time-shared quite efficiently. If however, the tasks impose demands on the same resources, interference will result, causing a deterioration in the performance of one or both of the tasks. Since secondary task techniques are intended to evaluate the effects of primary task performance on the secondary task efficiency, it is important that the demands imposed by the secondary task do not draw upon equivalent resources and thus cause artificial changes in primary task performance through interference. Such intrusiveness would result in erroneous estimates of imposed workload. Secondary task design is therefore vital to the viability of Spare Capacity Techniques as measures of mental workload. A number of methods have been developed for the purpose of minimizing

secondary task interference, these include:- modality specification, task complexity, temporal structure, constancy of secondary task requirements and priority specification.

Wetherell (1981) argues that when the primary and secondary task share the same sensory or response modality, dual task interference will be high. Although inter-sensory modulation complicates the control of interference effects (Taylor and Campbell, 1976), Wetherell (1981) proposes that by selecting the modality for the secondary task that is least used in the primary task, such effects can, to some extent, be reduced.

A second factor requiring consideration during the construction of a secondary task is the level of task complexity. Secondary task complexity will determine the amount of processing capacity required regardless of modality. Highly complex secondary tasks will therefore reduce the resources available for primary task performance (Wickens, 1987), resulting in unacceptable levels of interference and thus invalidating the measures (Moray, 1982). Similarly, where the secondary task is too simplistic, subject response will quickly become automatic (Vidulich and Wickens, 1986), irrespective of the nature of the primary task, resulting in low method of sensitivity.

The third secondary task factor to be considered is the specification of its temporal structure. Temporal structure can be specified along two dimensions, i.e., forced paced or self paced. A forced paced task structure requires the subject to perform the secondary task at set intervals. This may be achieved by either the presentation of stimuli at regular intervals, or by specifying that the secondary task must comply with certain temporal requirements. Self pacing refers to conditions where the subject performs at his/her own pace. Brown (1978)

emphasises that since subjects set their own response rate, self pacing prevents between subject comparisons. Furthermore, between subject variations in response rate may well be due to natural pacing rather than an increase in experienced workload. As a result, Brown (1978) concludes that for secondary tasks to be useful workload indices, they should be force paced.

The fourth secondary task factor requiring specification is the constancy of the human information processing demand of the additional task from one discrete stimulus to the next (Brown, 1978). This criterion provides a reliable scale of measurement within different levels of difficulty, and facilitates the moment to moment equation of task performance.

The final criterion for secondary task specification concerns the relative priorities the subject is instructed to assign to each of the two tasks. In Spare Mental Capacity Techniques the subject should be clearly instructed to regard the additional task as being of secondary importance to primary task performance. Failing to heed such instructions will result in unacceptable levels of secondary task interference, thus lowering the validity of the method. Albeit that in many applications task priorities are inherently clear (Brown, 1978), instructions to the subjects remain necessary to ensure the prevention of confusion over task priorities.

The preceding discussion of those issues relevant to the use of secondary tasks as mental workload measures, has served to highlight the complexity of the Secondary Task method. It was emphasised that the detailed specification of the subsidiary task can facilitate the successful use of reserve measures as indices of mental workload.

Physiological Measures

A further attempt at objective measurement is evidenced by the frequent use of physiological indices of mental load, which to a large degree can be seen as a result of the physical approach to mental workload measurement (see the section The Concept of Mental Workload earlier in this chapter for a discussion of this approach).

Research incorporating the measurement of physiological parameters is based upon two assumptions. Firstly, that workload has the effect of imposing some form of internal 'stress' (Weiner, 1982) upon the subject; and secondly, that fluctuations in workload will cause certain involuntary changes to take place in the physiological processes of the human body (Ettema and Zielhuis, 1971). As a result, numerous workload researchers (e.g., Ettema and Zielhuis, 1971; Bauer et al., 1985; Goldstein et al., 1985; Stern, 1987; Wilson et al., 1987) have argued that measures of physiological change can be used as an index of cognitive activity and mental workload imposed by task demands. Certain researchers (e.g., Mulder and Mulder, 1983) have even attempted to identify the processing stage being loaded.

Physiological measures taken for workload research purposes include:- heart rate, sinus arrhythmia, adrenaline and catecholamine excretion, electrical brain activity, etc. However, the theoretical and methodological problems associated with the use of such indices have served to restrict their use as effective mental workload measures (Hicks and Soliday, 1972).

The first major theoretical flaw associated with the use of physiological indicators as indices of mental workload concerns the assumption of a direct relationship between cognitive arousal and physiological reaction. However, the

multi-dimensional nature of the two variables makes any attempt at a meaningful interpretation concerning the link between cognitive activity and recorded physiological data extremely difficult. For example, research conducted by Ettema and Zielhuis (1971) revealed similar physiological changes in a number of workload situations. Moreover, the likelihood exists that similar task demands may induce a range of physiological changes, thus introducing a further problem of distinguishing between such changes. Finally, the very nature of cognitive functioning, i.e., the rapidity and subtlety thereof, would suggest that physiological measures are sensitive to cognitive processing only at the most general level.

The principal of Graceful Degradation provides evidence of the gradual build up of mental workload. Primary Task measurement techniques show measured deviations in primary task performance gradually returning to baseline levels, thus indicating a similar effect in the reduction of workload levels. This cumulative, rather than instantaneous nature of workload, referred to as the Time Lag effect, is relevant to the physiological measurement of mental workload.

In relation to cognitive flux, physiological change is an extremely slow process (Strasser, 1977). Consequently there will be a considerable time lag between the onset of cognitive change and its corresponding physiological reaction. In addition, variables such as cognitive process, physiological function and characteristics of individual subjects will cause variation in this time lag.

A third problem encountered when using Physiological Measures concerns the procedure, which typically involves the scoring of physiological reactions in terms of deviations from baseline levels. Casali and Wierwille (1983) argue that the wide range of normal physiological

baselines and flux ranges have the effect of invalidating between subject comparisons. However, the value of physiological measures in differentiating between intraindividual states has not been discounted.

It has been argued that physiological measurement is readily available, unobtrusive, and is related to the expenditure of both mental and physical effort by the subject (Rolfe and Lindsay, 1973). The unobtrusive nature of physiological measurement must be questioned when considering the nature of the equipment used to measure physiological responses. Not only is the physical presence of this equipment disturbing, uncomfortable and artificial, some equipment may have the effect of actively distorting primary task performance. Intrusion effects are thus a particular problem with physiological measures. Furthermore, the measurement of certain physiological reactions requires the use of sophisticated medical equipment, which can only be operated by highly qualified personnel.

Ease of implementation therefore becomes a factor to be considered when evaluating the viability of physiological measurement techniques.

Villegas and Wierwille (1979) assert that no single physiological measure is a sufficiently reliable workload index, and advocate the use of multiple measures. The true meaning of physiological responses can only be interpreted if there is a comprehensive knowledge and understanding of the task situation (Rolfe, 1976). Rolfe and Lindsay (1973) assert that a physiological response should perhaps be considered merely as an index of subjective load.

Commenting on the viability of physiological measures, O'Donnel (1979) argues "For many applications, there is no reason to resort to the procedures and technical

sophistication required to utilize physiological measures, when in relation to other techniques they are less powerful, needlessly complicated, expensive and insensitive" (p.63).

Subjective Opinion Measures

Subjective Opinion measures of mental workload involve eliciting subjective information about the demands imposed during task performance by administering either post-task rating scales or questionnaires to subjects, or alternatively by means of conducting interviews with subjects. The method assumes, however, that the subject can analyse and describe task demands.

Despite the creativity displayed by researchers in the field of subjective workload, in generating assessment techniques (Moray, 1982), the success attained in validating many of these measures against concurrently collected performance measures has unfortunately been severely limited. This dissociation between subjective and other measures of mental workload has been attributed to a lack of attention to the validation of subjective techniques and the development of standardized methods for their use (Vidulich and Tsang, 1986).

Many theorists base their validity judgements on the fact that subjective measures access 'operator orientated' determinants of mental load (Rolfe et al., 1974) (see the section Operator Orientated Factors earlier in this chapter for a discussion thereof).

The assumption is made that since subjective techniques evaluate workload determinants not considered by other techniques, they are therefore valid, and are subsequently incorporated into a test battery as a theoretical 'coverage of all options'.

Earlier attempts at subjective workload assessment suffered from a lack of theoretical background. Vidulich and Wickens (1986) argue that while the dissociation between subjective assessments and concurrently collected performance measures may well be attributed to the insensitivity of the former methods, if subjective techniques are "lawfully related to some aspects of the human information processing system, they can provide useful information" (p.291). As a result, researchers should concentrate on the specification of the appropriate uses for subjective measures rather than on the assessment of their overall validity. In line with this argument, later subjective techniques have come to acknowledge workload as being cognitively based, multidimensional and different across individuals, for example, the subjective workload assessment technique (SWAT) developed by Reid et al. (1981, 1981a), Casali's (1982) Multi-Descriptor Scale and the NASA Bipolar Technique developed by Hart et al. (1984).

While acknowledging that the major area of interest in cognitive task workload is at the internal processing stages, Vidulich (1986) argues that a lack of sensitivity to response demands could result in a reduction in the overall sensitivity of subjective techniques. In addition, Ericsson and Simon (1980) proposed that subjective methods are sensitive only to conscious mental processes, since unconscious processes are not perceived by the subject at all but the most general level (e.g., feelings of stress). Unconscious processes include many of those in long term memory as well as those procedural processes which are well practised and have become automatic. Given the many unconscious processes necessary for successful task performance, it would seem reasonable to assume that a considerable proportion of a subject's cognitive task experience would be rendered unavailable for subjective evaluation. Research conducted by Vidulich and Wickens

(1986) produced findings in support of these proposals.

A further threat to the sensitivity of subjective techniques is the representation of time varying demands in Subjective Workload Score (Vidulich and Tsang, 1986). Research findings concerning this issue have been contradictory. For example, whereas Thornton (1985) demonstrated that a short peak of very high workload within a longer task was likely to bias the final rating towards the peak value, both Yeh and Wickens (1985) and Vidulich and Tsang (1986) found that the overall Subjective Workload Ratings tended towards the average level, thus masking transient peaks and low levels of workload. Such contradictions indicate a need to conduct further investigation into the conditions under which each effect is likely to occur.

Probably the most obvious shortcoming of subjective assessments is between subject variability. In a study conducted by Vidulich and Pandit (1987) investigating the variability of a number of subjective techniques, it was demonstrated that all of the techniques investigated were highly susceptible to individual differences in interpretation.

Another problem concerning the scoring of subjective opinion measurements of mental workload is the development of a suitable scale. Although the majority of subjective techniques assume workload to be multidimensional, Vidulich and Tsang (1986) point out that a divergent number of dimensions are assumed to define workload adequately in each technique.

Furthermore, after deciding upon those workload dimensions to be evaluated, it is necessary to consider the number of levels to be discriminated. Too few rating choices may present too narrow a range of options to encompass all the

levels of workload experienced by subjects, whereas highly complex scales may require subjects to make fine discriminations between category descriptions. Vidulich and Wickens (1986) consider the selection of scoring systems for Subjective Opinion Techniques to be an issue of critical importance. In spite of this, the current trend in subjective workload research is to address the issue of dimensions to be assessed, rather than paying attention to the definition of the appropriate composition of scoring scales.

A major advantage associated with subjective measures, concerns their sensitivity to 'operator orientated' determinants of load, which are not reflected in other measures of mental load. For example, Vidulich and Wickens (1986) identified motivation as a factor strongly influencing the amount of effort a subject will exert for task performance, and suggested that highly motivated subjects would maintain good performance standards even under conditions of high load. Such increases in load will not be reflected by either Performance or Secondary Task measures. However, since increased motivation elicits a conscious decision to expend greater effort, higher workload should be well represented by Subjective Opinion Techniques (Vidulich and Wickens, 1986). Furthermore, subjective assessments have served to highlight the interesting paradox that periods of high workload may serve to increase job satisfaction and sense of fulfilment, thereby enhancing intrinsic motivation (Rolfe, 1973).

Further advantages of subjective measures include their ease of administration and low intrusion into primary task performance (Casali and Wierwille, 1983, 1984).

Vidulich and Wickens (1986) argue that, if appropriately designed to include relevant dimensions, rating scales with associated instructions and some method of reducing between

subject variability, Subjective Opinion Techniques could prove valuable in the assessment of mental workload.

Moreover, subjective assessments should be used in conjunction with other indices to provide a broader basis for evaluation and comparison (Williges and Wierwille, 1979).

The preceding discussion of mental workload measurement techniques has served to highlight the difficulties encountered when evaluating mental workload. These difficulties are attributed to both the complexity of the concept of mental workload itself; and the inabilities of workload estimation techniques to provide sensitive and accurate indications of load. Each of the measurement categories discussed is based within a certain theoretical framework, the aim of which is to achieve maximum sensitivity within that framework. Although this may prove advantageous for the validity of each method, it unfortunately results in the technique being insensitive to those workload dimensions excluded from that particular theoretical paradigm. Many workload researchers (e.g., Rolfe and Lindsay, 1973; Chiles and Alluisi, 1979; Moray, 1982) have stressed the inadequacy of using any single measure for assessing mental workload. A comprehensive evaluation of workload can be obtained only by combining information recorded at different stages and levels of the complex human information processing system. This is best achieved through the use of a battery of estimation techniques, each assessing a particular source of demand. For research purposes then, the issues requiring consideration include the composition of the battery, as well as the specific design of each technique to be used.

Chapter Summary

In the present chapter, Mental Workload was defined as 'the

extent to which an individual is occupied with meeting the demands made on his/her cognitive functions by the task in which he/she is engaged'. The nature of those demands, and of the individual's cognitive functions were explored in an attempt to provide some indication of the factors to which mental workload measurement techniques should be sensitive. The techniques themselves were discussed, - in terms of their sensitivity, intrusion and overall viability as measures of mental workload. It was concluded, because of the inability of a single technique to access all of the components of mental load, the best strategy would be to use a composite battery of measures for workload evaluation.

While taking cognisance of the points discussed and conclusions drawn in this chapter, the present research shall bring together the previously unrelated research paradigms of mental workload and componential analysis in an attempt to investigate the extent to which Sternberg's (1977) theory of analogical reasoning, in particular the linear componential models he proposed, provide information regarding the nature of task structure, information which could serve to enhance the present understanding of the concept of mental workload and its measurement.

By facilitating the possible identification of inappropriate work routines at the cognitive level of task performance, the present research may contribute to the development of a more sophisticated notion of task structure, which together with an enhanced understanding of the concept of mental workload, could stand to benefit future research into the development of strategies for reducing automated system user workload, experienced during the performance of complex cognitive tasks.

Having discussed the concept of mental workload, as well as the various mental workload estimation techniques, for

present purposes it remains necessary to consider the componential approach to mental abilities, in particular Sternberg's (1977) componential theory of analogical reasoning. This shall be the main focus of Chapter 3 (Cognitive Task Structure and Strategy).

CHAPTER 3

COGNITIVE TASK STRUCTURE AND STRATEGY

COGNITIVE ABILITIES AND PSYCHOMETRICS

The vast majority of theory and research on the nature of intelligence and mental abilities has been conducted within the context of a measurement paradigm (Pellegrino, 1987). The essential feature of this approach is to present a variety of cognitive tasks to a large sample of individuals. Scores on these tasks are then correlated and the resultant correlation matrix reduced by a multivariate procedure, typically factor analysis. The goal here being to obtain the simplest possible mathematical solution for the data set. The latter reflects varying consistency in individual differences over tasks. The basic assumptions underlying this approach are threefold: Individuals differ in performance on virtually any intellectual task they are required to perform; the differences on a given task are attributable to one or more underlying abilities; and the correlation of performance scores across tasks reflects both common and distinct abilities.

Over the past few decades this approach has been refined and applied to data sets of varying size and complexity, resulting in a variety of factor theories on structure of human intelligence, for example: Spearman's Two Factor Theory (Spearman, 1904, 1927); Holzinger's Bi Factor Theory (Holzinger, 1938); Thurstone's Multiple Factor Theory (Thurstone, 1938, 1947); Guildford's Structure of Intellect (SI) Model (Guildford, 1967).

In the past the psychometric approach has been hampered by

problems, stemming from both the contrasting interpretations of different sets of results and, the frequent inconsistency of results from different data sets.

Two reasons could possibly account for the occurrence of such problems (due to the fact that these reasons hinge upon the use of factor analysis as a method of data reduction, and have a significant bearing on the present study, an in depth discussion of the statistical concept of factor analysis will be presented in the section The Differential Approach to Cognitive Abilities later on in this chapter). First, the existence of numerous ways to factor analyse a single data set. Different apparent solutions can be obtained for the same data, which lead to different psychological interpretations. Second, the data obtained in separate studies differ. The critical data for any type of factor analysis are correlations between test scores, the patterns of which are dependent upon, for example, the actual tests administered and the characteristics of the group taking the test. Because studies vary with respect to inter alia the group tested, the total number of tests administered, and the specific tests included in the test battery, there exists a lack of comparability of separate studies, something which is further complicated by alternative methods of factor analysis.

Despite such problems, Pellegrino (1987) argues that certain general facts and conclusions cannot be overlooked. One pervasive fact being that when a range of intellectual problems (tests) varying in both content and format are administered to a sample of individuals the result is a matrix of positive test score correlations. The pattern of data supports a hierarchical model with a general ability, underlying all cognitive performances as well as several specific abilities (Spearman, 1904, 1927). Theorists pursuing this approach have endeavoured to

simplify the interpretation of correlations by way of general and specific abilities organised in a hierarchical or multidimensional scheme. "One way to view factor theories is in terms of a structural model of human intellect" Pellegrino (1987, p.268). Unfortunately factor theories are unable to provide a satisfactory means of defining exactly what a factor or ability might be. Essentially a factor is a hypothetical entity, an abstraction, that an individual possesses to varying degrees. But what is it that an individual actually possesses when he or she has high levels of general ability, or more importantly, scores well on tests of some specific ability? It is not clear what "g" is other than a mathematical abstraction. However, it is possible to be more precise about the tasks that tend to be loaded on this general factor and some of the underlying sources of individual differences. There is a relatively circumscribed class of cognitive tasks that significantly contributes to defining the general factor and that frequently constitutes all, or a significant portion of, tests of general intelligence (Jensen, 1980). These tasks all involve inductive reasoning, one of the major types of human problem solving (Greeno, 1978). All inductive reasoning tasks have the same general form: a set of elements of a pattern is presented and the task is to infer the rule structure in the pattern and to select the best possible completion. This generic task structure is manifest in different task forms such as classification, analogy, series completion and matrix problems using different types of content, for example, letters, numbers, words and geometric shapes.

At present there exists a comprehensive understanding of the cognitive processes and knowledge underlying performance on these various tasks. Theorists have attempted to develop and test explicit models of performance on these types of tasks and to use such models

as the basis of developmental and individual differences analysis (e.g., Evans, 1968; Sternberg, 1977; Mulholland et al., 1980; Pellegrino and Glaser, 1982; Goldman and Pellegrino, 1984). There exist a variety of sources of age and ability differences in performance on inductive reasoning tasks, ranging from speed of executing certain processes such as relational inference, to semantic and quantitative knowledge differences, to overall strategy differences.

Hunt (1974) suggests the possibility that differences in performance on certain inductive reasoning tests reflect a major difference in the strategy used to approach the task, and that strategy differences further reflect differences in understanding the constraints of the task. Studies of developmental and individual differences in performance on analogical reasoning tasks (e.g., Sternberg, 1977) clearly indicate that many individuals fail to evidence an understanding of the strategies necessary for task solution, strategies that are supposed to hold in formal analogical reasoning. In fact the instructions provided on any standardised test fail to make explicit the criteria to be used in performing the task (Pellegrino, 1987). Thus what is implicit and tacit knowledge for the test constructor may very well be unknown for certain individuals taking the test. Due to the fact that various types of knowledge differences, declarative and procedural, can be involved in producing overall performance differences exhibited by groups of individuals on tests of inductive reasoning ability, research needs to be conducted to ascertain the sources of such differences.

The issue is not whether a test is good or bad nor whether it serves to differentiate among individuals or groups of individuals, but rather how these differences are interpreted and the need to conduct analysis of the sources of differences that do exist. To achieve this would

require empirically validated theories of the processes and knowledge necessary to perform specific tasks, theories that would facilitate the prediction and explanation of performance patterns within and across individuals (e.g., Sternberg's (1977) componential theory of analogical reasoning). In short, what is needed, is to demystify existing measures of cognitive abilities. Merely labelling them as Level I/Level II (Jensen, 1968) or rote learning vs verbal ability vs g etc., is not an effective substitute for understanding exactly what it is that is being measured.

McNemar (1964) provided a summary criticism of psychometric research and theory on the nature of intelligence and mental abilities:-

"Studies of individual differences never come to grips with the process or operation by which a given organism achieves an intellectual response. Indeed it is difficult to see how the available individual difference data can be used even as a starting point for generating a theory as to the process nature of general intelligence or of any other specified ability" (p.881).

This general problem of measurement of cognitive abilities in the absence of understanding the measures themselves, namely the knowledge, cognitive structures and processes contributing to the performances being measured, severely inhibits our ability to explain the bulk of research on group and individual differences. This is not to say that psychometric research has failed to produce a wealth of information about developmental, individual and group differences in performance on a variety of cognitive tasks. However, the interpretation of such results as well as the limitations on such interpretation are another matter altogether.

Cognitive Style and Strategy Choice

Recent research in cognitive psychology has demonstrated that people exhibit significant individual differences not only in level of information processing ability, but also in the information processing strategies that they adopt during problem solving and other decision making activities (Robertson, 1985).

One of the most significant pieces of research on human information processing strategies was the work conducted by Bruner et al., (1956). In many respects this study marked the beginning of contemporary attempts to examine individual differences in information processing strategy. For the purpose of the present discussion, the major importance of the work of Bruner et al. (1956), lies not in the different strategies that they identified, but in the fact that they did demonstrate individual differences. However, it is important to emphasise that their research failed to produce data suggesting that people's strategies were consistent either over many trials or in a range of different problem solving situations.

Other investigators (e.g., Messick et al., 1976; Witkin, 1976, and Witkin et al., 1977) have, however, identified individual differences in information processing that display stability as far as the individual person is concerned and in some cases apply across a wide range of situations. The majority of those working in this area of individual differences in human information processing refer to such differences as differences in "cognitive style".

Extensive research has been conducted on cognitive style and, as a result, a variety of cognitive styles have been proposed and investigated, for example, Kagan's reflective-impulsive approach (Kagan et al., 1964), Pask's holist and

serialist strategy (Pask and Scott, 1972; Pask, 1976) and Witkin's field dependence-independence dimension (Witkin, 1976; Witkin et al., 1977).

Robertson (1985) notes that the literature on cognitive style consistently identifies cognitive styles as individual difference factors which are not merely different types of ability. Thus, the two poles of any cognitive styles dimension are thought to represent differences in information processing strategies and do not merely reflect whether someone displays more or less of a certain type of ability. Consequently there is assumed to be a qualitative difference in the type of processing that takes place. However, Robertson (1985) points out that the clear conceptual difficulty is difficult to reconcile with much of the research.

Another point of confusion identified by Robertson (1985) concerns the relationship (or lack of it) between cognitive style and cognitive strategy. The point he emphasises here is that although the various differences identified may in some ways appear similar and in others quite different, some of these apparent similarities or differences are likely to be, in part, a function of the specific measurement techniques or circumstances under which the differences in cognitive processes are exhibited. Messick et al. (1976) make this point clearly:-

"It is important to distinguish cognitive styles, which are high level heuristics that organise and control behaviour across a wide variety of situations from cognitive strategies, which are ... a function of the conditions of particular situations" (p.6).

Assumptions of Psychometrics and Strategy Choice

Differential methodology assumes that an ability or

achievement test measures the same general processes and structures for all examinees, and that between subject differences arise due to certain individuals having more of the ability or alternatively because they are able to use it more effectively. Consequently the differences are construed as quantitative rather than qualitative. Just and Carpenter (1985) disagree with this assumption and argue that its violation may well account for many of the confusions in the psychometric literature. For example, French (1965), in a study conducted to investigate the relationship of problem solving styles to the factor composition of psychometric tests, showed that different self-reported strategies (loosely characterised as global or analytic) in some psychometric tests resulted in different factor loadings. Based on the outcome of his study, French (1965) concluded that "there seems to be extensive evidence ... that tests, even simple 'pure-factor' tests do not measure the same thing for all people" (p.26). Furthermore, Just and Carpenter (1985) maintain that test items may sometimes systematically vary with respect to which strategy they evoke and, are of the opinion that there is likely to be both within subject and between subject contamination of the single strategy assumption in all but the simplest tests (referring specifically to tests of spatial ability). Such contamination could result in a test sometimes loading on one factor and sometimes on another factor if the two populations tested had different strategy preferences (French, 1965).

Moreover, it will be recalled from the Section Operator Orientated Factors in Chapter 2 that Bainbridge (1978) suggested the possibility that an increase in the demands imposed by a cognitive task could influence strategy choice. This assumption is supported by Just and Carpenter (1985), who emphasise that (psychometric) item and test complexity could well be major determinants of the information processing strategy elicited.

INVESTIGATING THE NATURE OF MENTAL ABILITIES

Numerous approaches have been taken towards explaining the structure and content of mental abilities. In the past, two major approaches have emerged, namely, the differential approach and the information processing approach. In their endeavour to understand the nature of mental abilities those pursuing the differential approach have concentrated on studying relationships between patterns of individual differences in scores on psychometric ability tests. The rationale motivating this approach is that a relatively small number of latent abilities are responsible for observed variation among individuals performance on these tests: Understanding of abilities is gained by identifying in the test data the sources of individual difference variation. Identification of these sources has typically been by means of correlational techniques, in particular factor analysis. Those following the information processing approach sought understanding of the nature of mental abilities by studying subjects' performance patterns on cognitive tasks with systematically varied attributes. These variations in task attributes, rather than variations in subject attributes, have provided the basis for inferring the nature of mental abilities. Methods for drawing inferences typically include content analysis of written or oral protocols, mathematic modeling of response time or error data, and simulation of cognitive processes on computer.

In recent years, various offshoots of these two approaches (i.e., cognitive correlates approach (Hunt, Frost and Lunneborg, 1973; Hunt, Lunneborg and Lewis, 1975), and componential analysis (Sternberg, 1977)) have sought to integrate the approaches in ways that would combine the best features of each. These approaches provide alternative means to investigate both between-subject and between-stimuli sources of variation, the respective

specialities of the differential and information processing approaches (Sternberg, 1979).

Since componential analysis constitutes the focal point of the present study, an indepth investigation of both the differential and information processing approaches is warranted, as this would surely facilitate a more comprehensive understanding of the development and composition of the componential approach.

THE DIFFERENTIAL APPROACH TO COGNITIVE ABILITIES

It was proposed in the preceding section that factor analysis contributed largely to the problems experienced by the psychometric approach to intelligence. This, together with the fact that factor analysis bears particular relevance to the present study, warrants a detailed discussion of this method of data reduction and its limitations.

Multiple Criteria

It frequently occurs that no single criterion is to be had which can be regarded as an unbiased indicator of the characteristic a certain psychometric test is designed to measure, but there are a number of available measures which may reasonably be regarded as partially related to it. In such cases a study of the correlations of the test with these partial criterion measures is often informative. In as much as the criteria are likely to differ with respect to their self-correlations, the correlation for attenuation should be applied in order to show the relative degree of relationship of the test to the different criteria apart from errors of measurement.

If the various partial criteria are of diverse kinds, multiple correlations may be used to show the extent of the

combined relationship to the test, in other words, the extent to which the test variance can be accounted for in terms of these partial factors taken in combination. It is of importance here, to emphasise that no assumptions as to causation are warranted on the basis of such a procedure. Correlation, whether simple or multiple, merely indicates the existence of common factors within the measures correlated.

Principles of Factor Analysis

In general, the position taken by factor analysts is that what appears to be an almost infinite variety of forms and levels of behaviour may be only the permutations and combinations of a relatively small number of underlying factors or directional tendencies which, taken in their totality, describe the individual in all his uniqueness.

Because any act (for example, responding to a mental test item) is the result of a number of these factors acting in combination, mental factors cannot be isolated and recognised by surface inspection nor measured in terms of superficial assumptions as to how and where they are manifested. But, as stated by Goodenough (1960), if a comparatively large number of measurements of related forms of behaviour or types of ability are made, such, for example, as would be given by a series of different kinds of intelligence tests, and the intercorrelations of these measures for a particular group of subjects are calculated and arranged in the form of a correlation matrix, "A certain orderliness in the numerical relationships of homologous terms can be noted" (Goodenough, 1960, p.226). Of significance, is that fact that although these observed relationships are mathematically sound, they bear no psychological relevance. Essentially, these relationships are unnatural, being dependent only upon the measurement scales used by the investigator.

The Purpose of Factor Analysis

Thurstone (1947) stated that "factorial methods were developed primarily for the purpose of identifying the principle dimensions or categories of mentality" (p.55). There have been two opposing schools of thought as to how this purpose should be accomplished. Both distinguish between stages of hypothesis formation and hypothesis testing. They differ, however, with regard to the stage at which it is believed that factor analysis is most useful.

On the one hand, there are those theorists who are of the opinion that factor analysis is a "useful tool in hypothesis formation rather than hypothesis testing" (Humphreys, 1962, p.475). Advocates of this viewpoint consider factor analysis as being a useful exploratory technique, but not a useful technique for choosing among the theories suggested by initial exploration (Sternberg, 1977).

On the other hand, there are those theorists who regard factor analysis "not as a source of hypotheses, but merely as a method of comparing, confirming or refuting alternative hypotheses initially suggested by nonstatistical arguments or evidence" (Burt, 1970, p.17). Advocates of this viewpoint regard the utility of factor analysis as not lying in the initial exploration of hypotheses, but rather in the selection of a preferred hypothesis from those suggested by initial exploration (Sternberg, 1977).

The intermediate position, namely, that factor analysis is useful both as a means of suggesting theories and as a means of choosing among them, appears to be the most common one.

Fundamental Unit of Analysis

The fundamental unit of analysis in factor analysis is the factor. However, there has always existed, among the advocates of factor analysis, a lack of consensus as to just what a factor is. Thurstone (1947) noted that "factors may be called by different names, such as causes, faculties, parameters, functional unities, abilities, or independent measurements" (p.56). Just as there have been differing opinions regarding the definition of factors, so too have there been opposing points of view regarding the nature of factors, namely, the descriptive and the causal. Burt (1940) took a descriptive view, arguing that "the primary object of factorial methods is neither causal interpretation, nor statistical prediction, but exact and systematic description" (p.63). Vernon (1971) likewise adopted a descriptive view, regarding factors as "categories for classifying mental or behavioural performances" (p.8).

Taking a causal viewpoint, Cattell (1971) has referred to factors as "source traits", and Guilford (1967) has described a factor as "an underlying, latent variable along which individuals differ."

The Objective of Factor Analysis

Factor analysis begins with a matrix, usually of intercorrelations among tests. The matrix represents the interrelationships between values (usually scores) on a number of different variables (usually tests). The main aim of factor analysis is to represent the set of variables in terms of a smaller number of underlying factors or hypothetical constructs, namely, to obtain the simplest possible mathematical solution for the data set (Pellegrino, 1987).

Major Decisions to be made in Factor Analysis

Method of Factor Analysis

There are numerous methods of performing factor analysis, for example, the principal-factor method, the principal-component method, the centroid method, etc. However, it is beyond the scope of this discussion to present an in depth review of the various factor analysis methods. Rather, for present purposes, it shall suffice to note that the choice of method will depend on the investigator's goals in performing the analysis. The choice of method, however, is generally less important than the choices described below.

Number of Factors

Usually the aim of an investigator will be to retain fewer factors than there are tests, since a major purpose of factor analysis is simplification. Difficulty may arise when deciding on how many factors to retain. Sternberg (1977) lists four criteria that are often used, the first three criteria being validity based, whereas the fourth criterion is reliability based. The first is a prior theory, which may specify the number of factors to be retained. The second criterion is the proportion of variance accounted for by each factor. Factors accounting for trivial proportions of variance in the data are generally of little interest. The third criterion is interpretability. In using this criterion, one retains only factors to which a meaningful substantive interpretation can be assigned. The fourth criterion is factorial invariance. If a certain factor reappears in study after study, it is reliable and probably of psychological interest, whereas if a factor appears only in an isolated study, it is less likely to be of psychological interest.

Type of Rotation

The orientation of factorial reference is not unique and, furthermore, an infinite member of orientations are mathematically permissible. It is agreed upon by factor analysts that part of their job is to decide upon a psychologically appropriate orientation, but there has been considerable disagreement as to what rotation should be performed - or whether one should be performed at all.

The most commonly used factor-analytic methods generally yield a factor structure in which the first factor is a general one, with most factor loadings having the same sign, and subsequent factors are bipolar, with approximately half of the loadings positive and half negative. Factor theorists appear to have followed either of two approaches to providing a basis for psychological theorising. The first approach involves using the unrotated factor solution as the basis for psychological theorising. Such solutions tend to support the notion of a strong general factor pervading virtually all mental ability tasks. The second approach involves using the rotated factor solutions as bases for psychological theorising.

Theorists preferring rotation (e.g., Guilford, 1967) have argued that the unrotated factor structure is indicative of mathematical rather than psychological regularities. These theorists maintain that psychologically meaningful factors can be obtained only through rotation.

By selecting the option to rotate, a theorist will be required to make two fundamental decisions. The first is whether to use an objective or subjective form of rotation. The second is whether to require orthogonal factor axes or to permit oblique ones. Horn (1967) has made a useful distinction between objective and subjective

rotation. He describes an objective rotation as one in which all investigators who use the method on the same data will arrive at the same result, regardless of differing a priori hypotheses they may have regarding the outcome, and a subjective rotation as one in which investigators with different hypotheses may get different results from the same set of data.

Factor axes may be either orthogonal or oblique. When orthogonal axes are used, factors are uncorrelated. When oblique axes are used, factors may be correlated. Performance on one factor is not independent of performance on every other factor.

The most common solution to the rotational problem is rotation to what Thurstone (1947) called simple structure. For many years, a precise mathematical algorithm for rotation to simple structure eluded factor analysts. However, within the space of a year, Carroll (1953) and Neuhaus and Wrigley (1954) independently arrived at mathematically equivalent algorithms that provided an approximation to simple structure. Their technique which simplified the rows (tests) of the factor structure matrix, was called the quartimax method. Kaiser (1958) developed an alternative, the varimax method, which simplifies the columns (factors) of the structure matrix. Since this outcome more closely approximated simple structure and tended to yield more interpretable solutions than the quartimax method, it has become the most widely used form of rotation. It is an objective method, and results in orthogonal axes.

Sternberg (1977) maintains that the choice of factor rotation method is of particular importance from a psychological viewpoint, because it constitutes a major distinguishing element among various factor-analytic theories of intelligence.

The differential approach to intelligence was for many years a most productive one. Through factor analysis, it "Provided the characteristic trend of thought for two generations of ability researchers; permitted the investigation of complex psychological constructs; generated theories at a macrotheoretical level; provided a systematic means for studying individual differences, fostered a symbiotic relationship between theories of measurement and intelligence that lead to greater sophistication in each; and proved itself relevant to the needs of applied settings" (Sternberg, 1977, p.35). Because of these contributions, and because there was for many years no viable alternative to the differential approach, it was in no way surprising that it thrived for as long as it did.

The Decline of the Differential Approach

Despite the earlier successes of the differential approach, in recent times interest in the approach has declined considerably. The decline has been attributed to the fact that its practical tool, factor analysis, could no longer meet the numerous demands placed upon it. Consequently the entire differential research enterprise became strained. The sources of strain were twofold:- misuses of factor analysis methods and limitations of these methods.

Misuse of Factor Analysis

Many misuses of factor analysis stem not from failing to inspect the data, but from using poor judgement in inspecting them. McNemar, (1951) described ten of the most common problems:-

1. Use of samples too small to permit stability;
2. Overinterpretation of small factor loadings and disregard for sample size in deciding upon the minimum loading to interpret;

3. Irrational decisions regarding rotation of factors;
4. Use of unreliable tests;
5. Lack of experimental independence among variables that are factor analysed;
6. Selection of tests so as practically to guarantee a predetermined factor structure;
7. Use of tests from domains that are too limited to demonstrate anything and use of tests from separate domains known to be uncorrelated;
8. Scanning of tests in a biased fashion when interpreting and naming factors;
9. Struggling to make sense out of results;
10. "Varying all over the map in the use of hypotheses from those who started with no hypotheses, used none in the process, and ended up with none, to those who start with hypotheses, use them en route, and end with hypotheses" (McNemar, 1951, p.358).

Guilford (1952) has likewise described "some common faults in factor analysis". He too noted ten problems:-

1. "Too many factors are often extracted for the number of experimental variables;
2. Too many experimental variables are factorially complex;
3. Sometimes a common factor fails to come out because it is substantially represented in only one experimental variable;
4. Not enough factors are extracted;
5. Correlation coefficients used in analysis are often spurious;
6. Correlations of ipsative scores are sometimes used in an analysis;
7. A pair of factors is very much confined to the same experimental variables;
8. The population on which the analysis is based is heterogeneous;

9. Not enough attention is given to requirements for correlation coefficients;
10. Difficulty levels of tests often vary substantially" (Guilford, 1952, pp.7-33).

When comparing these lists of McNemar and Guilford it becomes apparent that the actual items are virtually nonoverlapping. Thus, it is not surprising that, with twenty common faults, factor analysis ran into trouble. However, a number of theorists (e.g., Vernon, 1970, 1971; Guilford, 1974; Sternberg, 1977) believe that the most serious trouble resulted not from misuses of the method, but from limitations of factor analysis.

Limitations of Factor Analysis

The limitations of factor analysis have constituted the main source of strain upon the differential approach to intelligence. The limitations to be described pose serious problems only when factor analysis is used for purposes for which it is not suited. Sternberg (1977) proposes four reasons for its (factor analysis) lack of suitability as an appropriate method for discovering the components underlying intelligence, each corresponding to an intrinsic limitation of factor analysis. The first two limitations are statistical in nature and because of these limitations, factor analysis lacks inferential power for distinguishing among theories. The second two limitations are psychological in nature and because of these limitations, factors could not be components of intelligence.

Lack of control over mathematical realisation of psychological theory: The machinery of factor analysis rather than the investigator formulates and thus has control over the mathematical theory, resulting in a reduced ability of the investigator to perform theory comparison operations. The problem is that a rather wide

variety of psychological theories may be compatible with the results of a single factor analysis and it is not possible to distinguish among them mathematically.

Solution indeterminacy: The rotation dilemma: Guilford (1974) has referred to the need for rotation as "the most serious weakness" (p.498) of the factor analytic method. The crux of the problem being that while factor space is unique, the orientation of axes is not. The axes may be rotated in an infinite number of ways, each defining factors along different dimensions. The different dimensions have different psychological implications. For example, an unrotated solution will often support a theory of a general factor followed by bipolar group factors and possibly specific factors. Rotation to simple structure will often support a theory of group or multiple factors in the absence of a general factor. Subjective rotation to minimise discrepancies between predicted and observed factors can yield still different outcomes. As a result, solutions are indeterminate because of the infinite number of possible axis rotations.

Failure to discover or explicate process: Factor analysis provides no way to discover or explain the processes that in combination constitute intelligent behaviour. The problem is that factor analysis is capable only of analysing interitem structure in a test, whereas a process analysis would require analysis of intraitem structure. It would require clarification of the steps involved in the solution of a single item. Factor analysis "deals only with the end products of human thinking and behaviour, and throws little light on how these products come about in individual human beings" (Vernon, 1971, p.7). Because of this, factorial studies do not "enable us to decide what the basic components of mental organisation are" (Vernon, 1970, p.100).

Interindividual nature of analysis: The components of intelligence are intraindividual - they exist within individual subjects. Factor analysis, however is generally interindividual - it analyses patterns of individual differences across subjects. Since individual differences are meaningless in the context of one individual, it is not clear how factor analysis could enable us to discover what the components within an individual are. While Sternberg (1977) acknowledges that certain modes of factor analysis could be used intraindividually, he argues that it has not been shown that they could discover underlying components of intelligence.

The statistical limitations of factor analysis, described above, can be overcome. The advent of confirmatory maximum likelihood methods of factor analysis made it possible to use factor analysis for sophisticated testing of prior hypoth. and the attainment of a unique solution (not subject to rotation). It is possible, even likely that the second two substantive limitations can be overcome as well, however, "it remains to be shown, though, that factor analysis can be performed so as to dispense with all four limitations simultaneously" (Sternberg, 1977, p.29).

Although factor analysis can be of considerable value in the study of human abilities, it is not necessarily a direct means towards the discovery of the basic underlying components of mental ability. Unfortunately, this has been one of the most common purposes for which factor analysis has been used.

THE INFORMATION PROCESSING APPROACH TO COGNITIVE ABILITIES

The information processing view of human abilities as dynamic information processing capabilities provides an

alternative to the differential view of abilities as static psychological factors. However, as Sternberg (1977) emphasises, information processing methodology was developed not as an alternative to the factor-analytic concepts of differential psychology, but as an alternative to the stimulus-response concepts of behaviourism. The research of information processing theorists has taken a variety of forms, from the creation of detailed computer programmes to the formulation of complex mathematical models. "The common new emphasis was the exploration of complex processes and the acceptance of the need to be explicit about internal, symbolic mechanisms" (Newell and Simon, 1972, p.4).

The Fundamental Unit of Analysis

The fundamental unit of analysis in information processing psychology is the elementary information process (EIP) (Newell and Simon, 1972). The assumption here, being that all behaviour of a human information processing system is the result of sequences of these elementary processes. They are elementary in the sense that they are not further broken down into simpler processes by the theory under consideration. The processes must be well defined, and the collection of them must be sufficiently powerful to compose all microscopic performances of the human information processing system (Newell and Simon, 1972).

PRINCIPAL ORGANISING CONCEPTS WITHIN THE INFORMATION PROCESSING APPROACH

The Structure-Process Definition

As is the case in all scientific theories, the structure versus process distinction is apparent in information processing models of human intelligence. However as Estes (1977) points out, in this context it is less clear than in

the case of the physical sciences just what sense of the term structure is intended. The prominent viewpoint in experimental psychology has been that cognitive activities are manifestations of, or consequence of, processes in the brain. As a result there has been a natural desire to identify the neurophysiological processes responsible and to specify their logic in the nervous system. However, progress towards this goal has been meager, with only a few segments of research in cognitive psychology being directed to this problem (e.g., Nebes, 1974; Harnad, Doty, Goldstein, Jaynes and Kraithamer, 1976).

In general, "structural concepts in information-processing models, and more broadly in nearly all cognitive theories, are purely hypothetical" (Estes, 1978, p.7). Structural concepts are introduced in an attempt to provide characterization of the state of information preserved by an individual at particular points in time during the performance of a cognitive task. The goal being to provide a characterization that is of sufficient theoretical significance to help in the prediction of the further course of information processing. Therefore, the structural concepts need to "entail the analysis of stored information into constituents and the expression of relations between these constituents in some systematic way" (Estes, 1978, p.7).

Attempts at achieving this goal have been guided by two principal lines of thinking. The first is that of analogies to physical structures whose description is possible in logical or mathematical terms. The most common example is the analogy to electrical circuitry by way of network models in which nodes or control elements are defined to correspond to units of stored information and pathways between these nodes provide for probing of the network and retrieval of the stored information. The second analogy is that of computer programmes, with

structural concepts taking the form of lists, vectors, or arrays of stored information.

The processes of interest in information processing models are, most importantly, those responsible for changes in structure (that is, for learning) and for the accomplishment of tasks by way of the structures (cognitive operations).

Structure of Memory

Estes (1978) maintains that some of the most influential and pervasive changes in thinking about human cognition associated with the rise of the information processing approach have to do with the structure of memory.

Within the information processing approach, the traditional distinctions between perception, memorising and thinking are clouded. Information input is viewed along the lines of a computer analogy, the assumption being that the individual imposes elaborate organisations on material during the transition from sensory registration to memory storage. At the heart of memory organisation, in this view, is the concept of coding. What is stored are not replicas of experiences, but rather representations encoded in terms of features, that is, values of attributes on various dimensions. Features, however, are not just any attributes, but rather are elements of organised processing systems.

Corresponding to the notion of features is that of detectors, which in the information processing literature, are elements of a well established long-term memory system. These elements are conceived to be organised in terms of a more or less hierarchical arrangement or levels, the detector elements at each level serving as points through which incoming information is filtered and encoded (Estes, 1975, 1977).

Incoming sensory information from the receptors is believed to be projected to the lowest level of the filter system, if an attribute of the sensory input matches an element at the feature level, then a message is transmitted to the next higher level, but if no match occurs, the transmission is blocked. A consequence of this processing is that the record stored in memory for a given experience is conceived to take the form of a list or vector containing a record of the features or feature combinations that were activated at each of the levels of the memory system during processing of the input. The system is regarded as dynamic in that filter elements corresponding to newly experienced combinations of features at any level are added to the system and then enter into processing of later inputs.

Most important, is to note that within information processing theory, the contents of memory are taken to include not only records of past experiences, but also organised lists of instructions to be carried out on future occasions that satisfy particular conditions. This property of the system enables the individual to operate in a manner similar to powerful computer programming models based on the conception of production systems (Newell, 1973; Anderson, 1976).

Operations and Stages

The utilisation of memory structures in information processing models of cognition is conceived to be accomplished by collections of elementary information processing operations, including, least of all, those performed by a digital computer in carrying out similar tasks:- information input is encoded and associated with addresses; memory is searched for items of information processing specified attributes; the retrieved units or the results of computations performed on them are compared with aspects of the input, and decisions and actions are based on the outcomes of the comparisons.

If, according to Estes (1978), it is assumed that, in general, all of these operations proceed simultaneously when an individual is carrying out a cognitive task, it would prove difficult to gain a foothold on the experimental analysis and measurement of the various components. Estes goes on to argue that regardless of the ultimate theoretical standing of the conception of the individual as primarily a serial processor, it has had beneficial heuristic value. Most critical is the working assumption that each member of a sequence of elementary operations takes an appreciable amount of time and that these times can be ascertained by subtracting observed reaction times for the accomplishment of various portions of the task.

It has been argued that the serial processing conception is oversimplified and that actually much parallel processing goes on in the human problem solver, in contrast to the digital computer (Townsend, 1974; Anderson, 1977). Nonetheless, the serial conception has proven enormously fruitful. Besides providing the basis for reaction-time methods to explore the organisation of elementary information processing operations, the serial conception has provided a way to co-ordinate structural with processing concepts by way of stage models (Estes, 1978).

Current information processing models are characteristically organised in terms of an assumed sequence of processing stages corresponding to the levels of organisation of feature detectors and higher order units in the memory system through which the information is filtered (Estes, 1976, 1977). Just as the conception of a serial concatenation of elementary operations, the notion of a succession of discrete stages of processing is without doubt oversimplified, but nonetheless, "this approach is proving enormously fertile in opening up the possibility of exact measurement of component cognitive processes and must

be the starting point for anyone now interested in understanding the research strategies and tactics of current information processing approaches" (Estes, 1978, p.11).

INFORMATION PROCESSING METHODOLOGY

Artificial Intelligence versus Simulation

The information processing approach encompasses a broad range of perspectives and research techniques. The broadest distinction being between artificial intelligence and simulation. Artificial intelligence research is directed toward the creation of models (usually implemented on a computer) that perform complex tasks usually considered to require intelligence. The objective here being to optimise performance, with little or no regard to whether the optimal way of performing a task is also the one that humans use. Simulation research on the other hand, focuses on the creation of models that simulate the processes a human subject is theorised to perform mentally. Computer simulation has been used in two essentially different ways to investigate mental abilities. The first form of investigation using this approach involves the compilation of programmes designed to simulate human performance on actual ability test items. In such cases, the computer programme itself serves as the theory of the mental abilities under investigation. In the second form of investigation using this approach, Polson and his colleagues (Atwood and Polson, 1976; Jeffries, Polson, Razran and Atwood, 1977) have used computer simulation of complex problem solving tasks to test an information processing theory whose complexity does not permit an analytic test. In these cases, the computer programme tests an information processing theory by simulating human behaviour, but the programme is not itself the theory.

The two perspectives (i.e., artificial intelligence and simulation) may be seen as end points of a continuum rather than as dichotomous. However, since the present study is concerned with human abilities, attention will be directed toward simulation models of human behaviour.

Slow Process versus Fast Process Research

The simulation approach may be further dichotomised into what Sternberg (1977) refers to as slow process and fast process research. As was the case with the artificial intelligence versus simulation distinction, so too is this distinction continuous rather than discrete. Slow process research deals with tasks that are usually performed in the order of minutes, and that are somewhat accessible to introspection. The cryptarithmic, logical proof, and chess tasks studied by Newell and Simon (1972) fall into this domain, as does the hobbits and orcs (missionaries and cannibals) task studied by Greeno (1974) and Thomas (1974). Fast process research deals with tasks that are usually performed in the order of milliseconds, or at most a few seconds. Tasks of this nature include the memory scanning task studied by Sternberg (1969), the sentence-picture comparison task studied by Clark and Chase (1972), and retrieval from semantic memory (Collins and Quillian, 1969). Generally, such tasks are not readily accessible to introspection.

Slow process research has made heavy use of protocol analysis and computer simulation as methods for studying behaviour. Fast process research on the other hand has relied on response time analysis and mathematical modeling as methods for studying behaviour, since these tasks are performed with such speed that subjects are usually unaware of how they proceeded.

The Subtraction Method versus Additive-Factor Method

Fast process research involving the modeling of human behaviour often proceeds along the lines of either the subtraction method of Donners (1848) or the additive-factor method first proposed by S Sternberg (1969). The distinction between these two methods is stated in some detail by Sternberg (1969) and by Pachella (1974). However, for the purpose of present study, these methods will be described only in summary fashion. The subtraction method, which has been the most commonly used,

"... is applicable when the performance of an experimental task involves the sequential action of a series of discrete mental events. In order to measure the duration of one of these mental events, the reaction time for an experimental task containing the event as a subprocess is compared to that for a comparison task that differs from the examination task only by the deletion of the process of interest The difference in reaction time for these two conditions will then be equal to the duration of the isolated process" (Pachella, 1974, p.46).

In actual use, multiple processes will often be of interest, and several tasks may be included in the experimental design that systematically vary the processes required. The critical feature of the method is a set of tasks that, by successive deletions, enables the investigator to estimate the amount of time spent on each component process of interest.

The assumptions of additive-factor methodology are threefold: Information processing comprises a sequence of independent stages; each of these stages receives an input from the preceding stage, transforms it, and then passes it along to the next stage; and the transformation produced in

each stage is independent of the duration of the prior stage.

The assumptions regarding stages have several important implications for the relationship between stage durations and experimental manipulations:

"First, total reaction time is simply the sum of the stage durations. When an experimental manipulation affects the reaction time for a particular information processing task, it does so by changing the durations of one or more of the constituent stages of processing. Second, if two different experimental manipulations affect two different stages, they will produce independent effects on total reaction time. The effect of one manipulation will be the same, regardless of the level of the other variable. In other words, the effects of the two experimental factors mutually modify each other's effect, that is, if they interact in a statistical sense, they must affect some stage in common" (Pachella, 1974, p.52).

The additive-factor method is utilised by studying an information processing task in the context of a multifactor experimental design. The implications previously mentioned then enable the investigator to make inferences about stages of information processing. Specifically, the investigator examines whether the effects of each pair of experimental factors are additive or interactive. Additive factors suggest different stages of processing, whereas interactive factors are hypothesised to affect stages in common.

Advantages of the Information Processing Approach

The information processing approach suffers from none of the limitations of the factor analytic methods highlighted

in the section The Decline of the Differential Approach earlier in this chapter. Rather, the advantages of this approach are:-

The investigator has control over the mathematical (or computer) model. This facilitates the comparison of multiple models on the same set of data, or the application of the same model (or a controlled variant of it) to different sets of data.

There is no rotation dilemma, because unless confounded, the parameters in a given parameter space are unique.

Information processing theories specify the processes that are used in the solution of single items.

Information processing theories are intraindividual. They are based either upon the data for a single individual, or upon the data for the average of multiple individuals. Multiple individuals and individual differences are not required either for theory formulation or theory testing.

The first two advantages bestow upon information processing methodology a far greater degree of inferential power than is available through differential methodology. As a result it is possible to make strong inferences regarding the relative superiority of one theory over another and to reject theories that are clearly inappropriate (Sternberg, 1977).

The second two advantages facilitate the utilisation of information processing methodology for the purpose of studying the "latent traits" that underlie performance in tasks requiring various mental abilities. The elementary information process is a suitable unit for characterising the latent traits.

Although information processing methodology enables the investigation of a wide variety of abilities as well as the detailed specification of the elementary processes and accompanying mechanisms that are responsible for task performance, the methodology does suffer from certain limitations.

Limitations of Information Processing Methodology

Here attention will be focused on simulation methodologies, which include the computer simulation methods used to study slow processes, and the mathematical methods used to study fast response times. Response times, in turn, have been studied both by the subtraction method and by the additive-factor method.

Computer Simulation

Sternberg (1977) argues that computer theories are inaccessible to the scientific public: in practice, computer theories have a degree of inaccessibility that may inhibit the public nature of scientific enterprise. The increasing sophistication of psychological theory is due at least in part to the opportunities for psychologists to replicate or refute the results of others. But opportunities for progress are diminished when the theories are inaccessible, buried in computer programmes that are difficult to acquire and understand.

Computer theories lack parsimony: "Parsimony for its own sake may be of questionable value. But ignoring parsimony can lead to unfortunate results" (Sternberg, 1977, p.55). Computer theories contain many statements, most of them interactive with other statements, so that it becomes difficult to assess what effect each one has, and whether a given statement is really a necessary ingredient of the theory. Simon and Kontovsky (1963) and Kontovsky and Simon

(1973), for example, describe the differences in performance among more and less powerful variants of their theory, but they do not clarify what distinguishes one variant from another. With computer programmes, simple characterizations are difficult to come by. Lack of parsimony further contributes to the difficulty other members of the scientific community have in testing the theory.

Computer theories do not generally provide process parameter estimates: An investigator may want to know the degree to which each elementary information process contributes to the latency or difficulty of a given problem. This information is difficult to obtain from a computer-based theory because of the large numbers of interactive variables involved in most computer programmes.

Mathematical Models of Response Time

The Subtraction Method

Parameters are often confounded: Explicit and implicit confounding of parameters are fairly common in response time studies using the subtraction method (e.g., Clark and Chase, 1972).

Alternative models are often indistinguishable: Alternative models become indistinguishable when the parameters that could distinguish them are confounded.

Parameter estimates are based upon too few degrees of freedom for residual: The degrees of freedom problem has become particularly severe in recent years, as research into more complex thought processes has made it less unfashionable to use large numbers of parameters with few data points. Parameter estimation techniques (such as multiple regression) are very prone to capitalisation upon

error in data. Unfortunately, cross-validation of parameter estimates is rare, although not entirely neglected (e.g., Anderson, 1974). In short, increasing the degrees of freedom for residual is a responsible way of guarding against spurious "good fit".

The "degrees-of-freedom" problem is of less concern when investigators conduct proper goodness of fit tests for their models. However, when goodness of fit is not tested, the percent variance accounted for by a given model can be a misleading descriptive statistic unless the statistic is estimated with a large number of degrees of freedom to spare. And these tests do nothing to solve the problem of unstable parameter estimates.

Ordering of parameters is not mathematically specified: Information processing models always specify the order in which elementary information processes are executed, but the mathematical realisation of these models may have no order constraints.

Results of external validation may be distorted: When relating individual differences in parameter values either to each other or to individual differences in other variable values, correlational patterns will be distorted if there are confounded parameters, and these are not very highly correlated with each other. Furthermore there is no a priori reason to expect confounded parameters to be highly correlated, and they may even be negatively correlated.

The criticisms just mentioned are not those usually made of the subtraction method, but are those made by Sternberg (1977). There are, however, two criticisms that are frequently made (Pachella, 1974; S Sternberg, 1969).

1. "The starting point for the application of the method is a relatively sophisticated one: In order to construct a comparison task, one must already know the sequence of events that transpires between stimulus and response. Such sophisticated knowledge is rarely available Obviously, the conclusions reached on the basis of the application of the method can be no stronger than the substantiation of the initial conceptualization of the experimental tasks" (Pachella, 1974, p.47).

It is true that the subtraction method requires a strong model. However, Sternberg (1977) views this as an advantage rather than a disadvantage. Firstly, because the method requires a strong model that is inferentially powerful, it is thus capable of rejecting the model. Secondly, a good experiment will often be designed so that it is capable of rejecting models or classes of models other than the investigators preferred one.

The criticism regarding the sophisticated starting point of the subtraction method provided one of the motives behind S Sternberg's development of the additive-factor method. Yet, Sternberg (1977) argues that it is not clear whether in practice there is as great a difference between initial conceptualizations as has been claimed.

Haphazard selection of factors for a multifactor experiment cannot be expected to yield enlightening results. The investigator should choose factors that (s)he believes are best able to differentiate stages of processing, and in order to choose them in this way, the investigator must have at least some prior idea of what stages different patterns of additivity and interaction would imply. As Calfee (1976) points out,

"The first step in this paradigm is the analysis of the underlying cognitive operations required to perform a task" (p.24)."

The investigator should at least have an idea of what the maximal stage model will look like, because there is no point to including in the experimental design a factor that could not conceivably be additive with the others. This requirement is comparable to that for an investigator designing an experiment using the subtraction method: He/she should have an idea of the maximum number of parameters and what they might mean. It is always possible to test alternative models that differ solely in number of parameters, and thereby to discard superfluous parameters.

2. A second general criticism of the subtraction method concerns the comparability of the experimental and comparison tasks, or as Sternberg (1969) has called it, the assumption of pure insertion. This refers to the assumption that it is possible to delete (or insert) complete mental events from an information-processing task without changing the nature of the other constituent mental operations.

"Experimental operations ... which might be thought of as deleting entire stages without altering the functions of other stages, are probably very rare; they should be considered special cases" (Sternberg, 1969, p.280).

While the validity of the assumption of pure insertion has been challenged for some time now (e.g., Kulpe, 1895), it seems that the criticism should be properly directed not at the assumption but at the failure of many users of the assumption to test it. Perhaps it was true in former years that little was known "about

how to test the validity of any particular application of the subtraction method" (Sternberg, 1969, p.277). But this is no longer the case.

This criticism of the assumption of pure insertion provided a second motive for the formulation of the additive-factor method. The additive-factor method was designed to make the weaker assumption that factors would change only the duration, but not the inclusion, of various stages. As Pachella (1974) has observed, however, "The manipulation of factor levels may cause a fundamental change in the processing sequence as may happen with the deletion of an entire stage within the Subtraction Method. From a procedural point of view, the difference between the Subtraction Method and the Additive Factor Method can be quite subtle" (p.57). As Pachella points out, the distinction between methods is a matter of degree rather than kind. It is by no means clear that experimental factors will always affect only duration and not inclusion.

With regard to the subtraction method, it is not clear how one can demonstrate the validity of the assumption of pure insertion, if the model is true or close to the true model.' However, as Sternberg (1977) points out, it is not clear how the user of the additive factor method can demonstrate that only duration and not inclusion of stages is affected. If stage inclusion is affected, interpretation of the pattern of factor additivity and interaction will be compromised.

The Additive Factor Method

Sternberg (1977) suggests four limitations of the additive-factor method: It does not reveal stage duration: it provides no indication of stage order; it provides no direct indication of the substantive

interpretation of any stage and the fundamental assumption regarding the means of identifying separate stages is questionable. "In situations where stages have some independent definition, it is perfectly conceivable that two factors might affect a single stage in an additive manner or interact" (Pachella, 1974, p.58).

Common Limitations

Certain limitations are common to both the subtraction and additive-factor methods, and to computer simulation methodology as well.

Information processing methodology does not provide a means for systematically studying correlates of individual differences in performance. In order to examine either consistencies in patterns of individual differences across different task parameters, or between task parameters and external measures of performance, the correlational methods of differential psychology are needed to accomplish this goal.

Information processing methodology does not provide a common language across tasks (and investigators): As information processing methodology has become more widely used, the list of parameters corresponding to hypothesised elementary information processes has grown steadily (Sternberg, 1977). Thus the problem may arise of establishing whether the a parameter of one task is the same or different from the x parameter of another task. In most cases differences will be clear, but identity is difficult to establish. The correlational methods of differential psychology are useful in solving this dilemma. If values of the two parameters are very highly correlated across individuals as well as tasks, chances of the parameters being equivalent are increased, although a high correlation is only a necessary but not sufficient

condition for parameter equivalence. If individual differences in the parameters are unrelated or weakly related, the parameters cannot represent the same underlying process.

Information processing methodology does not prevent overevaluation of task specific components: A particular parameter may account for individual differences in performance of a particular task. But if it accounts for individual differences in performance only of that task, it is of limited interest. A shortcoming of information processing research has been that despite the extensive study of highly specialised information processing tasks there have been almost no attempts to show that the parameters that account for individual differences in single specialised tasks are generalisable to other tasks. Correlational methods provide a way of demonstrating generalisability (Sternberg, 1977).

It will have been noticed in the preceding discussion that the information processing view of cognitive abilities as dynamic information processing capabilities provides an alternative to the differential view of abilities as static psychological factors. The emphasis in information processing research is on the exploration of complex processes and on being explicit about internal mechanisms. Although the information processing approach does not suffer from any of the intrinsic limitations of factor analysis, it is not without its own shortcomings.

Having discussed both the differential and information processing approaches to investigating cognitive abilities, highlighting their respective strengths and weaknesses, attention will now be turned to more contemporary approaches, that have sought to integrate the differential and information processing approaches in ways that would combine the best features of each, i.e., the cognitive

correlates approach (Hunt et al., 1973, Hunt et al., 1975) and componential analysis (Sternberg, 1977).

CONTEMPORARY APPROACHES TO UNDERSTANDING COGNITIVE ABILITIES

Recently, a number of different approaches have been pursued by cognitive psychologists in their quest to gain a comprehensive understanding of mental abilities. Sternberg (1981) lists as examples: The cognitive correlates approach; cognitive components approach; cognitive training approach; and the cognitive contents approach. The most viable contemporary approaches to mental abilities all integrate differential and information processing techniques in some way. However, differences tend to emerge depending on whether they are outgrowths primarily of the differential or information processing approach (Sternberg, 1979). For the purposes of the present study attention will be focused on two contrasting approaches to the experimental analysis of individual differences, namely, the cognitive correlates approach and the cognitive components approach. Albeit that the former approach has been classified as being an outgrowth of the differential approach, whereas the latter approach has been classified as stemming from information processing techniques (Sternberg, 1979), Pellegrino and Glaser (1979) note that both approaches attempt to understand the mental activities that contribute to individual differences as measured by psychometric instruments. The cognitive components approach will, however, be discussed in far greater detail than the cognitive correlates approach, due to the central role that componential task analysis will play in the present investigation.

THE COGNITIVE CORRELATES APPROACH

The cognitive correlates approach is best typified by the

work of Hunt and his colleagues (Hunt, Frost and Lunneborg, 1973; Hunt, Lunneborg and Lewis, 1975; Hunt and Lansman, 1975; Hunt, 1978).

This approach involves the testing of subjects' ability to perform tasks that contemporary cognitive psychologists believe measure basic human information processing abilities.

Sternberg (1981) describes the proximal goal in cognitive-correlates research as being the estimation of parameters (characteristic quantities) representing the duration of performance for the information processing components constituting each task, and then the subsequent investigation of the extent to which these components correlate across subjects with each other and with scores on measures commonly believed to assess intelligence (e.g., Raven's Progressive Matrices). Most commonly, correlations between parameter estimates and measured intelligence are statistically significant, but moderately low - unusually around .3 (Sternberg, 1981).

The distal goal of cognitive correlates research is the integration of individual differences research - in particular by providing a theoretical grounding from cognitive psychology for differential research (Hunt et al., 1973). Thus, Sternberg (1981) maintains that instead of trying to draw theoretical conclusions from correlating scores on one empirically derived test (e.g., reasoning) with scores on another empirically derived test (e.g., vocabulary) as differential researchers have done, the cognitive correlates approach advocates the drawing of theoretical conclusions from correlating scores on an empirically derived test with parameters generated by a cognitive model of some aspect of mental functioning (e.g., memory scanning).

While giving credit to cognitive correlates researchers for providing a cognitive-theoretical base for individual differences research to supplement the psychometric theoretical base that had existed earlier, at the same time, Sternberg (1981) questions whether cognitive correlates researchers are providing the optimal cognitive theoretical base. He suggests that the relatively low correlations attained between task parameters and psychometric test scores may be attributed to psychometric test performance drawing on lower level perceptual and memory abilities of the kinds studied by cognitive correlates researchers, but only in a peripheral way. Moreover, Sternberg emphasises that it is not clear whether the rather simple kinds of information processing required by many of the perceptual and memory tasks cognitive psychologists have studied, do justice to the rather complex kinds of information processing required by many psychometric tests of mental abilities.

Sternberg (1981) contends that until a better understanding of the .3-level correlation between parameter estimates and measured intelligence is acquired, a final evaluation of the cognitive correlates approach will not be possible. To date, no research in this regard has been conducted. Nevertheless cognitive correlates researchers have managed to bring individual-differences research closer to mainstream cognitive psychological research.

THE COGNITIVE COMPONENTS APPROACH

In the preceding two sections, dealing with the differential and information processing approaches to cognitive abilities, it has been demonstrated that the respective strengths and weaknesses of these approaches are complementary. Each approach being capable of providing useful, but incomplete explanations of human behaviour.

Given the complementary strengths and weaknesses of the two approaches, "it should be theoretically possible to develop an approach that would capitalise upon the strength of each approach, and thereby share the weakness of neither" (Sternberg, 1977, p.65). During the past few decades there have been calls for investigators to do just that (Cronbach, 1957; Messick, 1973), and in recent years several syntheses have been made (e.g., Day, 1973; Hunt, Frost and Lunneborg, 1973; Carroll, 1974). The componential approach described here, i.e., componential analysis continues and draws upon the syntheses developed by Day, Hunt, Simon, Carroll and others.

The Objectives of Componential Analysis

The overall objective of componential analysis is the identification of the component mental operations underlying a series of related information processing tasks and the identification of how these component operations are organised in terms of their relationships both to each other and to higher-order constellations of mental abilities.

Fundamental Unit of Analysis

The fundamental unit of analysis in componential analysis is the component. A component is an elementary information process that operates upon internal representations of objects or symbols (Newell and Simon, 1972). The component may translate a sensory input into a conceptual representation, transform one conceptual representation into another, or translate a conceptual representation into a motor output. Componential analysis is therefore of similar character to information processing analysis in that its elementary unit is a process rather than a construct representing a static source of individual differences.

Each component has three important properties associated with it that may be measured by mathematically estimated parameters: duration, difficulty, and probability of execution. In other words a given component consumes a certain amount of real time in its execution, has a certain probability of being executed correctly, and has a certain probability of being executed at all (Sternberg, 1980).

Components perform at least five kinds of functions (Sternberg, 1979, 1980): Metacomponents are higher-order control processes that are used for planning a course of action, for making decisions regarding alternative courses of action during reasoning or problem solving, and for monitoring the success of the chosen course of action. Performance components are processes that are used in the execution of a reasoning or problem-solving strategy. Acquisition components are processes used in learning how to reason or to solve problems. Retention components are processes used in retrieving previously stored knowledge, whether it be knowledge needed during reasoning or problem solving or knowledge regarding the reasoning or problem solving algorithm itself. Transfer components are processes used in generalisation, that is, in carrying over knowledge from one reasoning or problem-solving task to another.

The component is not necessarily, and usually will not be, the most elementary unit of behaviour that might be studied. Operations that are thought to be theoretically unimportant will be specified in the information processing model of task performance, but will not be identified as separate components. The rationale underlying this selectivity is that complex tasks may require a considerable number of operations, most of which are theoretically uninteresting. Componential analysis, like all of psychology, is reductionist in that it selects for examination only those aspects of behaviour regarded as

being theoretically important. These aspects of behaviour are understood through componential theories and models.

Componential Theories and Models

A componential theory comprises two parts: Identification of components involved in task performance; and specification of the combination rule for these components (Sternberg, 1977). Componential theories may differ from one another because they differ with regard to either the components theorised to enter performance, or the combination rule for components.

As a rule componential theories do not fully account for information processing behaviour under consideration. As a result, it is necessary to propose one or more componential models, each of which is consistent with the componential theory. A model further specifies both the order of component execution, and the mode of component execution.

Order of component execution refers to how components are sequenced. Mode refers to whether the execution of a given set of components is serial or in parallel, exhaustive or self-terminating, and independent or non-independent. In serial processing, components are executed sequentially; in parallel processing, they are executed simultaneously. In exhaustive processing, all possible executions of a given component or set of components are performed; in self-terminating processing, execution of components terminates before all possible executions have occurred. In independent processing, the execution of a given component has no effect upon whether any other component is executed; in dependent processing, execution of one component does affect whether one or more additional components are executed (Sternberg, 1980).

Componential models are expressed in information processing terms, usually as flow charts. The flow chart should provide a full description of the processes involved in task solution. The information processing model must be translated into some form that is suitable for empirical testing. Sternberg (1977) lists two conventional ways of doing this. The first is by simulating the model on a computer and then determining whether the computer found difficult (or easy) the same items that human subjects found difficult (or easy). A second test requires formulation of a mathematical model of performance, with parameters of the mathematical model corresponding to components of the information processing model. The model is then validated in terms of its fit to experimental data.

Table 3.1 provides a summary of the aspects of performance that distinguish among theories and models.

When all four aspects of performance (see Table 3.1) are specified, a more comprehensive understanding of the behaviour under consideration is facilitated.

Individual Differences

In componential analysis, individual differences may occur in either the components involved in task performance or the combination rule for components. If individual differences derive from either of these two sources, then one theory is inadequate for explaining the behaviour of all subjects.

At the model level, individual differences may occur in either the order of component execution or the mode of component execution. Again, if individual differences derive from either source, then one model is inadequate for explaining the behaviour of all subjects.

<i>Theory</i>	<i>Identification of Components</i>	<i>Specification of Components Combination Rule</i>
<i>Model</i>	<i>Stipulation of Component Execution Order</i>	<i>Stipulation of Component Execution Mode</i>

TABLE 3.1 ASPECTS OF THEORIES AND MODELS IN COMPONENTIAL ANALYSIS^a

(Sternberg, 1977, p. 67)

At the component level, individual differences may occur in component time or difficulty. In other words, some subjects may perform certain operations more quickly or more easily than other subjects. When individual differences are at the component level only, then one theory and one model are applicable for all subjects.

The five sources of individual differences are hierarchically organised into three levels: Theory, model, component, which are summarised in Table 3.2. Sternberg (1977) suggests that individual differences at higher levels affect interpretation of individual differences at lower levels, and offers several examples to help clarify how these effects operate:

At the Theory Level: Should one subject use components a, b and c and another use components c, d and e, then it will be possible to compare component orders, modes and values only for one component, the overlapping c. Alternatively, should both subjects use components a, b and c, but one subject uses an additive combination rule ($a + b + c$) and the other subject uses a multiplicative combination rule ($a \times b \times c$), then component values must be considered in light of the number of times each component is executed. The first subject, for example, executes component a once while the second subject executes it bc times.

At the Model Level: Should one subject execute components a, b and c in that (forward) order and another subject execute them in the reverse order, component values may be affected by the efficiency of one or the other strategy. Alternatively, should the difference be in mode of execution, such as in self terminating versus exhaustive search, faster execution of individual components in the exhaustive search may be more than offset by the greater number of times the components must be executed. Thus, although amount of time per

LEVEL	SOURCE	EXPLANATION	EXAMPLE
Theory	Components	Some subjects use more components, fewer components, or different components from those used by other subjects.	Subject 1 solves problem using components a,b,c; Subject 2 solves problem using components a,d,e.
	Combination Rule for Components	Some subjects combine components according to one rule, others according to a different rule.	Subject 1 combines components additively: a b c. Subject 2 combines components multiplicatively: a b c.
Model	Order of Component Processing	Some subjects order components in one sequence, others in a different sequence.	Subject 1 orders components a, then b, then c; subject 2 orders them c, then b, then a.
	Mode of Component Processing	Some subjects process particular components in one mode, others in another mode.	Subject 1 processes component a in self-terminating mode; Subject 2 processes component a exhaustively.
Component	Component time or power	Some subjects process particular components more quickly.	Subject 1 is quicker in processing component a than is Subject 2.

TABLE 3.2

SOURCES OF INDIVIDUAL DIFFERENCES IN COMPONENTIAL ANALYSIS

(Sternberg, 1977, p. 68)

component execution may be shorter, total amount of time spent on a given component may be longer (owing to repeated executions of that component).

In general then, the effects of individual differences in component processes higher in the hierarchy will affect interpretation of individual differences lower in the hierarchy. In fact, Sternberg (1977) claims that the reverse may also be true.

Structure of Componential Analysis

In its barest outline (shown in Table 3.3), a componential analysis comprises a series of intensive task analyses that, taken together, form the basis for an extensive task analysis.

The goal of an intensive task analysis is to gain a comprehensive understanding of the psychology of one particular task. In particular, it is designed to identify and understand the performance components specific to that task. The goal of an extensive task analysis is to integrate the findings of a series of interrelated intensive analyses. In particular, it is designed to demonstrate the psychological reality and generalisability of the identified components.

Each intensive analysis may be dichotomised into an internal and external validation. Internal validation consists of the attempt to reduce composite task performance into underlying components. External validation is the attempt to understand how the identified components of task performance generate individual differences in performance on external tasks. It is of importance to demonstrate through external validation that patterns of individual differences in components for the particular task under investigation are not specific to that task, but are instead of general interest.

Extensive Analysis	Intensive Analysis 1	Internal Validation	External Validation
	Intensive Analysis 2		
	⋮		
	Intensive Analysis N		

TABLE 3.3

COMPONENTIAL TASK ANALYSIS

(Sternberg, 1977, p. 71)

INTENSIVE TASK ANALYSIS

Composite Task or Test Score

In the situation where a number of tasks have been administered to a particular group of subjects, and a task score for each subject and an index of difficulty (such as proportion correct or response time) for each item type have been computed, but the primary concern is with only one of these tasks, from the componential analysis viewpoint, how would the data be analysed? Taking this viewpoint, variance across subjects as well as items will be of interest, but the final strategy will differ from that of both the differential or information processing approaches. The composite task score will be regarded as a measurement of the manifest trait of personality theory, which represents observable behaviour, however, primary interest is not so much in this behaviour as in the latent traits that underlie this behaviour. In order to reach the level of the latent traits, Sternberg (1977) points out that it is necessary to go through an intermediate level between the manifest and latent traits, which are measured by composite task scores and component scores respectively. This intermediate level is the level of the interval score.

Interval Scores

The first step in a componential analysis comprises the formation of interval scores from the decomposition of the global task into a series of subtasks, as was done by Johnson (1960) in his pioneering method of serial analysis. Each interval score is a score on one of the series of subtasks, and measures performance on a subset of information processing components required for completion of the total task. Each subtask in the series of subtasks requires successively less information processing, and

hence should involve reduced processing time and difficulty (Sternberg, 1978). The basic strategy described above is referred to as precueing (Sternberg, 1977).

For the purposes of the present study attention will be focused on the method of precueing as it has been applied in the presentation of verbal and geometric analogy problems of the form A:B::C:D (Sternberg, 1977). In order to decompose the task, it is necessary to eliminate from the subject's information processing successive terms of the analogy. Since the analogy comprises five terms, up to five subtasks can be formed, although, according to Sternberg (1978), there is no reason for splitting up the two answer options. Thus, the precueing method involves only four subtasks. In each case, presentation trials are dichotomised. In the first part, the experimenter presents the subject with some amount of precueing to facilitate solution of the analogy. In the second part, that of primary interest, the experimenter presents the full analogy. Solution of the analogy, however, is assumed to require merely a subset of the full set of components (i.e., to be a subtask of the full task), because the experimenter assumes that the subject utilised the precueing presented in the first part of the trial to reduce his/her processing load in the second part of the trial (Sternberg, 1977).

In the first subtask (which is identical to the full task), the subject is presented with a blank field (no precueing). The subject indicates when he/she is ready to proceed, and then the full analogy appears. The subject solves the analogy, and answers accordingly. In the second subtask, the subject receives a large subset of the task. The first part of the trial consists merely of precueing with the first term of the analogy. The subject then indicates that this term has been processed, and then the whole analogy is presented. The subject solves it, and

then indicates his/her response. It is important to note that although the full analogy was presented in the second part of the trial, only the last four terms needed to be processed, since the first term had been preprocessed during precueing. The third subtask involves a smaller subset of the total task. The first part of the trial involves the presentation of the first two terms of the analogy; the second part consists of full analogy presentation. The fourth subtask involves a very small subset of the full task in the second part of the trial. The first part of the trial consists of presentation of the first three terms of the analogy; the second part consists of full presentation, but requires processing of only the last two terms.

The general strategy, then, is to provide subjects with successively greater amounts of prior information (precueing) in the first part of the trial, which serves to reduce the amount of processing required in the second part of the trial. Subjects are afforded as much time as they require for both parts, to ensure that maximum information processing is encouraged both before termination of precueing and before response selection (Sternberg, 1977).

Types of Interval Scores: Two types of interval scores stem from the breakdown of the task, one type for the first part of the trial and one type for the second part of the trial. Both are interval scores in the sense that they represent processing in some time interval during task solution. The interval score for the first part of the trial is called a cue score because it is based upon performance in the precueing part of the trial. The cue score represents the time from the onset of precueing until the subject indicates his readiness for the second part of the trial. The interval score for the second part of the trial is called a solution score because it is based upon performance during that part of the trial requiring item

solution. The solution score represents the time from the onset of the second part of the trial (usually the full problem) until the subject indicates his solution.

Assumption of Additivity: Only one assumption is made in the breakdown of the composite task and the formation of interval scores, i.e., that of additivity: Interval scores reflecting lesser amounts of processing are "contained in" interval scores reflecting greater amounts of processing. "The interval scores may be viewed as a series of concentric circles, with the composite-task interval score encircling interval scores for all sub-tasks, and scores for each successive sub-task, and scores for each successive sub-task encircling all those that are based upon less processing" (Sternberg, 1977, p.74).

In order for the additivity assumption to be justified, the precueing procedure must not alter the nature of the task. The succession of sub-tasks must actually represent processing that occurs in successively later intervals of the full task.

It is important to note one assumption that is not made, i.e., that interval scores correspond to distinct stages of processing (Sternberg, 1969). It is possible for multiple stages to occur within one interval, or for a particular stage to occur during several intervals. In short, there are no assumptions made regarding stages.

Internal Validation : Simplex Model: The assumption that interval scores requiring less processing are contained in interval scores requiring more processing is empirically testable. Should the assumption be correct, then the succession of interval scores should form a simplex (Guttman, 1954). Internal validation of the interval scores is accomplished by testing this assumption. One very simple test is performed by inspecting the

intercorrelation table of solution scores with themselves. Should there be a general decrease in the magnitude of correlations away from the principal diagonal of the intercorrelation matrix, then the solution scores are probably related simplicially.

The scores of primary interest are the solution scores and these are formed into a simplex model, which comprises a set of equations in which each solution score is predicted by regression from the immediately adjacent solution scores. Such a model, according to Guttman (19*), can be used as a basis for further testing of simplicial structure.

In practice, the succession of interval scores will rarely show perfect additivity. Sternberg (1977) attributes this to two reasons: First, interval scores will generally be somewhat unreliable, and to the extent that measurement error in one interval score is uncorrelated with measurement error in another interval score, additivity will be lacking. Second, it is possible that the breakdown of the task into sub-tasks will at least slightly alter the nature of processing in one or more sub-tasks relative to the full task. Again, additivity will be affected. In general, however, the procedures of componential analysis can withstand additivity violations, although large violations will obviously compromise results (Sternberg, 1977).

External Validation: External validation of the interval scores begins by correlating the interval scores from reference ability tests. The objective here is to demonstrate both convergent and discriminant validity.

Convergent validity is demonstrated by showing that the interval scores correlate highly with well-established tests or other performance criteria that are hypothesised to measure the same thing as the interval scores.

Discriminant validity is demonstrated by showing that the interval scores correlate only trivially with tests measuring abilities hypothesised to be unrelated. If there are confoundings of particular concern, discriminant validity may be shown by holding constant statistically the scores on the confounded tests(s).

Component Scores

Formation: Component scores are parameter estimates for individual subjects. They are hypothesised to measure the durations or difficulties of the components of the information processing model of task solution. Sternberg (1977) lists three key ways in which these scores differ from composite task scores and interval scores:

First, the scores differ in terms of the level of measured behaviour. The composite task score is viewed as measuring the overt or manifest trait. The interval scores are all at intermediate levels. The component scores are viewed as measuring latent traits, the root processes underlying observable behaviour. Second, the scores differ in the purity of measured behaviour. The composite task score is a measurement of a mixture of various elementary information processes. Interval scores formed from the composite score are also themselves composites. However, successive interval scores may contain different numbers of elementary information processes. Each component score is hypothesised to be a measure of the duration or difficulty of a single elementary information process. It is a "pure" measure at the level of interest of the investigator. Third, the scores differ in their dependence upon a theory of measured behaviour. The composite task score may be theory-free, with no theory behind it all. It is for this reason that it has been possible to go into a factor analysis without a theory and to emerge from it with a theory: The computation of scores from which the input

correlation matrix is calculated may be theory free. The interval score is bound to a theory about the measurement process. The theory is that the succession of interval scores is additive, and is mathematically formulated in terms of a simplex model. The component score is bound to a theory about psychological process. The component scores will change if the theory (or specific model) changes. The theory is the component theory together with the specific model under that theory.

Internal Validation : Component Model. Internal validation of the component scores is by the component model. An estimate of goodness or badness of fit is calculated for each individual subject under each model. The best fitting model will then be used as the basis for computing component scores.

Once again it is worth mentioning that no assumptions are made about stages. In some models, certain component operations may be repeated after other component operations intervene, so if some kind of stage representation is used, it is advisable to allow for the possibility that a stage may be re-entered after it has been executed

External Validation: Procedures for external validation of the component scores are similar to those for external validation of the interval scores. The component scores are correlated with the reference ability scores and thereby assessed in terms of convergent and discriminant validation.

Reference Ability Scores

Formation: Reference ability scores are based upon conventional tests of mental ability that are selected so as to provide criteria for convergent and discriminant validation of the interval and component scores. When more

than one test of each ability is used, a factor score (derived from factor analysis) may be used to summarise the information from the battery of tests.

In componential analysis, reference abilities are defined as "constellations of components (latent traits) that in combination form stable patterns of individual differences across tasks" (Sternberg, 1977, p.78). It is important to note that in this definition, reference abilities are defined in terms of multiple tasks, individual differences, and constellations of components. These three properties suggest the role factors might play in componential analysis.

Componential view of factors: In componential analysis, factors are mathematical representations of reference abilities, just as parameters are mathematical representations of components. The limitations of factors that make them unsuitable as representations of latent traits (see the section The Decline of the Differential Approach earlier in this chapter) cease to be problems when factors are viewed as representations of reference abilities rather than latent traits. First, both factors and reference abilities are defined in terms of multiple tasks and individual differences. Their intrinsic properties are therefore congruent, and it is quite appropriate to use factor scores as measures of reference abilities. Second, both factors and reference abilities may be understood in terms of constellations of components. When they are defined in this way, non-uniqueness ceases to be a problem, because it is simply a manifestation of the infinite number of ways in which components can be combined to form constellations. A single component may enter into many factors and may be of differential importance in different factors.

Relationship between tasks and factors: A significant

outcome of linking task scores to manifest traits and factor scores to reference abilities is that both tasks and factors come to be seen as quite similar in one fundamental respect: Both represent constellations of components. Given this similarity they also differ in one fundamental respect: The components entering into tasks are held together by information processing constraints - a certain set of components is required to process the task; the components entering into factors are held together by individual differences constraints - certain sets of components form stable patterns of individual differences.

By their nature, both tasks and factors are arbitrary. The number of tasks that can be devised requiring different combinations of components is infinite, as is the fineness with which sources of individual differences can be identified. As new tasks requiring new combinations of components are devised, new sources of individual differences will be found, if only the distinctions are drawn finely enough. Sternberg (1977) maintains that for this reason differential psychology has seen a proliferation of factors.

Components are non-arbitrary, and their number is finite. Although they can be combined in an infinite number of ways, and thus appear in many guises, they remain the same. They are the latent traits that account for both task scores and factor scores.

Componential view of factor analysis: Sternberg (1977), identifies two major uses of factor analysis in componential analysis.

First, factor analysis is used in the attempt to recover the hypothesised reference abilities from the manifest trait data. There is no guarantee that when subjects are given a series of ability tests, the recovered factor

structure will correspond to the hypothesised categorisation of reference abilities. It is necessary to demonstrate that the observed pattern of individual differences corresponds to the predicted pattern. If it does not, the identities of the reference abilities may have to be reconceptualised.

Second, factor analysis is used to provide summary factor scores for each identified reference ability. It is always advisable to give at least two and preferably more tests for each ability in order to separate the common variance of interest in the tests from specific variance that is not a true part of the reference ability. Some a priori weighting of tests could be used to form a composite estimate of performance on the reference ability. The advantage of using a factor score is that it is the statistically optimum estimate of performance on the factor and therefore of performance on the reference ability. Factor scores thus have the dual advantage of being convenient summary scores and at the same time, optimum estimates of reference ability performance.

Sternberg (1977) argues that the role played by factor analysis in componential analysis is much less ambitious than its role in traditional differential psychology. In componential analysis it is used to identify stable patterns of individual differences and to quantify these patterns by assigning one or more numbers to each subject. When factor analysis is used for these more modest purposes, the problems associated with its use (highlighted in the section The Decline of the Differential Approach earlier in this chapter) disappear because such problems occur only in the conventional use of factor analysis to uncover the latent structure of human abilities.

Reference Ability Models: Reference abilities are defined in terms of constellations of components. Given

their definition, it should be possible to dissect these abilities into their underlying components. This dissection is attempted through reference ability models.

In reference ability modeling, the objective is to account for individual differences in subject's reference ability scores in terms of individual differences in their component scores. Full understanding of a given reference ability would be attained when all the systematic between-subjects variance in the reference ability scores was accounted for by a combination of components.

Integration : Structural Regression Model

Thus far, three different model types have been described. The simplex model (for solution scores), the component model, and the reference ability models. The outcomes of these three different model types can be integrated through a structural regression model (Sternberg, 1977). This model gives an overall account of the interrelationships of individual differences in the various types of scores. The design of a structural regression model (as applied to componential analysis) is shown in Figure 3.1.

A structural regression model is a causal account of interrelationships. In componential analysis, the causal elements are the components. Parameters representing hypothetical component durations or difficulties are shown in the third row of the model in the figure. These components combine to form two types of entities. The first is embodied in the sub-tasks formed by breaking down the total task. These sub-tasks are shown in the second row of the model. Arrows are shown leading upwards from the parameters to the sub-tasks. Not all parameters are shown for all sub-tasks.

LEVEL

FACTOR

SUBTASKS

PARAMETERS

FACTOR

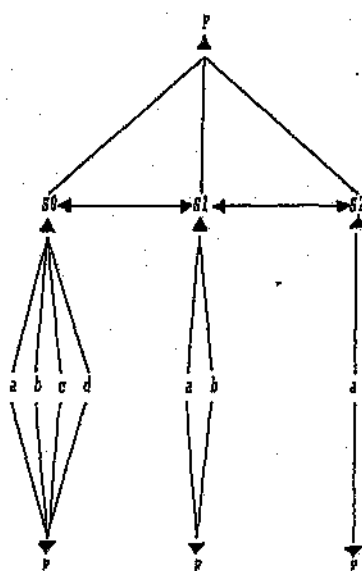


FIGURE 3.1 DESIGN OF A STRUCTURAL REGRESSION MODEL

(Sternberg, 1977 p. 81)

Parameters are shown only if the corresponding components enter into performance on the sub-task. The second type of entity is the reference ability. A hypothetical factor is shown in both the first and fourth rows of the model. In Figure 3.1, arrows are shown leading downward from the parameters to the factor representing the reference ability.

The relationship of the reference ability to the components is mediated by the use of sub-tasks. Access from the components to the reference abilities is possible via the sub-tasks that form the basis for componential modeling. Operationally, therefore, sub-tasks form a mediate link, and reference ability scores can be accounted for in terms of solution scores, which are themselves accounted for by the component scores. Thus, arrows are drawn from the sub-tasks to the factor, although it is understood that these sub-tasks represent only mediating operational links in the chain. From a theoretical point of view, the solution scores can be dispensed with once the component scores are estimated. In practice however, Sternberg (1977) points out that they are very useful even after the component scores have been estimated.

The structural regression model summarises in compact form a large quantity of data from the intensive analysis. It may be seen as forming the basis for construct validation of the system of models and of the task used to validate these models. Sternberg (1977) suggests that high construct validity is indicated by a high level of prediction for: Interval scores from component scores; reference ability scores from component scores; reference ability scores from interval scores and interval scores from each other.

OVERVIEW

The structure of the entire intensive analysis is shown in Figure 3.2. The composite task score, measuring the manifest trait, is broken up into interval scores. Internal validation of the interval scores is through the simplex model. External validation is through correlations with scores on reference abilities tests and through the reference abilities models. In the correlations, the aim is to show both convergent and discriminant validity. The interval scores are used as a basis for computing component scores, which are hypothesised to measure latent traits. Internal validation of the component scores is through the component model. External validation is through correlations with the reference abilities and through the reference abilities models. Again, the aim is to show both convergent and discriminant validity. The various types of models are combined into a structural regression model, which forms the basis for intensive construct validation of the system of models and of the task used to validate the models.

Extensive Task Analysis

An extensive analysis comprises a series of related intensive analyses. The prime objective thereof is to demonstrate through a sequence of converging operations (Garner, Hake and Eriksen, 1956) that the components identified in one particular intensive analysis have psychological reality, and that they are generalisable to a variety of related tasks.

Sternberg (1977) maintains that information processing psychology has been caught in a web of specificity, the reason for this he believes is that its very nature lends itself to specific theories. Little attempt has been made to demonstrate that the parameters that are applicable to

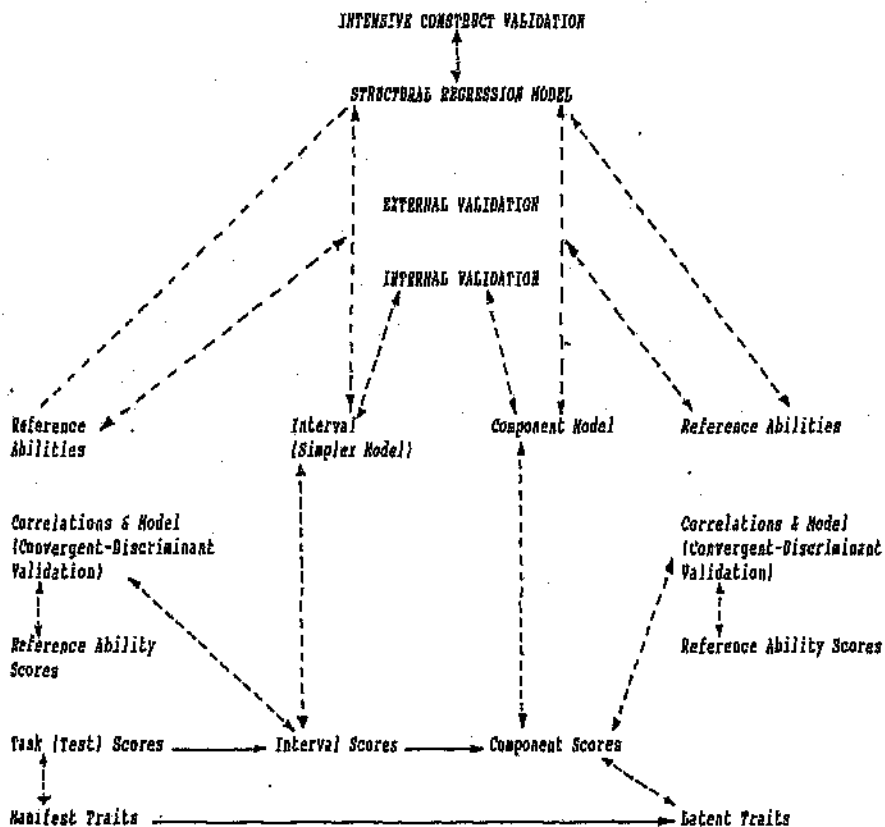


FIGURE 3.2

COMPONENTIAL THREE-LEVEL INTENSIVE TASK ANALYSIS

(Starnberg, 1977, p. 82)

one task are applicable to related tasks as well. In general the tendency has been toward the identification of very specific parameters for even more specific tasks, with no attempt to show that the hypothesised parameters are of any relevance or interest beyond a given task or even task material.

The extensive analysis strategy selected will depend upon the task domain under study, and so it is necessary to be somewhat specific in showing the course an extensive analysis might take. Table 3.4 provides a summary of the extensive analysis strategy used for the series of analogy studies conducted by Sternberg (1977).

In the extensive analysis for the analogies studies, five task variables were systematically varied: Dependent variable, response format, type of content, difficulty and definition of attributes. Three intensive analyses used response time as the primary dependent variable, while one used proportion correct. In two studies, the response format was true-false, in two, it was forced choice. In two experiments, the content of the analogies was verbal; in two, it was nonverbal. Items across studies ranged from very easy to very difficult. In one study, stimulus attributes were very well defined. Sternberg (1977) argues that if the same theory can be found to apply across all these different task manipulations, then it can be concluded that the theory is a general one.

Advantages of Componential Analysis

From the preceding discussion on the componential approach to cognitive activities, it is apparent that componential analysis has a number of advantages over both the psychometric techniques of differential psychology and the modeling techniques of information processing psychology. Table 3.5 presents a summary of seventeen of these

Type of Analogy	Dependent Variable	Response Format	Type of Content	Difficulty
People Piece Analogies	Response Time	True - False	Non - Verbal	Very Easy
Verbal Analogies	Response Time	True - False	Verbal	Fairly Easy
Geometric Analogies	Response Time	Forced - Choice	Non - Verbal	Fairly Difficult
Animal Name Analogies	Proportion Correct	Forced - Choice	Verbal	Very Difficult

TABLE 3.4 OVERVIEW OF CONVERGING OPERATIONS IN ANALOGICAL REASONING EXTENSIVE TASK ANALYSIS : EXTENSIVE CONSTRUCT VALIDATION

(Sternberg, 1977, p. 24)

INTERVAL SCORES	COMPONENT SCORES	REFERENCE ABILITY SCORES
1. Separate components which otherwise would have been confounded.	Estimate scores by inferentially powerful model controlled by investigator.	Identify correlates of individual differences in component performance.
2. Compare models which otherwise would have been indistinguishable.	Represent scores uniquely within parameter space - no 'rotation dilemma'.	Prevent overvaluation of components with task-specific individual differences.
3. Increase degrees of freedom for residual in prediction.	Interpret scores in terms of mental process.	Provide a common language across different tasks and researchers.
4. Specify precisely temporal orderings and location of components.	Provide scores based upon a well-specified theory and model derived from individual behaviour.	
5. Prevent distortion of results from external validation.	Permit systematic study of individual differences at theory, model and component levels.	
6. Provide component-free estimates of performance for nested processing intervals.	Pinpoint individual sources of particular strength or weakness for diagnosis and training.	
7.	Fit into a systematic framework for understanding intellectual development.	
8.	Derive estimates of measurement error from data for individual subjects.	

TABLE 3.5

ADVANTAGES OF INTERVAL, COMPONENT AND REFERENCE ABILITY SCORES

(Sternberg, 1977, p. 85)

advantages, which result from: The breakdown of the composite task into sub-tasks and the formation of interval scores; the formation of a component model and component scores, and the use of reference abilities and reference ability scores.

Problems with Componential Analysis

Pellegrino and Lyon (1979) have provided a comprehensive and perceptive analysis of componential analysis. Although the main focus of their analysis is on Sternberg's pioneering work in this domain of cognitive research, it takes into account more recent work as well, and thus serves as an evaluation of the analytic system as a whole. Sternberg (1980) subsequently responded to this critique. His response is organised along more or less the same lines as the original review.

Potential Methodological Problems

Parameter Confoundings: Pellegrino and Lyon (1979) note that parameters estimated from a componential analysis may be confounded - that is, contain embedded within them latencies or difficulties assignable to more than one component process. In componential analysis, each parameter is estimated on the basis of the manipulation of a single independent variable. Therefore, the possibility exists that more than one psychological process could be affected by the manipulation of a single variable, such that the latencies or difficulties of two or more processes may be estimated as a single parameter. More generally, the manipulation of a single independent variable may equally affect multiple psychological processes. Sternberg (1980) acknowledges the applicability of this criticism to componential analysis, however, he points out that it also applies equally to all psychological experimentation that ever has or will be done. It is never known for sure that

the psychological process (or state) assumed to be affected by a variable manipulation is the only process (or state) affected by that manipulation. In fact it is never known for certain whether a given manipulation affects the intended psychological process (or state) rather than some other one(s).

Given that no methodology is immune from this criticism, the most that can be expected from a methodology is that it: Minimises confoundings of psychological processes (or states), and maximises the probability that such confoundings will be discovered if they indeed exist. Sternberg (1980) argues that componential analysis approaches (although it does not reach) both of these goals.

Methods of componential task decomposition, in particular the method of precueing, permit separation of components that would be confounded in standard information processing paradigms. Although the methodology goes beyond certain standard paradigms in preventing confoundings, it is not immune from such confoundings - hence the necessity for discovering them when they exist. It is stated by Sternberg (1980) that external validation of the sort used in componential analysis often facilitates the discovery of such confoundings when they exist, and, that in fact, Pellegrino and Lyon (1979) rely upon the results of external validations presented in his (Sternberg's) 1977 research, to infer that confoundings exist.

Confoundings are indicated by either: Patterns of correlations between parameters and external criteria, or patterns of correlations whereby a single parameter is significantly correlated with two or more reference abilities that are themselves uncorrelated. It is important that a method be sensitive to confoundings, in part because the discovery of confoundings often serves the

heuristic function of suggesting future directions for research.

In summary, Sternberg (1980) maintains that componential analysis is less susceptible to confoundings than most other information processing paradigms, and it is better able than they are to discover and ultimately to correct such confoundings when they do exist.

Evaluation of Model Fits: Pellegrino and Lyon (1979) argue that, taken by itself, percentage of variance accounted for is an inadequate criterion for assessing model fits. They suggest that componential analysis relies very heavily on this criterion for assessing fit of componential models. They also point out, however, that this is not the sole criterion for assessing model fits in componential analysis. Sternberg (1980) believes that Pellegrino and Lyon overstate the importance of this criterion in componential analysis. In fact, he claims that the tests to which models have been subjected in componential analyses of human abilities are more stringent and more diverse than the tests that have been applied in many other kinds of abilities research.

Reliability of single-subject parameter values: Pellegrino and Lyon (1979) question the reliability of individual parameter estimates in componential analysis. Indeed, the reliabilities of the various parameters vary greatly, and these reliabilities ought to be taken into account in interpreting the estimates for individuals. In fact, Sternberg (1979) cautions against the use of individual parameter estimates from componential analysis to make real-life decisions about individuals. Nevertheless, the reliabilities of some parameters appear to be high enough to yield rather striking and replicable correlations with external criteria. For example, in the analogies tasks conducted by Sternberg (1977), comparison

parameters were found to have lower correlations with measures of higher-order reasoning than did the response constant parameter. This result has now been replicated with geometric analogies (Mulholland, Pellegrino, and Glaser; 1980), as well as with several tasks other than analogies (a.g., Egan, 1976). Thus, Sternberg (1980) points out that although reliabilities may not be high enough for practical individual prediction and diagnosis, they do seem to be high enough to yield stable patterns of individual differences, as shown by the consistent patterns of correlations obtained across experiments and even across experimenters.

Having presented a comprehensive overview of the cognitive components approach to investigating mental abilities, for present research purposes attention shall not be focused on Sternberg's (1977) componential theory of analogical reasoning.

STERNBERG'S COMPONENTIAL THEORY OF ANALOGICAL REASONING

Sternberg (1977) undertook an empirical investigation of the performance components involved in analogy solution, employing the cognitive components approach described in the preceding section of this chapter. This research culminated in the formulation of a componential theory of analogical reasoning.

The remainder of this chapter shall comprise an in-depth discussion and evaluation of Sternberg's (1977) componential theory.

Sternberg (1977) lists five criteria, which he believes a theory of analogical reasoning should satisfy:

"Completeness. A complete theory is one that accounts for all processes involved in analogical reasoning,

from beginning to end. A complete theory explains analogical reasoning from the time the analogy is first perceived to the time a response is emitted.

Specificity. A specific theory describes in detail the workings of each process involved in solving analogies. A theory can be complete but not specific if it accounts for all processes, but does not describe the workings of the processes in detail. A theory can also be specific, but not complete if it describes in detail a subset of the processes used to solve analogies.

Generality. A theory is general if it is applicable across a wide range of analogy problems. A fully general theory would be applicable to all analogy problems, regardless of the content or mode of presentation. It is not obvious a priori that a fully general theory is possible; indeed most previous theories are not very general.

Parsimony. A theory is parsimonious if it can account for analogical reasoning with a relatively small number of parameters and working assumptions. Parsimony is difficult to evaluate, in part, because theories that appear parsimonious on their surface have hidden assumptions, while other theories that appear less parsimonious can be taken more easily at face value. As might be expected, there is often a trade off between parsimony on the one hand and completeness and specificity on the other. A difficult problem facing theorists is to strike a reasonable balance between them.

Plausibility. A theory is plausible if it is able to account for experimental (or other) data that provide a test of the theory. Plausibility also provides

intuitive judgements about the reasonableness of the theory. If one theory seems less reasonable on its face than another theory, sceptics may require more compelling evidence to convince them of the former theory than to convince them of the latter" (Sternberg, 1977, pp.100-101).

NATURE OF ANALOGY

An analogy is a hierarchy of relations taking the form A is to B as C is to D ($A:D::C:D$). An analogy exists when there is a higher-order relation of equivalence or near-equivalence between two lower-order relations. In the terminology used by Sternberg (1977), an analogy exists when there is a rule Y, that maps a domain rule, X, into a range rule, Z.

The domain refers to the A and B terms of the analogy. It is in the domain of the analogy that the first lower-order relation X is discovered. The range refers to the C and D terms of the analogy. It is to the range of the analogy that the X relationship is mapped through the formation of an isomorphic relation Z. The essence of analogy, according to Sternberg (1977), is in the higher-order relation, Y, that maps the domain into the range of the analogy. This higher-order mapping relation is absent in many other theories of analogical reasoning, i.e., differential theories (e.g., Spearman, 1927; Thurstone, 1938; Cattell, 1971) and information processing theories (e.g., Johnson, 1962; Shalom and Schlesinger, 1972; Rumelhart and Abrahamson, 1973).

THE COMPONENTIAL THEORY

Overview

The componential theory of analogical reasoning proposed by

Sternberg (1977) contains six information processing components, five of them mandatory and one of them optional. The components are of three general types: Attribute identification, attribute comparison, and control. The combination rule for the components is one of additivity. The structure of the theory is summarised in Table 3.6, and the functions of the five mandatory components are shown schematically in Figure 3.3 for a true-false analogy.

Components

Attribute Identification. There is one attribute-identification component, encoding. This component involves the translation of the stimulus into an internal representation upon which further mental operations can be performed. The internal representation is stored in working memory and is assumed to consist of an attribute-value list. Encoding involves the comprehension and storage in memory, of each of the four analogy terms.

Attribute Comparison. According to the theory there are three mandatory attribute-comparison components and one optional one:

"Inference is the process by which a rule, X is discovered that relates the A term of the analogy to the B term. Inference thus occurs in the domain of the analogy, and the outcome is stored in working memory.

Mapping is the process by which a higher-order rule, Y, is discovered that maps the domain of the analogy into the range. Mapping requires discovery of a rule that relates A (the first term of the domain) to C (the first term of the range). The rule is stored in working memory.

COMPONENT PROCESS	LATENCY PARAMETER	DESCRIPTION
I Components of Analogical Reasoning		
A. ATTRIBUTE IDENTIFICATION		
Encoding		Identify attributes and values for each term of the problem.
B. ATTRIBUTE COMPARISON		
Inference		Discover rule relation A to B. Domain.
Mapping		Discover rule relating A to C. Domain to range.
Application		Generate rule to form D' and evaluate D. Range
Justification		Optional. Test validity of operations as performed. Domain. Range.
C. CONTROL		
Preparation		Prepare for analogy solution.
Response		Monitor solution process. Translate solution into response.
II Combination Rule		
Additive		Solution time = Encoding time + inference time + mapping time + no application time + (justification time) + preparation-response time. Solution difficulty = Encoding difficulty + inference difficulty + mapping difficulty + application difficulty + (Justification difficulty) + preparation-response difficulty.

TABLE 3.8 COMPONENTIAL THEORY OF ANALOGICAL REASONING

(Sternberg, 1977, p. 135)

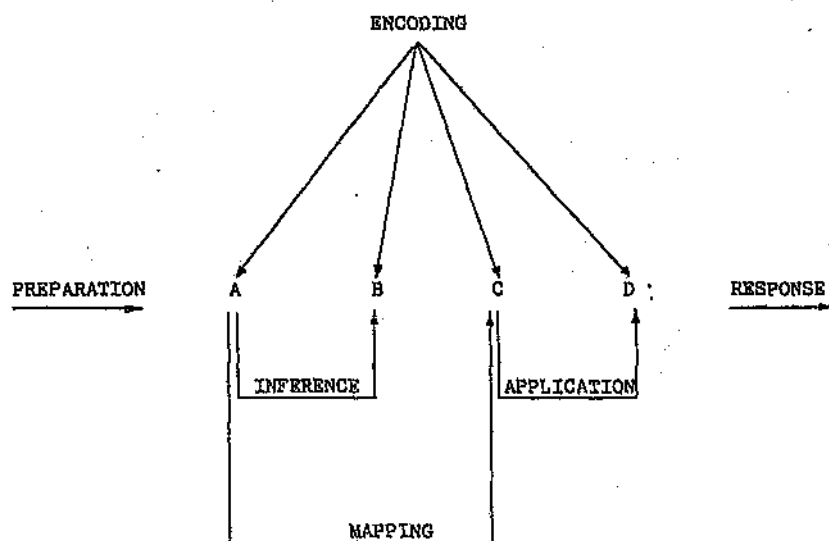


FIGURE 3.3 SCHEMATIC REPRESENTATION OF ANALOGICAL REASONING THEORY

(Sternberg 1977, p. 136)

Application is the process by which a rule, Z, is generated that forms D' (an image of the correct answer) and evaluates D. Application thus occurs in the range of the analogy, and the outcome is stored in working memory.

Justification, an optional component, is the process by which one of several answer options that are non-identical to D' is justified as closest to D'. The process is required only in forced-choice analogies in which none of the presented answers conforms to the visualised answer D'' (Sternberg, 1977, p.136).

Control. The theory includes one control component. This component includes the process by which subjects prepare for solving the analogy, monitor the solution process, and translate the solution into a response. The component, preparation-response, incorporates those operations that were not considered worthy of separate components, but were thought to be suitably represented in combination.

Combination Rule

It is hypothesised by Sternberg (1977) that response time is equal to the sum of the amounts of time spent on each component operation. Hence, a simple linear model predicts response time as the sum across the different component operations of the number of times each component operation is performed (as an independent variable) multiplied by the duration of that component operation (as an estimated parameter). Corresponding to each component operation, then, is a parameter representing its duration.

It is further hypothesised that proportion of response errors (item difficulty) is equal to the (appropriately scaled) sum of the difficulties encountered in executing

each component operation. A simple linear model predicts proportion of errors as the sum across the different component operations of the number of times each component operation is performed (as an independent variable) multiplied by the difficulty of that component operation (as an estimated parameter). This additive combination rule is based on the assumption that each subject has a limit on processing capacity (or space, Osherson, 1974). Each execution of an operation uses up capacity. Sternberg (1977) maintains that until the limit is exceeded, performance is flawless except for constant sources of error (such as motor confusion, carelessness, momentary distraction), but goes on to emphasise that once the limit is exceeded, performance is at a chance level.

In response-time models (with solution latency as a dependent variable), all component operations must contribute significantly to solution latency, since by definition each execution of an operation consumes some amount of time. However, in response-error models (with error rate as a dependent variable), all component operations need not contribute significantly to proportion of errors because, according to Sternberg (1977), some operations may be so easy that regardless of how many times they are executed their contribution to prediction of errors will only be trivial.

COMPONENTIAL MODELS OF ANALOGICAL REASONING

Four Basic Models

Overview

Sternberg (1977) proposes four alternative models of analogical reasoning processes, which share the same information processing components and combination rule, but differ in the way the components are organised into a strategy for solving analogy problems.

Encoding

Washington. ((President (first)), (Portrait on currency (Dollar))
(War Hero (Revolutionary)))

1. ((Counting number (one)), (Ordinal Position (first)),
(Amount (one Unit))).

Lincoln ((President (Sixteenth)), (Portrait on Currency (Five
Dollars)), (War Hero (Civil))).

10. ((Counting number (ten)), (Ordinal position (tenth)),
(Amount (ten Units))).
5. (Counting number (five)), (Ordinal position (fifth)),
(Amount (five Units))).

Inference

Washington-1. ((President (Ordinal position (first))),
(Portrait on currency (Amount (Dollar))), (0))

Mapping

Washington → Lincoln ((President (First, Sixteenth)), (Portraits
on currency (Dollar, Five Dollars)), (War heroes
(Revolutionary, Civil))).

Application

Lincoln → 10. ((0), (0), (0))

Lincoln → 5. ((0), (Portrait on Currency (Amount (Five
Dollars))), (0))

TABLE 3.7 ATTRIBUTE - VALUE LIST FOR ANALOGY EXAMPLE
WASHINGTON : 1:: LINCOLN: (a. 10, b. 5)

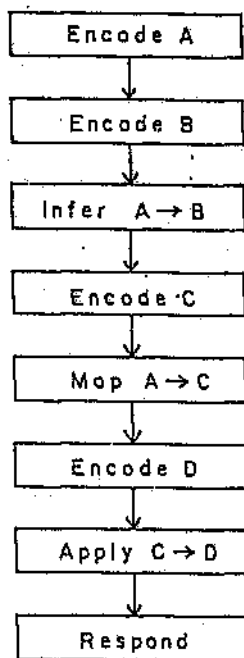


FIGURE 3.4 SCHEMATIC FLOW CHART FOR MODEL 1

(Sternberg, 1977, p. 138)

Although the models are discussed in terms of an attribute-value list representation for information, Sternberg (1977) points out that they could be mapped into other representations as well. He uses an example analogy, Washington : 1 :: Lincoln: (a.10,b.5), to illustrate the models. The results of operations upon this analogy are summarised in Table 3.7 and are described by Sternberg (1977) as follows:

Model I. A schematic flow chart for Model I is presented in Figure 3.4. Names of operations appear inside the boxes of the flow chart.

The subject begins analogy solution by encoding the first analogy term, Washington, and then the second analogy term, 1. In the example, the subject stores in working memory the facts that Washington was the first President, a portrait on a Dollar bill, and a revolutionary war hero. For the numeral 1, possible relevant facts are that it can represent the first counting number, the first ordinal position, or a unitary amount.

Next, the subject infers the relation between the first two analogy terms. In the example, two of the three attributes encoded for Washington can be used to relate that analogy term to the second term, 1. The inferred relation states that Washington was the first president and is the portrait on a Dollar bill. No relation to 1 could be found for the third attribute-value pair identifying Washington as a revolutionary war hero, and so that attribute-value pair is assigned a null value, 0, in the inference relation.

The subject then encodes the third analogy term, Lincoln, thus facilitating the mapping of the relation between the first and third terms. In the example, the subject

recognises that both Washington and Lincoln (a) were presidents, (b) are portraits on currency, and (c) were war heroes.

The final required encodings are of the answer options, 10 and 5. Once these terms are encoded, the subject attempts to apply from Lincoln to each answer option a relation analogous to the one previously inferred. An attempt to construct an analogous relation from Lincoln to 10 results in a null relation, because no analogy can be found: Lincoln was neither the tenth president nor is he the portrait on a 10-Dollar bill. An analogous relation can be constructed from Lincoln to 5, however, because Lincoln's portrait appears on a 5-Dollar bill.

The analogy is now solved. All that remains is for the subject to communicate the solution. The subject responds with option b, completing the analogy problem.

It is important to note that in Model 1, inference, mapping, and application (the attribute-comparison operations) are exhaustive processes. All attributes are compared to each type of operation. However, Sternberg (1977) emphasises that this characteristic fails to hold true for the subsequent models to be described.

Model II. The schematic flow chart for Model II, shown in Figure 3.5, is identical to that for Model I through the encoding of D (including D_1 and D_2).

At this point, Sternberg (1977) claims that an "attribute counter" is set equal to zero and then incremented by one. Only one attribute is applied. This attribute, that Lincoln was the sixteenth president fails to distinguish between answer options. As a result, the subject returns to the attribute counter, and again increments it by one. The subject now applies the second attribute. In doing so,

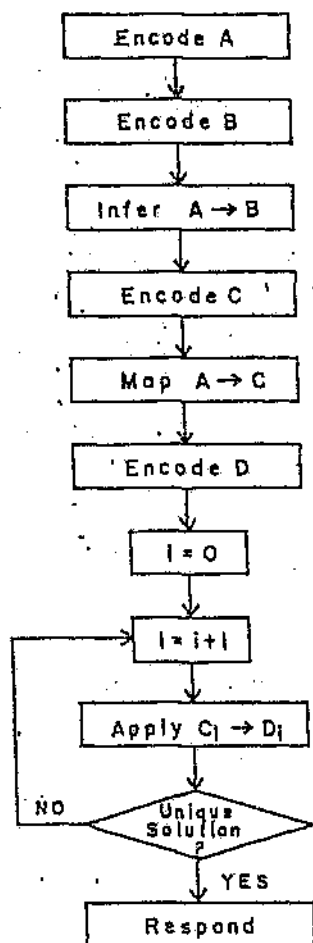


FIGURE 3.5 SCHEMATIC FLOW CHART FOR MODEL II

(Sternberg, 1977, p. 140)

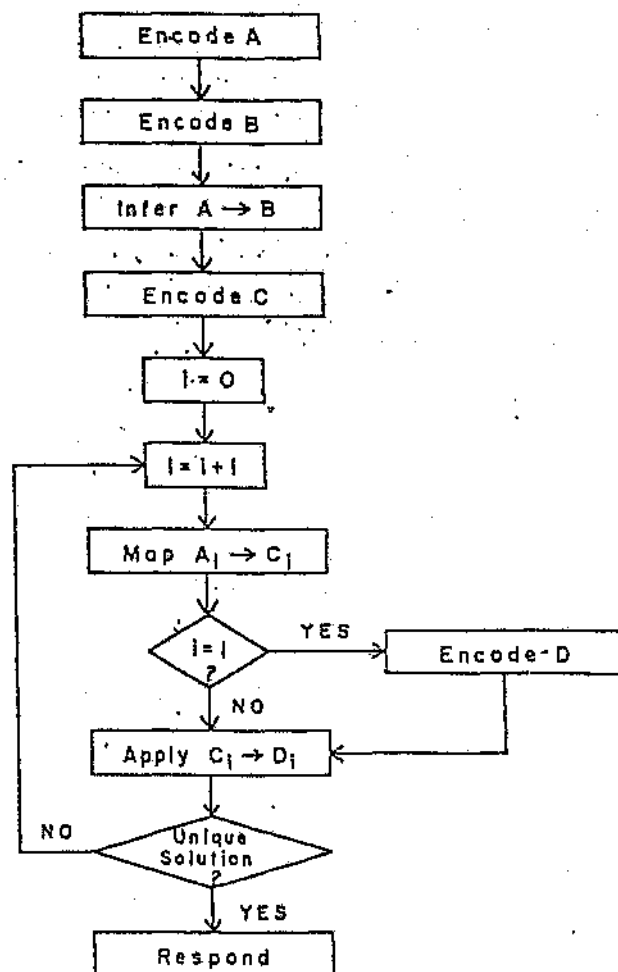


FIGURE 3.6 SCHEMATIC FLOW CHART FOR MODEL III

(Sternberg, 1977, p. 141)

he/she recognises that 5 refers to the amount of the bill on which Lincoln appears. Since the second option leads to an analogous rule, the subject responds b.

It is worth noting that in Model II, inference and mapping are exhaustive, but application is self-terminating. Whereas, in Model I, all attributes were applied, in Model II, only as many attributes are applied as are needed to reach a unique solution, i.e., the subject never reaches the third attribute-value pair identifying Lincoln as a civil war hero, because the second attribute permitted a unique solution.

Model III. The schematic flow chart for Model III, shown in Figure 3.6 is similar to that for Model II through the encoding of the first attribute.

At this point, the subject determines that the subject enters the attribute-test loop. In Model II includes only application. He maps one attribute, recognising that Washington and Lincoln were both presidents, the first and sixteenth respectively. Since $i = 1$ (the subject is travelling through the loop for the first time), the subject encodes D. However it is purported that the subject will not re-encode D should subsequent iteration through the loop be necessary, since, in subsequent iterations, $i = 1$. Next the subject applies the first attribute, that Lincoln was the sixteenth president. This application fails to yield a unique solution, so the subject returns to the beginning of the loop. As in Model II, the second trip through the loop does yield a unique answer.

It is important to note that in Model III, inference is exhaustive, but mapping and application are self-terminating. Whereas Model II required mapping of all three attributes to solve the analogy. Model III requires mapping of only two attributes.

Model IV. The schematic flow chart for Model IV, shown in Figure 3.7, is the same as that for Model III through encoding of the second analogy term.

At this point, the subject sets $i = 0$, and then enters the attribute-testing loop. Once again, the loop is arranged so that the analogy terms are encoded only on the first trip through the loop. The subject infers, maps, and then applies an attribute, trying to select a unique answer. If the subject cannot, he/she returns to the beginning of the loop, and infers, maps, and applies another attribute, continuing the iterative process until the analogy is solved.

In Model IV it is significant to note that inference, mapping, and application are all self-terminating. Whereas Model III required inference of all three attributes to solve the analogy, Model IV requires only two inferences.

TWO BASIC STRATEGIES IN APPLICATION

Sternberg (1977) maintains that in applying the analogous rule from C to D, two basic strategies may be followed, the essential aspects of which are presented schematically in Figure 3.8.

Sequential Option Scanning. The first of these strategies involves sequential scanning of the presented answer options. The subject applies attributes onto one option before applying any attributes onto another option. The assumption is made that as a safeguard against premature decisions, the subject checks all options.

Considering the example in Table 3.7, it is apparent that the subject initiates application by applying the first attribute-value pair for Lincoln, (President (Sixteenth)),

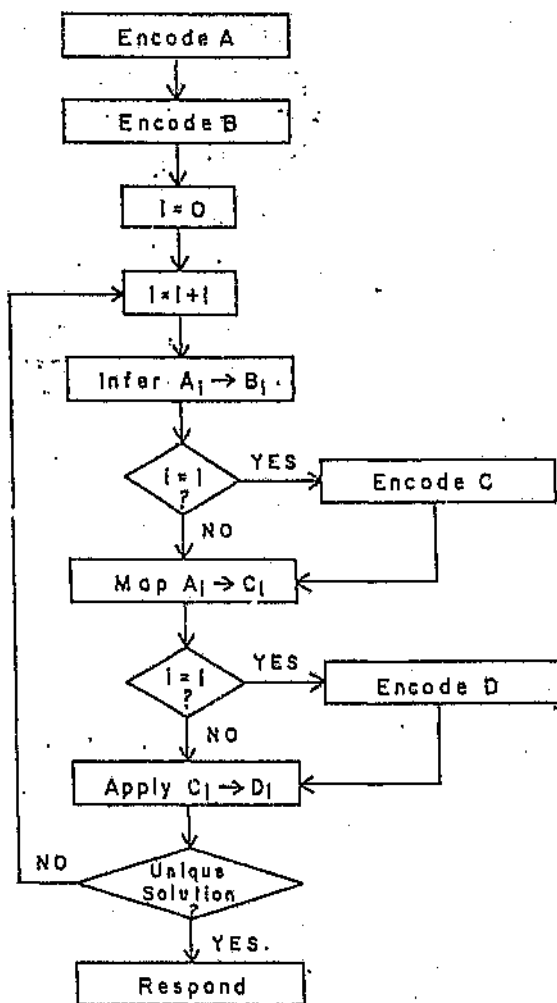


FIGURE 3.7 SCHEMATIC FLOW CHART FOR MODEL IV

(Sternberg, 1977, p. 142)

onto 10. When this fails to yield a match, as do subsequent applications onto 10, the subject applies each of the attribute-value pairs onto 5, finding a match for (portrait on currency (Five Dollars)).

Alternating Option Scanning. The second strategy involves alternating scanning of presented answer options. Here the subject applies an attribute onto each option before applying any further attributes. Again, the subject is assumed to check all options.

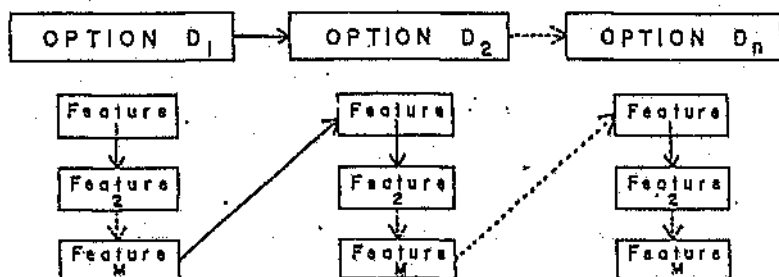
Focusing again on the example in Table 3.7, it is apparent that the subject initiates application by applying (President (Sixteenth)) first onto 10, then onto 5. When this fails to yield a match, the subject applies (portrait on currency (Five Dollars)), onto 10, and then 5, only then obtaining a match. It is of importance to note that whereas sequential option scanning required application of all attributes onto the first answer option, alternating option scanning required application of only two attributes onto this answer option.

SCOPE OF THE MODELS

Analogy Format. The models are applicable to analogies presented in either forced-choice or true-false format. Application for forced-choice analogies proceeds as described in the previous section, and may be extended to analogies with more than two answer options. In true-false analogies, Sternberg (1977) points out that the subject proceeds through the analogy and tries to determine whether a relation can be constructed from C to D that is analogous to the relation from A to B.

If such a relation can be constructed, the analogy is true; otherwise, it is false. The distinction between sequential

SEQUENTIAL OPTION SCANNING



ALTERNATING OPTION SCANNING

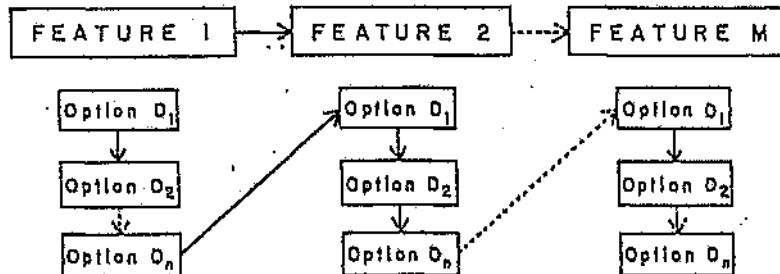


FIGURE 3.8 SCHEMATIC REPRESENTATIONS OF TWO TYPES OF FORCED CHOICE MODELS.

(Sternberg, 1977, p. 144)

and alternating option scanning becomes irrelevant, because there is only one answer option.

Analogy Content. Sternberg (1977) maintains that just as the models are applicable to verbal analogies of the type presented in Table 3.7, they are also applicable to geometric, animal name, and Miller analogies, using the same encoding schema.

Analogy Processes. While the models proposed by Sternberg (1977) specify in some detail alternative ways in which attribute information can be combined to arrive at a solution for analogy problems, they do not specify what the possible attributes are for different types of analogies, nor do they specify how subjects discover these attributes in the first place. A complete theory of analogical reasoning would have to specify this further information, something that Sternberg (1977) has failed to do.

EVALUATION OF STERNBERG'S COMPONENTIAL THEORY

Sternberg (1977) provides a short evaluation of the proposed componential theory in terms of the criteria he believes a theory of analogical reasoning should satisfy:

Completeness. It is believed that to a large extent the theory satisfies this criterion: The analogical reasoning process is described by detailed models from beginning to end. Each model falling under the theory can be represented fully in flow chart form, with all necessary processes explicitly stated and shown in their relations to each other.

Specificity. The theory is viewed as quite specific in describing the details of the three mandatory attribute-comparison operations, but less specific in

describing the details of the encoding process. The theory describes numerous minute monitoring operations, although "the internal mechanisms, behind 'preparation' and 'response' remain a mystery" (Sternberg, 1977, p.146).

Generality. The theory is regarded as very general, being applicable to a wide variety of analogy types: People piece analogies, verbal analogies, geometric analogies, and animal name analogies.

Parsimony. The theory is seen as achieving a considerable degree of parsimony by specifying all operations, but assigning separate information processing components only to psychologically significant operations. The theory thus manages to maintain completeness while at the same time retaining parsimony. The major aspects of analogical reasoning are accounted for in five mandatory and one optional component.

Plausibility. The results of a series of experiments, devoted to tests of the theory and models, conducted by Sternberg (1977), revealed that the theory accounted for the data very well, thereby satisfying the criterion of plausibility.

In conclusion, Sternberg states, a it tentatively, that "On balance, the componential theory does quite well by the five criteria" (1977, p.147).

FURTHER COMMENTS ON STERNBERG'S THEORY OF ANALOGY

Similar to their wide-ranging and perceptual analysis of componential analysis, Pellegrino and Lyon (1979) provided an evaluation of Sternberg's (1977) componential theory of

analogical reasoning. As was the case with their critique of componential analysis Sternberg (1980) has responded to their comments.

Difficulty of Items. Pellegrino and Lyon (1979) note that the analogies studied in the experiments reported by Sternberg (1977) were rather easy and suggest that the simple linear models that fit the response time data for simple analogies might not work for more complex problems, such as difficult verbal analogies (such as those contained in the Miller Analogies Test). Sternberg (1980) makes several points in response to this observation:

Firstly, he emphasises that while it would be unrealistic to believe that the simple linear models used to predict response time simulate in precise detail the more complex, possibly non-linear strategies subjects actually used in problem solving situations, they nevertheless provide reasonable approximations.

Secondly, it is noted that although the linear models for response time are rather simple, they have proved quite successful in predicting response times in tasks that are really quite complex, such as categorical and conditional syllogisms (Guyote and Sternberg, 1978). Thus, difficulty in itself does not preclude the use of a linear model for predicting response times.

Thirdly, Sternberg (1980) stipulates that componential analysis is not predicted solely upon the use of linear models. Non-linear models have been used to predict response choices in syllogisms tasks (Guyote and Sternberg, 1978) and in analogies, classifications and series completion tasks (Sternberg and Gardener, 1979). In componential analysis and presumably other forms of analysis as well, a form of a model is selected that is appropriate for the behaviour being modelled.

Finally, Sternberg (1980) acknowledges, as Pellegrino and Lyon claim, that componential analysis may not work well for certain kinds of difficult verbal analogies. However, he attributes this lesser success not to the proposed theory of analogical reasoning, but instead suggests that it might reflect a peculiarity in the nature of complex verbal analogies (e.g., the Miller Analogies Test). Specifically, complex verbal analogies are as much or more a measure of vocabulary as they are of reasoning (Sternberg, 1974, 1977). Although these items have the "surface structure" of analogy items, Sternberg (1980) believes that they have the "deep structure" of vocabulary items, in the sense that many subjects never know the meanings of words. The major source of individual differences therefore appears to be in vocabulary rather than reasoning. Thus it is unclear whether a theory of analogical reasoning is particularly relevant to the solution of complex verbal analogies since the major difficulty they present is not one of reasoning.

Taking cognisance of Sternberg's (1980) response to Pellegrino and Lyon's (1979) comment regarding the issue of item difficulty, it is a contention of this research that in any kind of cognitive process the probability of multi-phasic, overlapping events cannot be ignored in favour of the appeal to simplicity of a neat linear sequential model. Rather, what is required, is an attempt to formulate (a) more appropriate non-linear model/s for purposes of investigating subject performance on analogy problems.

Components as Black Boxes. Pellegrino and Lyon (1979) argue that some and possibly all of the components have the character of black boxes: They name processes without specifically describing what mental operations occur when these processes are executed. It is suggested by these authors that one future direction componential analysis

ought to take is toward identifying what is done in each of the components of analogical (or other) forms of reasoning. Sternberg (1980) is in full agreement with this analysis, and believes that to accomplish this goal, the following would be necessary: A first stage of research in which the components of information processing are identified, followed by one or more subsequent stages in which the nature of these components is further explicated.

Chapter Summary

This chapter commenced with a discussion and evaluation of the two traditional approaches employed to investigate the nature of cognitive abilities, i.e., the differential and information processing approaches, went on to examine two contemporary approaches to cognitive abilities, i.e., the cognitive correlates and cognitive components approaches, and culminated with a discussion and evaluation of Sternberg's (1977) componential theory of analogical reasoning, which has as its foundation the cognitive components approach.

By employing the research technique of componential analysis used by Sternberg (1977), the present study shall attempt to investigate the extent to which subject's performance on forced choice geometric analogy items of varying complexity can offer insights into the strategic influences of mental workload variation on cognitive task performance. The rationale and procedure for the present research is presented in Chapter 4.

PART 2

THE PRESENT STUDY

CHAPTER 4

RATIONALE AND PROCEDURE

Conceptual and Methodological Issues

It will be recalled from the section The Concept of Mental Workload in Chapter 2 that for measurement of mental workload, analytical success is dependent upon an accurate and robust model of task demands. It was noted further that models of this nature are to be found in the inductive reasoning paradigm (e.g., Evans, 1968; Sternberg, 1977; Mulholland, Pellegrino and Glaser, 1980). Inductive reasoning ability has been prominent in theories of human intelligence, playing a central role in its assessment. A variety of tasks have been used to assess this ability (e.g., classification problems, series completion problems, analogy problems and matrix problems). Of these tasks, Pellegrino and Glaser (1980) regard the analogy problem as the most pervasive.

Despite the long history of inclusion of analogy problems on tests of aptitude and intelligence, a combined theoretical and empirical analysis of the component cognitive processes involved in analogical problem solving has only recently become an active research area (e.g., Sternberg, 1977, 1978; Pellegrino and Ingram, 1977; Sternberg and Rifkin, 1979; Pellegrino and Glaser, 1979, 1982; Sternberg and Nigro, 1980; Mulholland et al., 1980). Sternberg's (1977) pioneering attempt to specify the components of performance in this problem type, in particular his investigation of the performance components involved in the solution of forced choice geometric analogies, constitutes the basis for the empirical and theoretical issues explored in the present research.

The rationale here being that, of the various sources of age and ability differences in performance on inductive reasoning tasks (see the section Cognitive Abilities and Psychometrics in Chapter 3, for a description of these sources of difference) the most interesting differences appear to be those associated with content knowledge (Pellegrino and Glaser, 1982; Holzman, Pellegrino and Glaser, 1983). In attempting to solve verbal and numerical analogy problems, the critical relational inference processes depend on the knowledge of the meanings of words and their relationships or the various quantitative relations that can exist between pairs of numbers. Pellegrino (1987) argues that developmental and individual differences in performance on verbal and numerical analogy problems are partially attributable to differences in the availability and/or accessibility of such knowledge. Therefore, it is contended here that by concentrating on Sternberg's (1977) Theory of Analogical Reasoning as it applies to forced choice geometric analogy solution, it will be possible to eliminate such extraneous influences from experimental task solution.

Sternberg's (1977) theory specifies several processes applicable to all analogical reasoning tasks of the form A is to B as C is to D. The component processes include: encoding; inference; mapping; application; justification, and response (see the section The Componential Theory in Chapter 3 for a detailed discussion of these processes). A major assumption of Sternberg's theory is that of additivity across performance components. Sternberg (1977) maintains that the terms of a geometric analogy are encoded as element lists. Comparisons between the A and B terms yield a list of transformations that occur from A to B. An element comparison of the A and C terms maps the A to B transformations onto the C term such that the ideal D term can be generated. The D options are then evaluated as to whether they represent the appropriate transformation modifications of C.

Sternberg (1977) tested several componential models, each representing different assumptions about serial versus parallel processing and exhaustive versus self-terminating processing in the element encoding and comparison components (see the section Componential Models of Analogical Reasoning in Chapter 3 for a discussion of these models). The model which accounted for the greatest portion of variance across different types of analogy items was one in which encoding and inference of the transformations from A to B were serial and exhaustive, whereas mapping and application of the elements with respect to C and D were serial and self-terminating (see Figure 3.6 in the section Componential Models of Analogical Reasoning in Chapter 3 for an illustration of this model).

Sternberg's (1977) research into analogy task solution focused predominantly on comparison processes and less on the particular content of the different analogies used. Mulholland et al. (1980) argue that in Sternberg's (1977) Geometric Analogy Experiment, the nature of the element encoding process was insufficiently detailed and no descriptions were provided of the elements encoded in different analogy types. In this regard, Sternberg (1977) has commented that critical issues relating to representation (i.e., encoding and discovery) of the elements comprising the separate terms of analogy items are left unexplored in his research.

This would appear troublesome since a great deal of the time associated with geometric analogy solution seems to involve the encoding process (Mulholland et al., 1980), and it seems particularly important to gain a more comprehensive understanding of this and other related processes. Taking cognisance of this shortcoming, the present study shall undertake to investigate forced choice geometric analogy solution as a function of systematic

variations in the information structure of individual items.

Sternberg's (1977) theory advocates that in general, the overall performance on the analogy task is a function of the amount of information that must be processed to solve any given item. In the case of geometric analogy items the amount of information to be processed is considered to be a function of the number of individual elements that comprise the separate analogy terms and both the number and nature of the transformations applied to each element (Pellegrino and Glaser, 1980; Mulholland et al., 1980).

It is a contention of this study that increasing the amount of information to be processed during geometric analogy task solution will produce a corresponding increase in the load on working memory, which may require the allocation of substantial processing resources to operations designed to maintain high levels of performance. Since the demands imposed by the geometric analogy task compete for the resources of a limited capacity information processing system (Wickens et al., 1986) it could well be that control and/or memory management processes are necessary to supplement the actual processing operations associated with element and transformation processing.

It would therefore not appear unrealistic to assume that performance on highly complex forced choice geometric analogy items may necessitate a certain amount of time-sharing (McCormick and Sanders, 1982). Should the findings of the present research serve to confirm this assumption, this would have wide ranging implications for future componential analysis research in particular and cognitive task research in general, since the notion of time-sharing is a rejection of a rigid sequential, additive strategy to task solution, such as that advocated by the simple linear componential models proposed by Sternberg (1977).

The principal design components of the present study shall comprise the construction of forced choice geometric analogy items that systematically vary on the two major dimensions of item content, i.e., the elements used to construct the individual terms, and the spatial transformations that define the rules relating the pairs of terms. It is intended that such manipulation will facilitate the verification of whether additivity of element and transformation processing is maintained for overall solution latency, under conditions of varying complexity, as assumed in Sternberg's (1977) experimental work. It is proposed that should violations of additivity be detected, depending on their severity and form, these would suggest the possibility that forced choice geometric analogy task solution may follow a more integrated strategy involving interactive element and transformation processing at high levels of item complexity.

The present research shall endeavour to investigate the extent to which findings from an apparently stable cognitive phenomenon, i.e., forced choice geometric analogy task solution, performed under conditions of varying complexity, can offer insights into the strategic implications of mental workload variations on cognitive task performance.

The Measurement of Mental Workload Associated with Forced Choice Geometric Analogy Task Performance

Performance on a complex cognitive task may require an individual to monitor several sources of information, and to act on them in a decision making role. The rate of information presentation may be very low, in which case problems may arise concerning the individual's need to maintain interest and alertness. The problems with which the present study are concerned, however, are those arising as a result of an individual performing at the

upper margins of his/her ability, i.e., under conditions of high mental workload.

It will be recalled from the section The Concept of Mental Workload in Chapter 2 that despite the widespread agreement about the importance of measuring mental workload, there is as yet no consensus on how to do this, and furthermore, no universally accepted definition of the concept. Researchers from different backgrounds would probably have different priorities and purposes and hence would adopt methods of assessment best suited to their objectives. Barber (1988) emphasises that there is no primacy of one set of methods over another, though it remains important to establish which set of methods is best suited for which purpose.

Given the sheer speed of solution of geometric analogy items (a few seconds at most (Sternberg, 1977)), together with the primary objective of the present study, i.e., the investigation analogy items of varying complexity, of the mental workload estimation techniques discussed in the section Mental Workload Estimation Techniques in Chapter 2, the technique best suited to present purposes would appear to be the Primary Task measurement technique. The remaining measurement techniques discussed, i.e., Secondary Task Measures, Physiological Measures and Subjective Opinion Measures, are more suited to the estimation of workload on those tasks that are performed in the order of minutes.

Primary Task measurement techniques assume that underlying changes in workload imposed by the task will be reflected by measured changes in primary task performance (Leplat 1978) (for present purposes the primary task index of response time will be used to assess variations in task performance). Task performance variation is quantified by comparison of actual primary task performance to pre-

assigned task performance to pre-assigned task objectives. Typically, task performance is broken down into a number of task dimensions for which it is possible to obtain quantitative performance data in terms of deviations from baseline performance levels. This facilitates the selective analysis of primary task dimensions. In this study, the precueing method proposed by Sternberg (1977) (described in detail in the section Intensive Task Analysis in Chapter 3), and employed in his Geometric Analogy Experiment, will be used for the purpose of primary task decomposition.

For present purposes, decomposition will be facilitated by means of eliminating from subjects' information processing successive terms of the analogy. This is achieved by dichotomising presentation trials for each analogy item. In the first part, the subject is presented with some amount of precueing to facilitate analogy solution. In the second part, that of primary interest, the full analogy is presented. Here, solution is assumed to require merely a subset of the full set of performance components considered necessary for task solution. The general strategy being to provide subjects with an amount of prior information (precueing) in the first part of the trial, in order to reduce the amount of processing required in the second part of the trial. A more detailed description of the precueing method to be employed for present research purposes is presented in the section Procedure later in this chapter.

As was the case in Sternberg's (1977) Geometric Analogy Experiment, three types of forced choice geometric analogy will be used in the present study: Degenerate, Semidegenerate, and Nondegenerate analogies. Degenerate analogies are those in which the number of element transformations from both A to B and A to C is zero. These analogies are of the form $A:A::A:A$. Semidegenerate analogies are those in which the number of element

transformations either from A to B, or from A to C (but not both) is zero. These analogies are of the form $A:A::B:B$ or $A:B::A:B$. In general, the fourth term of a Semidegenerate analogy should match either the second or the third term. Nongenerate analogies are those in which element transformations take place from A to B and A to C. These analogies are of the form $A:B::C:D$. The fourth term of a Nongenerate analogy will not match any of the other terms. At this juncture it is important to point out that since both variations of the Semidegenerate analogy will be employed in this study, reference will be made to the "four" forced choice geometric analogy types for the remainder of this thesis.

Two precueing conditions will be used to establish baseline performance levels on items from each of the abovementioned analogy types. Firstly, a zero cue condition (no precueing information provided) will be used to establish baseline performance levels. Here the whole forced choice geometric analogy item will be presented, including the two answer options. The subject will solve it and indicate his/her response. The recorded solution time will represent the baseline performance level. Secondly, a two cue condition will be used to investigate deviations from baseline performance levels. Here the trial will be dichotomised (as discussed earlier in this section). In the first part of the trial the first two terms of the analogy will be presented. Once the subject indicates that these terms have been processed, the cue time will be recorded, and the whole analogy (including the two answer options) will be presented, this constitutes the second part of the trial. The subject will solve the analogy and indicate his/her response, the solution time will be recorded (see the section Types of Interval Scores in Chapter 3 for a discussion of cue times (scores) and solution times (scores)). The difference between the zero cue condition solution time and the two cue condition solution time

recorded for a particular item will represent the deviation from the baseline performance level.

According to the Primary Task hypothesis, mental workload increases will produce a corresponding deterioration in task performance. However, Barber (1988) considers such an assumption to be overly simplistic, since it may occur that as a task becomes more complex and imposes greater demands (in terms of number and complexity of operations to be carried out per unit of time), performance may remain unaltered, and suggests that this may occur as a result of either the individual performing within his/her own information processing limitations; or because of a change in solution strategy to accommodate increased demands. This issue receives attention in the section Subjective Influences on Cognitive Task Performance next in this chapter.

Subjective Influences on Cognitive Task Performance

It was highlighted in the section Operator Orientated Factors in Chapter 2 that a common source of difficulty in relating processing capacity to task demand is that the precise ways in which tasks are performed differ widely between individuals and within any one individual on different occasions. There are usually different strategies available for achieving adequate performance, some being more efficient than others. It was argued that skill, in terms of cognitive task performance, would appear to lie in the selection and application of the most efficient out of range of available information processing strategies.

In addition, it was argued in the section Operator Orientated Factors in Chapter 2 that increased experience on the part of an individual would appear to enhance his/her level of proficiency on a specific task or group of tasks, and thereby reduce the mental effort expended in

order to meet task demands. It was stated further, that the same actual demand would impose different levels of operational workload dependent on the cognitive capabilities of the individual.

It will be recalled from Chapter 2 that Bainbridge (1978) was of the opinion that an increase in the demands imposed by a cognitive task should influence strategy choice. It was also noted that should an individual be unable to adopt a different working method (strategy variation) in order to overcome increased task demands, he/she would experience increased mental workload, which would result in a deterioration in task performance. If increased task demand can be dealt with only by a modification in processing strategy then, it would appear realistic to assume that strategic flexibility would serve to enhance an individual's reaction to those demands.

"Indeed, such strategic changes in how a task is conducted are likely to be of particular interest in the context of mental workload estimation" (Barber, 1988, p.103).

However, because changes of this nature are unlikely to be evidenced by simple Primary Task measures such as response time and error rate, due to their low overall sensitivity (Gartner and Murphy, 1976), for purposes of investigating strategy variation it would be necessary to have recourse to more subtle indices of task performance.

The speed of solution of forced choice geometric analogy items suggests the automatic allocation of cognitive resources at high speed, with little or no time to deliberate on an optimum solution strategy. The choice and control of a strategy can be either conscious or unconscious. Given the overt nature of Primary Task Techniques, they would fail to indicate a conscious

awareness on the part of subjects, of the strategies adopted for the solution of the different item types to be used in this study, i.e., Degenerate, Semidegenerate and Nondegenerate forced choice geometric analogies. It is a contention of the present research that, given the relative simplicity of the experimental items, subjective reflection will provide a suitable means of establishing the existence of a conscious differentiation between solution strategies for different item types. Therefore, in light of the shortcomings associated with the use of Primary Task workload assessment techniques (highlighted in the section Primary Task Measures in Chapter 2), and given the argument that no single measurement technique should be used in isolation, for present purposes, a debriefing technique, i.e., a Subjective Strategy Report will be incorporated into the present experimental design.

The Research Design of Sternberg's (1977) Geometric Analogy Experiment and its Relevance to the Present Research

The research procedure adopted in Sternberg's (1977) Geometric Analogy Experiment (in particular, the method of subject selection and number of subjects selected) does not facilitate the investigation of item complexity influences on the processes of solution strategy selection and administration.

In the Sternberg (1977) experiment subjects were selected from a population of 441 students studying Psychology I at Stanford University. All of the students were required to complete two tests. The first test was a Word Grouping Test. The second test was the Card Rotation Test (Test S-1) from the French Test of Reference Tests for Cognitive Factors (French, Ekstrom and Price, 1963).

Only the Word Grouping Test was used in subject selection

"... since selection was desired on reasoning but not spatial visualisation" (Sternberg, 1977, p.256). Subjects were then selected from two groups, i.e., those scoring between the 5th and 25th percentiles and those scoring between the 75th and 95th percentiles on the Word Grouping Test. Twelve subjects (six male and six female) were selected from each of the abovementioned groups. In total twenty-four subjects participated in Sternberg's (1977) Geometric Analogy Experiment.

Given the highly specific nature of the subject selection process used by Sternberg (1977), it is contended that such a selection procedure would be too restrictive for present purposes and furthermore, it suggests a strong a priori expectation on Sternberg's part. The solution strategies of two highly specific groups of subjects would in no way provide an accurate indication of the range of solution strategies available for forced choice geometric analogy solution. Establishing the existence of meaningful strategic variations within, as well as between subject performance on forced choice geometric analogy items would require a more general subject sample, from a more general subject population. Moreover, it is argued that a sample of twenty-four subjects would be insufficient to facilitate the application of a suitable data reduction technique for the purpose of differentiating between alternate solution strategies.

Of particular significance, is that at no time did Sternberg (1977) attempt to establish whether any of the subjects participating in the Geometric Analogy Experiment had previously participated in any form of experimentation involving the solution of analogy problems. As a result, it would not have been possible to establish whether past experience had a moderating effect on task performance as measured in the experiment, or whether the results yielded were merely attributable to variations in item complexity

(the effects of experience and inexperience on task performance were discussed in detail in the section Operator Orientated Factors in Chapter 2). For present purposes it would be necessary to select a sample of subjects on the basis that previous exposure to analogical reasoning tasks has been minimal, unstructured and certainly not subjected to psychometric screening. Consequently, their reasoning abilities will be subject to random variability. It can therefore be assumed that subjective task performance will be representative of the range of reasoning strategies available for forced choice geometric analogy solution, as well as highly responsive to workload influences.

It was noted in the section Conceptual and Methodological Issues earlier in this chapter that providing subjects with prior information in the first part of item presentation serves to reduce the amount of processing required in the second part of item presentation. It is proposed here that providing precueing information in the first part of item presentation serves to reduce the degree of task uncertainty, and therefore mental workload, experienced by subjects during task solution. Furthermore, it is contended that the incorporation of the precueing procedure described in the section Conceptual and Methodological Issues earlier in this chapter into the present experimental design, will facilitate the manipulation of task uncertainty. In the present study the highest degree of task uncertainty will occur with the zero cue precueing condition, whereas the two cue precueing condition will represent a state of greatly reduced task uncertainty.

The present experimental design will resemble that adopted in the Sternberg (1977) Geometric Analogy Experiment only in terms of the item selection, manipulation and presentation procedures, as these have been shown by

Sternberg (1977) to be both stable and reliable. This will permit the comparison of the results obtained from this study with those achieved by Sternberg (1977).

Assessment of Subjective Influences

Subjective Influences and Cluster Analysis

In terms of the central issues of the present research involving the inductive reasoning paradigm, investigation has mainly followed the "classical" approaches to research design and analysis (e.g., Fisher, 1942; Sheffe, 1959; Kerlinger, 1973; Pedhazur, 1982), which hinge upon the assumptions of within and between group variance and involve the conceptual manipulation and examination of statistical variance (e.g., analysis of variance, parametric correlation and regression techniques etc.) in dependent variable measures as the result of some controlled use of experimental independent variable(s). In most instances the analysis technique employed would be some variation or other of factorial analysis of variance. However, given the reservations expressed in Chapter 3 concerning the use of factor analysis in cognitive psychology, it follows that some reservations must be extended to the value and appropriateness of statistical techniques used in data analysis in many of the key inductive reasoning experiments (e.g., Sternberg, 1977, 1978; Pellegrino and Ingram, 1977; Sternberg and Rifkin, 1979; Pellegrino and Glaser, 1979, 1982; Sternberg and Nigro, 1980; Mulholland, Pellegrino and Glaser, 1980).

In particular, the use of small numbers of subjects in repeated measures designs involving large numbers of independent variables may result in a violation of the principle of providing sufficient scores within each cell of the experimental design to facilitate an appropriate amount of variation in dependent measures to support the

intended analysis. This is evidenced by Sternberg's (1977) statement, while commenting on the outcome of his Geometric Analogy Experiment, that:-

"The reliabilities were disappointing ... apparently 180 observations per subject were too few to obtain reliable parameter estimates for individual subjects" (Sternberg, 1977; p.276)

Furthermore, whereas techniques involving the multivariate analysis of variance are highly appropriate methods of assessing variance across pre-viously decided and designed experimental conditions, they do not necessarily facilitate the more qualitative assessment of partitioning such data sets into groups of individuals without a priori classification. This apparent shortcoming is compounded by the fact that a small subject sample would make it extremely difficult to search for groupings, clusterings or other patterns that might emerge in the experimental data set. However, with a more sizeable data set this obstacle could well be overcome, thereby providing additional insight into relationships between task, score and individual differences, while at the same time avoiding the trap of grouping individuals solely on the basis of a priori assumptions.

To pursue this rather speculative notion further, a truly exploratory approach to this classification problem would demand the relegation of classical multivariate analysis procedures in favour of a more flexible and interpretive cluster analytical technique.

In contrast to multivariate analysis procedures, in concept at least, clustering methods are used to discover groupings amongst subjects' performance on cognitive tasks without embracing unsatisfactory a priori assumptions. Aldenderfer and Blashfield (1984) have suggested that most

applications of cluster analysis can be placed into one of four categories, viz: the development of typologies for classification; investigation of useful conceptual schemes for grouping objects; hypothesis generation through data exploration; and hypothesis testing. Thus, there would appear to be sufficient precedent in the use of cluster analysis to justify its use in the present research context. At this juncture it would be appropriate to consider certain general points about cluster analysis techniques in order to assess both the strengths and weaknesses associated with their use.

In simple terms, a cluster analysis seeks to group objects on the basis of a set of variables, or attributes of the object. Variables may be any type of nominal, ordinal, interval or ratio measurement, and a complete row of variables is called an object's profile (Lorr, 1983). The outcomes of such analyses are referred to as groups, types, classes of clusters.

Although the strategy of cluster analysis is essentially cluster seeking, in practice the operation of a clustering algorithm is structure imposing. In this way clustering techniques seek to identify groups of objects which are similar to one another, without an a priori assumption of their existence. The pre-classificatory nature of the clustering transformation processes would appear most advantageous in the context of the present research considerations, given that typically neither the number or nature of such groupings are known prior to the analysis.

Ironically, this apparent advantage is also a major weakness, since cluster methods will yield clusters irrespective of the structure of the data, even if the distributions are random (Sneath and Sokal, 1973). Of major concern, is to establish why a given set of clusterings emerge and furthermore, to differentiate

between those groups which are meaningful and those which may have merely been imposed by the clustering method.

Essentially, clustering methods are relatively simple statistical procedures, intended to act mainly as data reduction heuristics and should thus not be viewed as a validation of any particular methodological approach but rather as a formal statement of relationships within a data set.

Of particular importance is to select from the range of clustering algorithms available, one that is most appropriate for the theoretical context of the research questions; an inappropriate technique will yield inappropriate clusterings (Aldenderfer and Blashfield, 1984). Since the choice of clustering algorithm is of prime concern in the context of the present research, discussion thereof will be resumed later in this section, as well as in the section The Cluster Analysis Procedures in Chapter 5 (Data Analysis and Results) when considering some different techniques in cluster analysis.

Given then that a clustering technique is essentially a simple data reduction heuristic for arranging objects into clusters on some basis of observed similarity, it would be appropriate to consider the related issues of concept of similarity measurement and the concept of transforming different variables within a profile that may potentially embrace each form of data (nominal, ordinal, interval and ratio) so that similarity might be established.

Since the object profiles to which cluster analysis will be applied in the present research will comprise only ratio data, the issue of transforming variables requires no elaboration. Further discussion of the implications of choosing variables for the object profile will be delayed until the sections Overview of Data Analysis and The

Cluster Analysis Procedures in Chapter 5 (Data Analysis and Results).

By contrast, the essential choice of similarity measure has proven to be a major conceptual problem inherent in cluster analysis. The concept of similarity is fundamental to the formation of clusters, with differing measures of similarity demanding differing transformation and comparison techniques. The quantitative estimation of similarity has been dominated by the concept of metrics, involving the representation of objects as points in co-ordinate space such that similarities and dissimilarities observed between these points correspond to distance representations between them (Tversky, 1977). Based upon an assessment of their inherent mathematical properties, Sneath and Sokal (1973) identified four principal measures of similarity, viz: correlation coefficients, distance measures, association measures and probabilistic similarity measures. However, to date, only correlation coefficients and distance metrics have been widely used in the social sciences (Aldenderfer and Blashfield, 1984). Since both association coefficients and probabilistic similarity coefficients are confined to binary data by virtue of their mathematical properties (Sneath and Sokal, 1973; Clifford and Stephenson, 1975), and given the properties of the data set planned to emerge from the methodological framework articulated in the following sections of this chapter, neither association coefficients, probabilistic similarity coefficients or correlation coefficients require elaboration. Rather, a brief discussion of distance measures and their potential limitations is appropriate at this point.

Aldenderfer and Blashfield (1984) propose that, technically, distance measures are dissimilarity measures in that objects are judged identical if each one is described by variables with the same magnitude, i.e., if

the distance between them is zero. The nature of distance measures is such that ordinarily they have no upper bounds and are scale dependent, resulting in their becoming more valuable as the variable reflects interval and ratio data scores.

Amongst the potential limitations of distance measures are the problems posed for cluster interpretation. Critically, estimation of similarity between objects is significantly affected by elevation (the mean score of the object over all the variables). Consequently, variables with both large size differences and large standard deviations can effectively swamp other variables with smaller absolute sizes and standard deviations. Moreover, distance measures are also affected by transformation of the scale of measurement of variables in a profile (Everitt, 1980) and this may require standardisation of all variables to unit variance and means of zero before the calculation of distance. However, since the object profile to which cluster analysis will be applied in this study will comprise only ratio data, this particular limitation would be irrelevant in the context of the present research.

Given the wide range of similarity measures, each with advantages and disadvantages, it would be unwise to undertake a cluster analysis without careful consideration of the measurement implications inherent in the chosen technique (Aldenderfer and Blashfield, 1984).

Accordingly discussion of the choice of cluster technique appropriate to the present research need will be delayed until after an explanation of the present research method and procedure, and will therefore be returned to in the sections Overview of Data Analysis and The Cluster Analysis Procedures in Chapter 5 (Data Analysis and Results).

Having thus far considered some conceptual aspects of clustering, of similarity measures and of the selection of profile variables, it is appropriate at this juncture to return to those issues raised earlier in this section concerning the range of clustering techniques and the problems of cluster validation. Aldenderfer and Blashfield (1984) identify and discuss seven major families of clustering method, viz; hierarchical agglomerative, factor analytic, iterative partitioning, density search, clumping and graph theoretic. Each family represents shared assumptions and objectives in clustering concept, mathematical assumptions and operation. However, by definition, these shared features do not necessarily extend to relationships between families. Consequently, results yielded by each type of method can vary greatly, even when applied to the same data set. This would suggest that the choice of cluster method should reflect a decision integral to the research need. Aldenderfer and Blashfield (1984) reinforce the importance of this choice by their insistence that:-

"... the ... choice of ... clustering method must be compatible with the desired nature of the classification, the variables to be used and the similarity measure used to estimate the resemblance between ... (objects)" (Aldenderfer and Blashfield, 1984; p.35)

As is the case with similarity measures, each different clustering method presents its own advantages and disadvantages. However, it is beyond the scope of this study to enter into a review of these features. Rather it will suffice to acknowledge that choice of technique should reflect the context and aspirations of the research and accordingly, discussion of the choice of clustering method appropriate to the present research objectives will be delayed until after an explanation of the present research

method and procedure, and will therefore be returned to in the section The Cluster Analysis Procedures in Chapter 5 (Data Analysis and Results).

The second issue reserved for consideration was that of validating emergent cluster solutions. As noted earlier in this section, cluster analysis techniques hinge around the application of cluster imposing algorithms and are therefore more a formal statement of relationships rather than an indication of "meaningful" groupings. Aldenderfer and Blashfield (1984) argue that most solutions to the problem of knowing how many clusters are revealed by an analysis are very subjective in appearance and are particularly dependant upon both the research context and the type of clustering method used. However, since the issue of deciding the number of clusters is contingent upon the clustering method used, discussion thereof will be reserved until the method and procedure of the present research has been examined in the remainder of this chapter. Accordingly this issue will be returned to in the section The Cluster Analysis Procedures in Chapter 5 (Data Analysis and Results).

Five techniques of validating cluster analysis solutions emerge from the literature (viz: the cophonetic correlation (Sokal and Rohlf, 1963), significance tests on the variables used to create clusters, replication, additional tests outside of the clustering technique carried out on the independent variables, and Monte Carlo (random number substitution procedures), each with its associated advantages and disadvantages and, given a particular research context each may either be applicable or completely inappropriate. Once again, since the choice of appropriate validation technique is influenced by the choice of clustering method, an in depth review of validation principles is unnecessary here. Rather, this issue will receive further consideration in the section

The Cluster Analysis Procedures in Chapter 5 (Data Analysis and Results).

The present section has presented a brief overview of the concept of cluster analysis as an exploratory method for forming groupings in multivariate data sets. In concept then cluster analysis would appear to offer a refreshing alternative in the investigation of cognitive inter and intraindividual differences, in that the techniques are preclassificatory and thus avoid the constraints associated with the classical multivariate analysis procedures reviewed in Part 1. Despite the practical and mathematical uncertainties of cluster analysis it is contended that its use in the investigation of mental workload influences on cognitive task performance could prove most insightful. An explanation of the aims, methodology and procedure of the present research will take place in the remaining sections of this chapter.

Subjective Strategy Report

The subjective strategy report will attempt to elicit from subjects a breakdown of the strategies used to solve the different item types presented in the experiment, i.e., Degenerate, Semidegenerate and Nondegenerate forced choice geometric analogies. The objective being to investigate whether a particular solution strategy can be generalised, or whether systematic differences exist within, as well as between subject performance. The subjective strategy report is intended to yield a partitioning of solution strategies.

For each geometric analogy type, the association between the outcome of the subjective strategy report and the outcome of the cluster analysis procedure will be investigated as this will serve to establish whether the subjective reports are in fact indicative of meaningful

strategic variations, i.e., whether they serve as an indicator of response time cluster membership.

Task Complexity Ranking

The task complexity ranking procedure will require subjects to rank the "four" geometric analogy types used in the study, i.e., Degenerate, Semidegenerate (both $A:A::B:B$ and $A:B::A:B$ variations) and Nondegenerate, in terms of their perceived complexity. This procedure will be conducted both before and after the presentation of the total item complement. The aim here is twofold: Firstly, to investigate whether strategic variations in task performance can be related to differences in the subjects' impressions of stimulus complexity and secondly, to determine whether performance on forced choice geometric analogy items has an effect on the subjects' perceptions of item complexity.

Aims of the Present Research

Sternberg's (1977) pioneering attempt at specifying the information processing components involved in geometric analogy solution constitutes the basis for the empirical and theoretical issues explored in the present study. However, it must be emphasised that by no means is the present study intended to dispute the findings of Sternberg's (1977) research. Rather, its aims are twofold. Firstly, to investigate whether the componential models Sternberg (1977) describes as best defining task performance on Nondegenerate forced choice geometric analogies are in fact as generalisable as he purports them to be or, alternatively, whether systematic individual differences exist within, as well as between subject performance. More specifically, whether meaningful strategic variations are apparent, and if so whether such variations could be attributed to the level of mental workload experienced by subjects as a result of their

exposure to the demands of the experimental tasks. And, secondly, based on the outcome of the present study, propose models that best define task performance on Degenerate and Semidegenerate forced choice geometric analogies, in terms of performance components adopted and order of component execution only.

Hypothesis

Increasing the complexity of forced choice geometric analogies, while simultaneously manipulating the degree of task uncertainty, will elicit meaningful strategic variations on the part of subjects in order to overcome associated increases in mental workload and maintain a high standard of task performance.

Subjects

Sixty subjects were tested; thirty one males and twenty nine females. Subjects were all volunteers and were all drawn at random from the population of undergraduates, postgraduates and staff of the University of the Witwatersrand and friends and acquaintances of the researcher. Most critically, from questioning during recruitment, it was firmly established that no subject was familiar with cognitive research methods in general or the cognitive components approach to analogical reasoning in particular (see section Procedure later in this chapter).

Handedness of subjects was assessed and recorded (see the section Procedure later in this chapter), but was not used as a factor in the sampling strategy. No subject was recorded as being ambidextrous. Three subjects (5%) were recorded as being dominantly left handed (one male; two females). The remaining fifty seven subjects were recorded as being dominantly right handed (thirty males; twenty seven females).

Age of subjects was recorded to age at nearest (next-last) birthday. Mean age for the sample was 22.96 years (males 24.19; females 21.65), standard deviation 3.32 years (males 3.84; females 2.0), range 16.0 years (males 14.0; females 8.0).

All subjects were thus of adult age and were assumed therefore to have acquired the normal adult reasoning capability of formal operations (Inhelder and Piaget, 1954; Flavell, 1963) though this assumption was not the subject of any independent assessment. No subject reported having been diagnosed as suffering from any form of cognitive impairment (see the section Procedure later in this chapter). For all subjects vision was confirmed as either normal or corrected to normal (see the section Procedure later in this chapter).

Design and Apparatus

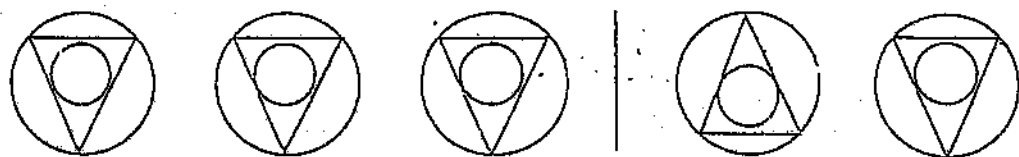
The experimental stimuli were sixty-six forced choice geometric analogies. Analogies were generated from six types of common elements found in geometric analogy test items: a line, triangle, circle, cross, rectangle and square. Each element had regular and irregular forms and areas and lengths of sides were varied to facilitate item construction. Elements were combined into configurations of one, two and three to form the terms of the analogies. A set of seven transformations representative of test item transformations were used: identity (no transformation), increase in size, decrease in size, rotation to the right, addition of an element, removal of one half of an element, and inversion of an element. Either zero (in the case of Degenerate geometric analogies), one, two or three transformations were applied to the elements comprising the terms of the analogies. In line with the hypotheses of Sternberg (1977) and Mulholland et al. (1980), increasing the number of elements and transformations within items was assumed to increase item complexity.

Of the sixty-six experimental analogies, thirty (45%) appeared in a Degenerate or Semidegenerate form. Of these analogies, six appeared in the form $A:A::A:D1,D2$; twelve appeared in the form $A:A::B:D1,D2$; and another twelve appeared in the form $A:B::A:D1,D2$. For these analogies and the thirty six Nondegenerate analogies (which appeared in the form $A:B::C:D1,D2$), the correct answer option was D1 for half the items and D2 for the other half. Items were presented in either a zero or two cue condition.

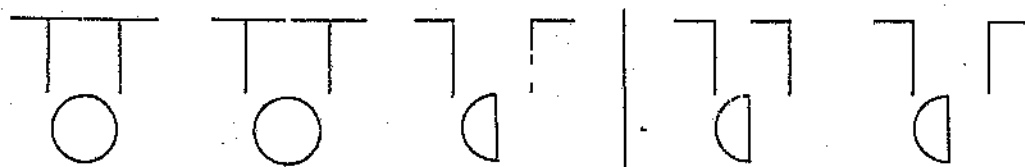
Figure 4.1 displays items similar to those used in the experiment. As in the actual items, a single vertical line separates the first three terms of the analogy from the two answer options. In the case of the experimental items the geometric figures were centered on the horizontal centre line of each slide, and the vertical line was drawn the full length of each slide so that it would not be viewed as part of the item. The first item presented in Figure 4.1 is Degenerate, the second Semidegenerate (of the form $A:A::B:D1,D2$) the third Semidegenerate (of the form $A:B::A:D1,D2$), and the fourth Nondegenerate. It will be noticed that the elements comprising the terms of the first item have undergone zero transformations, whereas those comprising certain terms (depending on analogy type) of the second, third and fourth items have undergone three transformations.

The following Aparatus was used:-

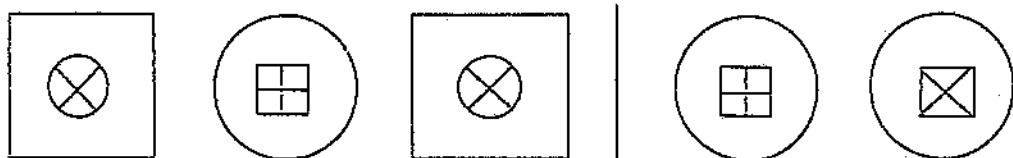
A Kodak Carousel Auto Focus Slide Projector Model 850H. The slide projector was linked to a Lafayette Instruments Company Limited projection tachistoscope Model 41010. The tachistoscope was linked to a Lafayette Instruments Company Limited "Voice Response Time Control Set", which in turn was linked to a throat microphone. The voice response time control set was set on a single cycle and linked to a Lafayette Instruments Company Limited millisecond



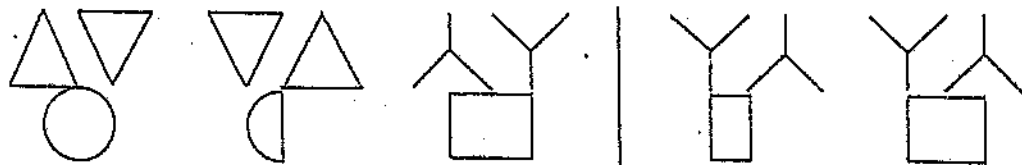
Degenerate Forced Choice Geometric Analogy : 3 Elements



Semidegenerate Forced Choice Geometric Analogy
(Format A:A::B:B) : 3 Elements - 3 Transformations



Semidegenerate Forced Geometric Analogy
(Format A:B::A:B) : 3 Elements - 3 Transformations



Nongenerate Forced Choice Geometric Analogy
3 Elements - 3 Transformations

FIG. 4.1. FORCED CHOICE GEOMETRIC ANALOGIES SIMILAR
TO THOSE INCLUDED IN THE EXPERIMENTAL
TRIALS.

clock/time/counter comparator model 54530. The clock was set to record in milliseconds.

Each of the sixty-six experimental analogies was drawn with black ink on white paper. The items were photographed separately, producing a single 35mm, positive transparency for each item presented in the zero cue condition and two 35mm, positive transparencies for each item presented in the two cue condition (i.e., one slide of the first two terms of the analogy, and one slide of the whole analogy). In total then, ninety nine stimulus slides were prepared for use as projection stimulus material. The size of the projected image measured 43cm horizontally and 27.5cm vertically. For each stimulus item the vertical line separating the first three terms from the two answer options was located 5.5cm to the right of the vertical centre line.

Procedure

In order to escape the problems of repetitive use of a small number of highly trained, hypothesis sophisticated subjects and to facilitate the application of robust exploration clustering techniques, a large sample (60) of naive subjects were recruited and tested. Subjects were approached individually and asked if they would care to participate in some cognitive psychology research.

In order to avoid the use of hypothesis sophisticated subjects (see the section Subjective Influences on Cognitive Task Performance earlier in this chapter) each participating individual was questioned as to the extent of their awareness of inductive reasoning, cognitive components theory and mental workload issues. Anyone showing even the slightest understanding was excluded from further stages of the research.

The experiment was conducted in a quiet laboratory cubicle set aside for that purpose. Each subject was tested individually, with only the researcher present to administer and conduct the research activity.

Subjects were firstly put at ease through general conversation, in no way related to the specific research to be done. Those subjects who raised the issue, were informed that the research was neither a competition or a personal test. When subjects confirmed that they were relaxed and physically comfortable the experimental procedure was initiated.

Subjects were asked to state their age at the nearest (next-last) birthday and this was recorded for confirmation of adulthood only (see the section Subjects earlier in this chapter). Gender was also recorded on the basis of *prima facie* evidence. Handedness was assessed by asking subjects to rate themselves as either dominantly left handed, dominantly right handed, or ambidextrous. No subject reported being ambidextrous.

Age, handedness and gender, though recorded, played no further part in the experimental procedure, but were noted with a view to any post hoc analysis need that might occur.

Subjects were then asked whether as far as they knew they had been clinically diagnosed as suffering any deficit or disorder of spatial ability. No subjects reported any known problem.

Subjects were then asked whether their vision was normal or corrected to normal, with contact lenses, spectacles, etc. Such aids were allowed to be worn for the experimental trials. All subjects conformed to a loose expectation of normal vision.

Subjects were then introduced to analogy problems. It was emphasised that the aim of the present research was to investigate task performance on analogy problems such as those found on psychometric tests. It was explained that analogy problems were used to assess the primary mental ability known as induction, which refers to the capacity to generalise from specific experiences and to form new more abstract concepts.

Using a simple Nondegenerate verbal analogy as an example, the basic format of analogy problems was then explained, until it was reported by subjects that they understood the nature of the task. Subjects were then informed that what they had just been shown was a verbal analogy, and that the present study involved the solution of geometric analogies which in contrast to verbal analogies comprised geometric shapes rather than words. The basic components comprising a geometric analogy were then explained using as an example a simple forced choice Nondegenerate geometric analogy similar to those used in the experimental trials. Here, subjects were told that all of the experimental analogies would comprise a first, second and third term, and two answer options, the first option being labelled A and the second option B. It was pointed out that each of the terms as well as the two answer options would comprise a certain number of geometric shapes, referred to as elements.

Subjects were informed that the number of elements comprising each of the terms making up the experimental analogies would vary, but that the number of elements per term in a single analogy would remain constant. It was then pointed out that the elements comprising the terms of the experimental analogies could undergo transformations in shape, size and orientation, but that the terms in which elements would be transformed would depend on geometric analogy type.

At this point the "four" types of forced choice geometric analogy used in the experiment were identified, using simple examples of each type. The basic format of each type was then explained in detail. Particular care was taken to ensure that subjects understood the basic format of each geometric analogy type, and could successfully differentiate between the types.

Subjects were then told that experimental analogies would be presented in one of two conditions. Simple forced choice geometric analogies similar to the experimental analogies were used to illustrate the two presentation conditions. It was explained that in the first condition, identified as the zero cue condition, the entire analogy (i.e., only one stimulus slide) would be presented. In the second condition, identified as the two cue condition, it was explained that two stimuli would be presented. The first stimulus would comprise only the first and second terms of the analogy, whereas the second stimulus would comprise the first and second terms (as presented in the first stimulus), as well as the remainder of the analogy. Once again particular care was taken to ensure that subjects could identify and differentiate between the two conditions of item presentation.

Subjects were then presented with one example of each of the forced choice geometric analogy types, each example being similar to the experimental items. The Degenerate example comprised three elements, whereas the Semidegenerate and Nondegenerate examples each comprised three elements and three transformations. Having considered each of the examples, paying particular attention to the component elements and the transformations, subjects were asked to rank the four analogies in order of complexity; a ranking of 1 was given to the least complex, and 4 for the most complex. This

ranking was recorded. Subjects were not asked to explain the basis of their ranking judgements.

The apparatus was then explained; how the slide projector and projection tachistoscope were linked, how item presentation would be controlled, how the voice key and timer were linked, etc. The throat microphone was then fitted and gain/sensitivity controls adjusted.

The operation of the throat microphone was then explained and subjects were warned against swallowing or making unwanted vocal noises during the duration of the experimental trials. Subjects readily understood and accepted these explanations.

The experimental task was then explained. Subjects were informed that once they had indicated their readiness to begin the experiment the experimenter would initiate the presentation of the analogies. The subjects' task was to solve the projected analogies as quickly as possible while at the same time trying to maintain a high standard of accuracy.

In the case of an analogy appearing in the zero cue condition, subjects were instructed to solve the analogy and give the verbal response "A" if the first answer option was considered correct, or alternatively the verbal response "B" if the second answer option was considered correct. In the case of an analogy being presented in the two cue condition, subjects were instructed to respond as follows: upon presentation of the initial stimulus comprising the first two terms of the analogy, the task was to examine the two terms and establish the relationship between them, viz: were they identical, or were they different, and if so to identify the transformations. Once this was accomplished the subjects were told to give the verbal response "go". This would immediately result in the

second stimulus appearing (comprising the whole analogy), the task are being to solve the analogy with the aid of the information gained from the previous stimulus, and give the verbal response "A" or "B" depending on which answer option was considered correct.

It should be noticed that this form of verbal response produces minor but consistent differences between "A" and "B" responses. Despite this measurement inaccuracy the method is widely used in experimental psychology and has distinct advantages over button press (Mulholland et al., 1980) and footpedal (Sternberg, 1977) responses, which through their dependence upon different spatial positioning of the response keys/footpedal, often involve added complexity in combining verbal modes of input-output and the uncertain relationship between spatial and motor representation (Le Doux, Wilson and Gazzaniga, 1977) together with the problem of motor output differences caused by hand/foot dominance. These are, of course, entirely unwanted and in the present context the exclusion of their contaminating influence is considered more critical than the known measurement inaccuracy of the voice key.

The subjects' task was then defined as to solve the projected analogy and respond as fast as possible and as accurate as possible. Each verbal response activated the voice key and stopped the timer. The response time (in milliseconds) was recorded by the experimenter. The slide was then replaced by the next in the series and the process was repeated

Subjects were then given a small number of pilot trials on analogies not used in the experimental trials, consisting of items representing each of the "four" types of analogy, in both zero and two cue conditions. For both pilot trials and experimental trials, subjects were seated 90cm from a

blank white wall on to which items were rear projected, at the subjects' seated eye level. Response times for pilot trials were not recorded. Following these pilot trials a short break was allowed whilst the experimental apparatus was checked. Experimental trials were confined to one session of ninety nine stimulus slides (sixty-six forced choice geometric analogy items), which were re-randomised for each subject. The elapsed time between a subject's response to one item and the appearance of the next item was approximately 5 seconds. A rest period was given during the change of the 80-item slide carousel.

After all experimental trials had been completed, subjects were asked to repeat the ranking of complexity of the same four examples presented prior to the experimental trials. These rankings were recorded on the same basis as the pre-experimental rankings described earlier in this section.

Finally, subjects were asked whether they had developed any conscious strategy for handling the different types of forced choice geometric analogies included in the experimental trials. All subjects reported that they had. Subjects were then presented with two examples of each of the geometric analogy types included in the experimental trials (as they would have appeared in the zero cue condition). The first Degenerate example comprised one element throughout, and the second example three elements throughout. In the case of the Semidegenerate and Nondegenerate examples, one example comprised one element per term with one transformation, whereas the other example comprised three elements per term with three transformation. Each of the eight examples presented had been included in the experimental trials.

It was explained to subjects that what was required of them here, was to give a step by step account of the solution

strategy they had adopted for experimental trials similar to each of the eight examples, paying particular attention to which of the terms had been compared and the order in which this had taken place. Following this, only for the Nondegenerate example comprising three elements and three transformations, subjects were questioned as to whether the elements comprising the terms had been compared individually, or as a whole. Then, for each of the eight examples subjects were asked to identify which of the analogy terms they had compared with the two answer options before responding. Again, only for the Nondegenerate example comprising three elements and three transformations, subjects were required to identify whether elements comprising the term(s) had been compared individually, or as a whole with the elements comprising the two answer options. Finally, subjects were questioned as to whether the answer options had been considered simultaneously, or one at a time. Each of the reported strategies was recorded and coded by the experimenter. This completed the experimental interaction. Subjects were thanked for their participation in the experiment.

Total time for the complete experimental interaction varied from subject to subject, taking between approximately 25 minutes and 45 minutes.

The Nature of the Data Yielded

Excluding the details of age, gender, handedness and subjective strategy indicator code classifications, two data sets emerge.

The first, subsidiary, set relates to the scores for subjective complexity ranking of the experimental stimulus analogies on a pre and post experiment basis. There were no missing values.

The second, and major, set consisted exclusively of response times, recorded in milliseconds, and encompassed subjects' performances across conditions made up of the variables of number of elements, number of transformations, cue condition, and analogy type, to form a classical $3 \times 3 \times 2 \times 4$ multivariate matrix (Edwards, 1960). There were no missing values.

Chapter Summary

This chapter had drawn together the major points to emerge from the discussion in Part 1 of the concepts of inductive reasoning, individual differences and mental workload. Moreover, it has outlined a research procedure to investigate more appropriately these relationships whilst seeking to expose the influence of experimental demand characteristics and subjective interpretation thereof upon the data.

The intention is to use a sequence of data analysis techniques on the two data sets yielded to investigate the following issues: The effects of the role of subjective stimulus complexity, the effects of the experimental task variables on performance variation, and the extent to which an appropriate clustering algorithm will lead to exposure of individual processing groupings not exposed by other forms of variance analysis.

Data reduction and analysis will be discussed in more detail in Chapter 5 (Data Analysis and Results).

DATA ANALYSIS AND RESULTS

Overview to Data Analysis

The research procedure outlined in Chapter 4 yielded two main data sets, together with the recorded details of subjects' age, gender, handedness and subjective strategy indicator codes. Applying a sequence of suitable data analysis techniques to these combined data will facilitate the investigation of a number of interrelated issues. However, since "four" types of forced choice geometric analogy were included in the experimental trials, each varying significantly in format and demand characteristics, for purposes of data analysis each analogy type will be considered separately.

It was emphasised in the section Procedure in Chapter 4 that the main experimental task required each subject to solve sixty-six forced choice geometric analogies incorporating the independent variables of number of elements per term, number of transformations between terms, cue condition and analogy type. However, since for purposes of data analysis performance on each analogy type will be considered separately, the original data set comprising subjects' response times will be subdivided into four separate data sets, each comprising response times for a particular analogy type, thus producing a separate factorial design for each analogy type. In view of the need to repeatedly address the same set of data to pursue different issues, it will be relevant to discuss the treatment of data sets in the order of the issues to be considered.

The first issue to consider is the performance of the experimental task per se. The Degenerate analogies included in the experimental trials incorporated the independent variables of number of elements per term and cue condition. The number of elements ranged from one to three and the cue condition was restricted to either zero or two, thus yielding a 3×2 factorial design. Both Semidegenerate geometric analogy types incorporated the independent variables of number of elements per term, number of transformations between terms and cue condition. Here the number of elements ranged from one to three, the number of transformations was limited to one or three, while the cue condition was restricted to either zero or two. This produced a $3 \times 2 \times 2$ factorial design for both Semidegenerate geometric analogy types. The Nondegenerate geometric analogies included in the experimental trials incorporated the independent variables of number of elements per term, number of transformations between terms and cue condition. Here, both the number of elements and number of transformations ranged from one to three, while the cue condition was either zero or two. This produced a $3 \times 3 \times 2$ factorial design.

The variables of number of elements per term and number of transformations between terms have been an integral part of previous research into the geometric analogy task (e.g., Sternberg, 1978; Mulholland et al., 1980; Pellegrino and Glaser, 1980), however, only Sternberg (1977) has considered the variable of cue condition. The variable of analogy type has not previously been considered. The forced choice geometric analogy task typically involves establishing the relationship between the first and second terms of the analogy, and then using that information to complete the sequence by means of identifying an analogous relationship between the third term and one of the two answer options. Deductive analyses similar to those standardly reported in geometric analogy experiments will be performed on these data.

However, performing a separate analysis on the data set pertaining to each of the "four" analogy types, together with the large number of subjects tested (relative to previous geometric analogy experiments) should facilitate further analysis of the extent of strategy variation on an intraindividual basis in a way not previously investigated.

An explanation of the statistical analysis involved in examining the issue of experimental task performance, together with summary tables of consequent data, is set out in the section The Main Experiment later in this chapter.

The second issue to be addressed is that of stimulus complexity. The effect of stimulus complexity on task performance has been widely discussed in previous investigations of geometric analogy solution (e.g., Evans, 1968; Sternberg, 1977; Mulholland et al., 1980), however, there is as yet no agreed upon measure of item complexity. From the points raised in the section Subjective Influences on Cognitive Task Performance in Chapter 4, there is good reason to believe that simple measures, such as number of elements, and number of transformations, will not necessarily determine item complexity.

Taking such problems into consideration, the present method for assessing the effect of stimulus complexity concentrates on the level of subjective assessment. By taking subjective rank order assessments of stimulus complexity both before and after the presentation of the full item complement it is intended to establish whether exposure to, and performance on the experimental trials had an effect on subjects' perceptions of stimulus complexity. These findings will be incorporated into the analysis of the research steps to follow.

An explanation of the statistical analysis involved in examining the stimulus complexity issue, together with the

summary tables of the consequent data, is set out in the section Stimulus Complexity and the Experimental Task later on in this chapter.

The third issue to consider is the extent to which the demand characteristics of each of the "four" geometric analogy types influences the particular solution strategy adopted by subjects. It is intended to establish for each analogy type whether a particular solution strategy can be generalised, or whether systematic differences exist within, as well as between subject performance.

An explanation of the statistical analysis involved in examining the influence of task demand characteristics on task solution strategy, together with the summary tables of the consequent data, is set out in the section Solution Strategy and the Experimental Task later on in this chapter.

The fourth and final issue to be considered is the extent to which appropriate clustering techniques can provide insight into the nature of the forced choice geometric analogy processing differences emerging in the main experimental task. To examine this issue more fully, an appropriate clustering algorithm will be applied to the reaction time data emerging from the experimental trials. Particular care will be taken to ensure that the selection and application of the clustering technique, as well as the interpretation and validation of any cluster solution are consistent with the theoretical and methodological context of the present research.

An explanation and justification of the statistical analysis involved in examining the clustering issue, together with the emergent data, are presented in the section The Cluster Analysis Procedures later in this chapter.

Finally, it will be the chosen style of this thesis to present the results of each data analysis in the form of tables only. Graphs have been excluded specifically to avoid the tendency to discuss results in terms of reaction time slopes, number of transformations between terms, intercepts, etc., as is typically the case in geometric analogy research. As discussed in the section Individual Differences and Cluster Analysis in Chapter 4, it is the fundamental rationale of this present research that such terms predispose interpretation of results and, moreover, that the concepts of variance and invariance alone provide the proper context for discussion.

The Main Experiment

This section is concerned with the analysis of the four separate data sets created by the subdivision of the major data set emerging from the experimental procedure and the effects of the independent variables in the design of the experimental structure. It will be recalled from the previous section Overview to Data Analysis that each of the four data sets was the product of a separate factorial design. Each subject was tested under each of the sixty-six experimental conditions (incorporating items from each of the "four" analogy types) with order of presentation being re-randomised for each subject. Data yielded is in the form of decision reaction times to determine which of two simultaneously presented answer options successfully completes a forced choice geometric analogy item.

Tables 5.1a - 5.1d show mean reaction times and standard deviations for task performance on each of the "four" geometric analogy types. Interpretation of the statistics in Tables 5.1a - 5.1d will be returned to in the section Stimulus Complexity and Task Performance in Chapter 6 (Discussion).

CONDITIONS, MEANS AND STANDARD DEVIATIONS (in Milliseconds)			
CUE CONDITION	NUMBER OF ELEMENTS	MEAN RESPONSE TIME (SOLUTION TIME)	STANDARD DEVIATIONS
ZERO	1	1726.62	576.34
	2	1961.57	712.19
	3	2495.03	753.97
TWO	1	1302.53	500.85
	2	1519.88	591.05
	3	1775.35	607.17

TABLE 5.1a CONDITIONS, MEANS AND STANDARD DEVIATIONS FOR THE FACTORIAL RESPONSE TIME DATA
SET RELEVANT TO DEGENERATE FORCED CHOICE GEOMETRIC ANALOGY ITEMS.

CONDITIONS, MEANS AND STANDARD DEVIATIONS (in Milliseconds)				
QUE CONDITION	NUMBER OF ELEMENTS	NUMBER OF TRANSFORMATIONS	MEAN RESPONSE TIME (SOLUTION TIME)	STANDARD DEVIATIONS
ZERO	1	1	2146.27	663.87
	1	3	2711.65	934.97
	2	1	2525.47	891.86
	2	3	3140.75	1367.50
	3	1	2855.87	909.46
	3	3	3757.40	1443.40
TWO	1	1	1517.85	551.08
	1	3	1681.17	945.02
	2	1	1786.98	725.82
	2	3	2348.98	1121.06
	3	1	2030.30	853.47
	3	3	2718.55	1250.92

TABLE 5.16 CONDITIONS, MEANS AND STANDARD DEVIATIONS FOR THE FACTORIAL RESPONSE TIME DATA SET RELEVANT TO SEMIOGENERATE (A:A::B:B) FORCED CHOICE GEOMETRIC ANALOGY ITEMS.

CONDITIONS, MEANS AND STANDARD DEVIATIONS (in Milliseconds)				
CUE CONDITION	NUMBER OF ELEMENTS	NUMBER OF TRANSFORMATIONS	MEAN RESPONSE TIME (SOLUTION TIME)	STANDARD DEVIATIONS
ZERO	1	1	2797.22	869.04
	1	3	3374.25	996.90
	2	1	3110.67	898.86
	2	3	4244.95	1391.29
	3	1	3773.43	1399.54
	3	3	5440.83	1685.67
TWO	1	1	1955.82	780.06
	1	3	2580.23	984.82
	2	1	2441.80	780.70
	2	3	3533.95	1061.70
	3	1	2798.73	918.32
	3	3	3928.90	1183.88

TABLE 5.1c CONDITIONS, MEANS AND STANDARD DEVIATIONS FOR THE FACTORIAL RESPONSE TIME DATA SET RELEVANT TO SEMIDEGENERATE (A:B::A:B) FORCED CHOICE GEOMETRIC ANALOGY ITEMS.

CONDITIONS, MEANS AND STANDARD DEVIATIONS (in Milliseconds)				
CUE CONDITION	NUMBER - OF ELEMENTS	NUMBER OF TRANSFORMATIONS	MEAN RESPONSE TIME (SOLUTION TIME)	STANDARD DEVIATIONS
ZERO	1	1	3412.46	1130.74
	1	2	4392.92	1210.40
	1	3	5216.08	1612.18
	2	1	4030.37	1149.89
	2	2	5901.05	1478.17
	2	3	7265.59	2105.21
	3	1	5059.99	1697.32
	3	2	7182.67	2219.20
	3	3	8407.23	2335.95
TWO	1	1	2241.23	976.95
	1	2	2780.52	1033.46
	1	3	3665.84	1073.50
	2	1	3040.84	909.54
	2	2	3566.18	1066.53
	2	3	4733.24	1462.36
	3	1	3407.86	1197.63
	3	2	4369.35	1469.57
	3	3	5346.80	1519.55

TABLE 5.1d CONDITIONS, MEANS AND STANDARD DEVIATIONS FOR THE FACTORIAL RESPONSE TIME
DATA SET RELEVANT TO NONDEGENERATE FORCED CHOICE GEOMETRIC ANALOGY ITEMS.

Since the reaction time data yielded in this experiment is assumed to be ratio data of normal distribution, it will be analysed by means of parametrical statistical procedures.

Each of the four data sets created by the subdivision of the major data set thus represents the results of a classic factorial repeated measures design with each subject being regarded as a block in a randomised block design (Edwards, 1960). It has been suggested that the randomised block design is well suited for the analysis of treatment effects, which are apt to be slight; as in reaction time research (Sternberg, 1966, 1969; Posner, 1978).

Tables 5.2a - 5.2d show the summary tables for the repeated measures analysis of variance procedures performed on each of the subsets of reaction time data. The tables show the source of variation, the sum of squares, the degrees of freedom, the mean square, source of error, F statistic, and corresponding probability level for each main effect and each interaction effect. A main effect represents a comparison between the means for each level of a given variable (e.g., A) averaged over the levels of each other variable (e.g., B, C, D). The interaction effect indicates the extent to which the difference between the means for each level of one variable (e.g., A1 - A3) measured at one level of an influencing variable (e.g., B1) will differ from the difference between the means for each level of the same variable (e.g., A1 - A3) at subsequent levels of the same variable (e.g., B2, B3, B4, etc.).

A non significant interaction suggests that the effects of the first variable (e.g., A) are independent of other influencing variables (e.g., B). A significant interaction, however, would suggest that the effects of the first variables (e.g., A) are not the same for the differing levels of the influencing variables (e.g., B).

In Tables 5.2a - 5.2d significance levels of $p \leq .05(*)$, $p \leq .01(**)$, $p \leq .001(***)$ and $p \leq .0001(****)$ are indicated. Interpretation of the statistics in these tables will be returned to in Chapter 6 (Discussion).

Data assessment was performed by use of the general linear model (GLM) repeated measures procedure of the SAS statistical computation package (SAS Institute, 1985). This procedure makes available the option to conduct single degree of freedom contrasts for factors parameterised in the REPEATED statement. The CONTRAST option generates contrasts between levels of a factor and a referent level, whilst a further option, PROFILE, generates contrasts between adjacent levels of the factor, thus facilitating post hoc examination of between level differences (SAS Institute, 1985).

Tables 5.3a - 5.3d show the results of fundamental single degree of freedom contrasts for key relationships within the factorial designs of the four response time data sets. The tables show the source of variation, sum of squares, degrees of freedom, the mean square, F statistic and corresponding probability level for each contrast.

In Tables 5.3a - 5.3d significance levels of $p \leq .05(*)$, $p \leq .01(**)$, $p \leq .001(***)$ and $p \leq .0001(****)$ are indicated. Interpretation of the statistics in these tables will be returned to in the section, Stimulus Complexity and Task Performance in Chapter 6 (Discussion).

Stimulus Complexity and the Experimental Task

As mentioned previously in this chapter in the section Overview to Data Analysis, the present method for assessing the effect of stimulus complexity concentrates on the level of subjective assessment.

ANOVA SUMMARY TABLE FOR DEGENERATE GEOMETRIC ANALOGY DATA SET

KEY: CUE = CUE CONDITION (ZERO vs TWO) (N = 2)

ELEMENT = NUMBER OF ELEMENTS PER TERM (N = 3)

LEVELS OF SIGNIFICANCE; $p < .05^*$; $p < .01^{**}$;

$p < .001^{***}$; $p < .0001^{****}$

SOURCE OF VARIANCE	SUM OF SQUARES	df	MEAN SQUARES	F VALUE	p
CUE	25138517.02	1	25138517.02	305.62	0.0001****
ERROR	4852561.47	59	82246.80		
ELEMENT	23576512.82	2	11838256.41	221.70	0.0001****
ERROR	6301055.17	118	53398.77		
CUE * ELEMENT	1649731.20	2	824865.60	20.42	0.0001****
ERROR	4767110.80	118	40399.24		

TABLE 5.2a ANOVA SUMMARY FOR DEGENERATE GEOMETRIC ANALOGY RESPONSE TIME DATA SET.

ANOVA SUMMARY TABLE FOR SENIDEGENERATE (A:A::B:B) GEOMETRIC ANALOGY DATA SET

KEY: CUE = CUE CONDITION (ZERO vs TWO) (N = 2)

ELEMENT = NUMBER OF ELEMENTS PER TERM (N = 3)

TRANS = NUMBER OF TRANSFORMATIONS BETWEEN TERMS (N = 2)

LEVELS OF SIGNIFICANCE; $p < .05^*$; $p < .01^{**}$; $p < .001^{***}$;
 $p < .0001^{****}$

SOURCE OF VARIANCE	SUM OF SQUARES	df	MEAN SQUARES	F VALUE	p
CUE	117783120.16	1	117783120.16	341.80	0.0001****
ERROR	20331349.74	59	344599.14		
ELEMENT	72314432.96	2	36157216.48	116.89	0.0001****
ERROR	36501727.86	118	309336.67		
TRANS	68295304.16	1	68295304.16	105.12	0.0001****
ERROR	38332371.08	59	649701.20		
CUE * ELEMENT	1405639.93	2	702819.96	8.91	0.0002***
ERROR	9308373.89	118	78884.52		
CUE * TRANS	1098320.33	1	1098320.33	11.81	0.0011***
ERROR	5485808.81	59	92979.81		
ELEMENT * TRANS	3345442.30	2	1672721.15	17.47	0.0001****
ERROR	11298480.19	118	95749.66		
CUE * ELEMENT * TRANS	239397.10	2	119698.55	1.73	0.1824
ERROR	8181826.39	118	69337.51		

TABLE 5.2b ANOVA SUMMARY FOR SENIDEGENERATE (A:A::B:B) GEOMETRIC ANALOGY RESPONSE TIME DATA SET.

ANOVA SUMMARY TABLE FOR SEMIDEGENERATE (A:B::A:B) GEOMETRIC ANALOGY DATA SET

KEY: CUE = CUE CONDITION (ZERO vs TWO) (N = 2)

ELEMENT = NUMBER OF ELEMENTS PER TERM (N = 3)

TRANS = NUMBER OF TRANSFORMATIONS BETWEEN TERMS (N = 2)

LEVELS OF SIGNIFICANCE: $p < .05^*$; $p < .01^{**}$;

$p < .001^{***}$; $p < .0001^{****}$

SOURCE OF VARIANCE	SUM OF SQUARES	df	MEAN SQUARES	F VALUE	p
CUE	151349933.16	1	151349933.16	300.63	0.0001****
ERROR	29702894.58	59	503438.89		
ELEMENT	205491196.71	2	102745598.35	234.36	0.0001****
ERROR	51732968.61	118	438414.98		
TRANS	193787364.01	1	193787364.01	263.62	0.0001****
ERROR	43371129.73	59	735103.89		
CUE * ELEMENT	10075402.78	2	5037701.38	22.21	0.0001****
ERROR	28766416.23	118	226834.03		
CUE * TRANS	1415563.36	1	1415563.36	7.25	0.0092**
ERROR	11515589.04	59	195179.47		
ELEMENT * TRANS	19522316.60	2	9811158.30	41.58	0.0001****
ERROR	27845411.39	118	239978.06		
CUE * ELEMENT * TRANS	2974163.96	2	1487081.98	10.46	0.0001****
ERROR	16779646.36	118	142200.39		

TABLE 5.2c ANOVA SUMMARY FOR SEMIDEGENERATE (A:B::A:B) GEOMETRIC ANALOGY RESPONSE TIME DATA SET.

ANOVA SUMMARY TABLE FOR NONDEGENERATE GEOMETRIC ANALOGY DATA SET

KEY: CUE = CUE CONDITION (ZERO vs TWO) (N = 2)

ELEMENT = NUMBER OF ELEMENTS PER TERM (N = 3)

TRANS = NUMBER OF TRANSFORMATIONS BETWEEN TERMS (N = 3)

LEVELS OF SIGNIFICANCE; $p < .05^*$; $p < .01^{**}$; $p < .001^{***}$; $p < .0001^{****}$

SOURCE OF VARIANCE	% OF SQUARES	df	MEAN SQUARES	F VALUE	p
CUE	2092925369.94	1	2092925369.94	539.73	0.0001****
ERROR	225657850.47	59	3876065.26		
ELEMENT	1464509526.92	2	732254763.46	312.06	0.0001****
ERROR	276887885.18	118	2150518.27		
TRANS	1808450158.32	2	904225079.16	420.47	0.0001****
ERROR	253761156.95	118	2150518.27		
CUE * ELEMENT	101830899.80	2	50915449.94	74.58	0.0001****
ERROR	80561353.11	118	682723.33		
CUE * TRANS	132734241.97	2	66367120.98	77.12	0.0001****
ERROR	101541278.52	118	860519.30		
ELEMENT * TRANS	79869528.78	4	19967382.19	22.67	0.0001****
ERROR	207834235.43	236	880653.53		
CUE * ELEMENT					
* TRANS	26415842.48	4	6603960.62	13.34	0.0001****
ERROR	116817456.85	236	494989.23		

TABLE 5.2d ANOVA SUMMARY FOR NONDEGENERATE GEOMETRIC ANALOGY RESPONSE TIME DATA SET.

ANOVA CONTRAST SUMMARY TABLE FOR DEGENERATE GEOMETRIC ANALOGY DATA SET

KEY: CDE = CUE CONDITION (ZERO vs TWO) ((N = 2)

ELEMENT = NUMBER OF ELEMENTS PER TERM (N = 3)

LEVELS OF SIGNIFICANCE: $p \leq .05^*$; $p \leq .01^{**}$; $p \leq .001^{***}$; $p \leq .0001^{****}$

CONTRAST	SUM OF SQUARES	df	MEAN SQUARES	F VALUE	p
CUE 1 vs CUE 2	150819102.15	1	150819102.15	305.62	0.0001****
ERROR	29115368.85	59	493480.82		
ELEMENT 1 vs					
ELEMENT 2	12274517.40	1	12274517.40	83.90	0.0001****
ERROR	8631478.60	59	146295.24		
ELEMENT 1 vs					
ELEMENT 3	92439611.26	1	92439611.26	281.02	0.0001****
ERROR	19407748.73	59	328944.89		
ELEMENT 2 vs					
ELEMENT 3	37344948.26	1	37344948.26	225.59	0.0001****
ERROR	9767103.73	59	165544.13		

TABLE 5.3a ANOVA CONTRASTS FOR THE DEGENERATE FACTORIAL RESPONSE TIME DATA SET.

ANOVA CONTRAST SUMMARY TABLE FOR SEMIDEGENERATE (A:A::B:B) GEOMETRIC ANALOGY DATA SET

KEY: CUE = CUE CONDITION (ZERO vs TWO) (N = 2)

ELEMENT = NUMBER OF ELEMENTS PER TERM (N = 3)

TRANS = NUMBER OF TRANSFORMATIONS BETWEEN TERMS (N = 2)

LEVELS OF SIGNIFICANCE: $p < .05^*$; $p < .01^{**}$;

$p < .001^{***}$; $p < .0001^{****}$

CONTRAST	SUM OF SQUARES	df	MEAN SQUARES	F VALUE	p
CUE 1 vs CUE 2	1413397442.01	1	1413397442.01	341.80	0.0001****
ERROR	243978198.98	59	4135189.77		
ELEMENT 1 vs					
ELEMENT 2	143267853.75	1	143267853.75	51.78	0.0001****
ERROR	163249539.25	59	2765924.39		
ELEMENT 1 vs					
ELEMENT 3	578511181.06	1	578511181.06	169.84	0.0001****
ERROR	200965808.93	59	3406200.15		
ELEMENT 2 vs					
ELEMENT 3	145994160.81	1	145994160.81	116.71	0.0001****
ERROR	73806386.18	59	1250955.69		
TRANS 1 vs					
TRANS 2	819543650.01	1	819543650.01	105.12	0.0001****
ERROR	459988452.98	59	7796414.45		

TABLE 5.36 ANOVA CONTRASTS FOR THE SEMIDEGENERATE (A:A::B:B) FACTORIAL RESPONSE TIME DATA SET.

ANOVA CONTRAST SUMMARY TABLE FOR SEMIDEGENERATE (A:B::A:B) GEOMETRIC ANALOGY DATA SET

KEY: CUE = CUE CONDITION (ZERO vs TWO) (N = 2)

ELEMENT = NUMBER OF ELEMENTS PER TERM (N = 3)

TRANS = NUMBER OF TRANSFORMATIONS BETWEEN TERMS (N = 2)

LEVELS OF SIGNIFICANCE: $p \leq .05^*$; $p \leq .01^{**}$; $p \leq .001^{***}$; $p \leq .0001^{****}$

CONTRAST	SUM OF SQUARES	df	MEAN SQUARES	F VALUE	p
CUE 1 vs CUE 2	1816199198.01	1	1816199198.01	300.63	0.0001****
ERROR	356434734.98	59	6041265.89		
ELEMENT 1 vs					
ELEMENT 2	413043843.75	1	413043843.75	214.82	0.0001****
ERROR	11340067.25	59	1922713.00		
ELEMENT 1 vs					
ELEMENT 3	1643926132.81	1	1643926132.81	311.53	0.0001****
ERROR	311336566.18	59	5278890.95		
ELEMENT 2 vs					
ELEMENT 3	408924384.06	1	408924384.06	125.08	0.0001****
ERROR	196018989.93	59	3322355.76		
TRANS 1 vs					
TRANS 2	2325448368.15	1	2325448368.15	263.62	0.0001****
ERROR	520453556.85	59	8821246.72		

TABLE 5.3c ANOVA CONTRASTS FOR THE SEMIDEGENERATE (A:B::A:B) FACTORIAL RESPONSE TIME DATA SET.

ANOVA CONTRAST SUMMARY TABLE FOR NONDEGENERATE GEOMETRIC ANALOGY DATA SET

KEY: CUE = CUE CONDITION (ZERO vs TWO) (N = 2)

ELEMENT = NUMBER OF ELEMENTS PER TERM (N = 3)

TRANS = NUMBER OF TRANSFORMATIONS BETWEEN TERMS (N = 3)

LEVELS OF SIGNIFICANCE: $p \leq .05^*$; $p \leq .01^{**}$; $p \leq .001^{***}$; $p \leq .0001^{****}$

CONTRAST	SUM OF SQUARES	df	MEAN SQUARES	F VALUE	p
CUE 1 vs CUE 2	75312913318.01	1	75312913318.01	539.73	0.0001****
ERROR	8232762616.98	59	139538349.44		
ELEMENT 1 vs					
ELEMENT 2	11189917596.15	1	11189917596.15	339.89	0.0001****
ERROR	1942611448.85	59	32922227.94		
ELEMENT 1 vs					
ELEMENT 3	34946080274.01	1	34946080274.01	413.48	0.0001****
ERROR	4986445844.98	59	84516031.27		
ELEMENT 2 vs					
ELEMENT 3	6586345099.26	1	6586345099.26	127.86	0.0001****
ERROR	3039016572.73	59	51510280.89		
TRANS 1 vs					
TRANS 2	11759776001.06	1	11759776001.06	301.67	0.0001****
ERROR	2299875966.93	59	38982643.50		
TRANS 1 vs					
TRANS 3	43378087744.26	1	43378087744.26	557.80	0.0001****
ERROR	4588180293.73	59	7765787.69		
TRANS 2 vs					
TRANS 3	9966341954.39	1	9966341954.39	261.66	0.0001****
ERROR	2247245389.60	59	38088704.90		

TABLE 5.3d ANOVA CONTRASTS FOR THE NONDEGENERATE FACTORIAL RESPONSE TIME DATA SET.

As part of the present experimental procedure, subjects were asked to rank four forced choice geometric analogy items (one from each analogy type) in order of subjective complexity. Subjects were required to do this both before and after completion of the experimental trials. It is the assumption that any pre-test/post-test differences in complexity rank order will reflect a modification of the subjects' complexity criteria caused by their exposure to, and performance on the experimental trials.

Analogy type complexity rankings thus comprised two subsets of four rank orderings of 1, 2, 3, and 4, applied to geometric analogy items A (Semidegenerate forced choice geometric analogy (format A:B::A:B)), B (Degenerate forced choice geometric analogy), C (Nondegenerate forced choice geometric analogy), and D (Semidegenerate forced choice geometric analogy (format A:A::B:B)). Total permutations are thus sixty-four $((4!) \times (4!))$.

Subjects' actual rankings are summarised in Table 5.4. Of particular significance is that only five of the potential sixty-four orderings emerge, each with the identical post-test complexity ranking set. In Table 5.4, the complexity ranking sets have been arranged in descending order of frequency and a complexity code ascribed accordingly. It must be emphasised that this ordering of ranking sets in no way reflects any 'real' or 'natural' order, and therefore the codes ascribed to these sets do not reflect any 'real' arithmetic or mathematical scale.

Approximately 82% of respondents are accounted for by their selection of code 1 or code 2 complexity sets; with 92.67% being accounted for by the first three code frequencies. Thus, the percentage of respondents accounted for by the first three code frequencies, together with the fact that only five complexity orderings emerged, suggests a non-random distribution of complexity judgements.

ITEM COMPLEXITY RANKING		FREQUENCY (N = 60) (Pa)	ASCRIBED COMPLEXITY CODE
PRE-TEST	POST-TEST		
A B C D	A B C D		
2 1 4 3	3 1 4 2	31	1
3 1 4 2	3 1 4 2	18	2
4 1 3 2	3 1 4 2	6	3
1 2 4 3	3 1 4 2	4	4
3 1 2 4	3 1 4 2	1	5
CHI-SQUARED (df = 4)		P < .01**	

TABLE 5.4 SUBJECTIVE ITEM COMPLEXITY CODING.

The resultant distribution of complexity judgements in Table 5.4 is ideally suited to analysis by the chi-squared one variable test. This test has been described as a "goodness-of-fit" technique in that it may be used to determine whether or not a significant difference exists between the observed number of cases falling into each category, and the expected number of cases, based upon the null hypothesis; which in the present context would suggest that there should be an equal number of responses within each category.

This concluded the analysis of complexity rankings per se. While the analysis conducted revealed the existence of pre-test/post-test differences in complexity rankings, the question remains as to why this should be so. It is thus preferred to retain the integrity of the judgement sets as distinct units of complexity-task association. Further subdivision and analysis of complexity code data (e.g., investigation by gender and handedness, etc.) was decided against given the lack of theoretical justification for such analysis, together with the risk of over fragmenting each of the four separate data sets (arising from the subdivision of the total data set on the basis of geometric analogy type) and thereby reducing the number of observations per cell (a procedure which was cautioned against in the section The Research Design of Sternberg's (1977) Geometric Analogy Experiment, and its Relevance to the Present Research, in Chapter 4).

However, complexity code groups were used as the basis for post hoc partitioning of the data sets pertaining to each of the "four" geometric analogy types, in order to determine whether subjective complexity code offered an associative explanation for the basis of both subjective strategy code group membership and cluster membership. These issues will be returned to in the sections Solution Strategy and the Experimental Task, and The Cluster

Analysis Procedures, later on in this chapter.

Interpretation of the statistics in Table 5.4 will be returned to in the section, Stimulus Complexity and Task Performance in Chapter 6 (Discussion).

Solution Strategy and the Experimental Task

This section is concerned with the extent to which variations in demand characteristics of forced choice geometric analogy items influences subjective solution strategy.

Previous investigations of geometric analogy solution have not appealed to subject's post hoc introspective reports for support for the interpretation of performance data. However, it is contended that this type of supplementary data may provide valuable insights into the extent to which variations in item complexity influence the order of execution of the cognitive components involved in geometric analogy solution.

As part of the procedure for the present experiment, subjects were asked to describe, in detail, the solution strategies they had adopted for each of two examples from each of the "four" types of geometric analogy included in the experimental trials. Then reason that subjects were required to provide strategy reports for two examples of each geometric analogy type was that one of the examples would comprise a minimum number of elements per term and transformations between terms (relevant to each analogy type), whereas the other example would comprise the maximum number of elements per term and transformations between terms (relevant to each analogy type). Thereby ensuring that subjects provided a strategy report for each analogy type in its most "simple" and "complex" form. It is, of course, the assumption that any variation in subjective

strategy reports for the two examples of a particular geometric analogy type will reflect a modification in the subjects' solution strategy brought about by the variation in the task demand characteristics.

Adopting the procedure described in the section Procedure in Chapter 4, each of the reported strategies was recorded and coded. Particular care was taken to ensure that any strategy report differing even slightly from those previously recorded, in terms of performance components adopted and order of component execution, was documented and assigned a separate code. To the extent that the solution strategy reports included descriptions of componential processes corresponding to those identified by Sternberg (1977) as being involved in forced choice geometric analogy solution, this was recorded.

The self reported solution strategies relevant to each of the geometric analogy types will be considered separately.

Degenerate Forced Choice Geometric Analogy Subjective Task Strategy Types

Strictly in terms of performance components adopted and order of component execution, two dominant subjective task strategy types were identified from subjects post hoc introspective reports relevant to the solution of Degenerate items with one element per term throughout. The first of these strategy types was coded 1.1a and the second 1.2a.

At this juncture it must be emphasised that neither these codes, nor those assigned to the dominant subjective task strategy types identified from the subjective strategy reports relevant to the solution of 3 element Degenerate items, both variations of Semidegenerate items, and Nondegenerate items, represent any "real" arithmetic or

mathematical properties of scale. These are purely nominal categories and any ordering of them is arbitrary.

Variations of the two dominant strategy types were identified however since these did not involve additional performance components or changes in the order of component execution, they were not coded as separate strategy types but, rather, were recorded as variations of the respective dominant strategy types.

The identical two dominant subjective task strategy types were identified from subjects post hoc introspective reports relevant to the solution of 3 element Degenerate items. The first of these strategy types were coded 3.1a and the second 3.2a.

Similarly, variations of these two dominant strategy types were identified, however, these variations were fewer in number than those identified and recorded for 1 element Degenerate item solution. Furthermore, since these variations did not involve additional performance components or changes in the order of component execution, they were not coded as separate strategy types but, rather, were recorded as variations of the respective dominant strategy types.

A detailed discussion of the dominant subjective task strategy types as well as their respective variations, identified from the subjective strategy reports relevant to the solution of 1 and 3 element Degenerate items, together with a summary of the numbers of subjects who reportedly adopted each of these strategy types, is presented in the section, Task Structure and Solution Strategy in Chapter 6 (Discussion).

Semidegenerate Forced Choice Geometric Analogy (Format A:A::B:B) Subjective Task Strategy Types

Strictly in terms of performance components adopted and order of component execution, two dominant subjective task strategy types were identified from subjects post hoc introspective reports relevant to the solution of 1 element - 1 transformation Semidegenerate items of the format A:A::B:B. The first of these strategy types was coded 1.1b and the second 1.2b.

Variations of these two dominant strategy types were identified, however, since these did not involve additional performance components or changes in the order of component execution, they were not coded as separate strategy types but, rather, were recorded as variations of the respective dominant strategy types.

The identical two dominant subjective task strategy types were identified from subjects post hoc introspective reports relevant to the solution of 3 element - 3 transformation Semidegenerate items of the format A:A::B:B. The first of these strategy types was coded 3.1b and the second 3.2b.

Similarly, variations of these two dominant strategy types were identified, however, these variations were fewer in number than those identified and recorded for 1 element - 1 transformation Semidegenerate item solution. Furthermore, since these variations did not involve additional performance components or changes in the order of component execution, they were not coded as separate strategy types but, rather, were recorded as variations of the respective dominant strategy types.

A detailed discussion of the dominant subjective task strategy types, as well as their respective variations,

identified from the subjective strategy reports relevant to the solution of 1 element - 1 transformation and 3 element - 3 transformation Semidegenerate items of the format A:A::B:B, together with a summary of the numbers of subjects who reportedly adopted each of the strategy types, is presented in the section Task Structure and Solution Strategy in Chapter 6 (Discussion).

Semidegenerate Forced Choice Geometric Analogy (Format A:B::A:B) Subjective Task Strategy Types

Strictly in terms of performance components adopted and order of component execution, two dominant subjective task strategy types were identified from subjects post hoc introspective reports relevant to the solution of 1 element - 1 transformation Semidegenerate items of the format A:B::A:B. The first of these strategy types was coded 1.1c and the second 1.2c.

Variations of these two dominant strategy types were identified, however, since these did not involve additional performance components or changes in the order of the component execution, they were not coded as separate strategy types but, rather, were recorded as variations of the respective dominant strategy types.

The identical two dominant subjective task strategy types were identified from the subjects post hoc introspective reports relevant to the solution of 3 element - 3 transformation Semidegenerate items of the format A:B::A:B. The first of these strategy types was coded 3.1c, and the second 3.2c.

Similarly, variations of these two dominant strategy types were identified, however, these variations were fewer in number than those identified and recorded for 1 element - 1 transformation Semidegenerate item solution. Furthermore,

since these variations did not involve additional performance components or changes in the order of component execution, they were not coded as separate strategy types but, rather, were recorded as variations of the respective dominant strategy types.

A detailed discussion of the dominant subjective task strategy types, as well as their respective variations, identified from the subjective strategy reports relevant to the solution of 1 element - 1 transformation and 3 element - 3 transformation Semidegenerate items of the format A:B::A:B, together with a summary of the numbers of subjects who reportedly adopted each of the strategy types, is presented in the section Task Structure and Solution Strategy in Chapter 6 (Discussion).

Nondegenerate Forced Choice Geometric Analogy Subjective Task Strategy Types

It was mentioned in the section Aims of the Present Research, in Chapter 4, that a major aim of this study was to investigate whether the cognitive component processes described by Sternberg (1977) as best defining task performance on Nondegenerate forced choice geometric analogy items are in fact as generalisable as he maintained them to be. Since Sternberg's (1977) componential models of analogical reasoning were applicable to Nondegenerate forced choice geometric analogy items comprising more than one element per term, the subjective strategy reports for the two Nondegenerate examples were obtained in different ways.

The subjective strategy reports for the solution of 1 element - 1 transformation Nondegenerate items were obtained in a similar manner to those relating to the solution of Degenerate and Semidegenerate items. In contrast, for the 3 element - 3 transformation

Nondegenerate example particular care was taken to ensure that in their subjective strategy report, subjects identified which terms had been compared with which; which term(s) had been compared with the two answer options; whether elements comprising the terms and answer options were compared individually, or as a whole, and whether the comparison of elements had been self terminating or exhaustive.

Strictly in terms of performance components adopted and order of component execution, four dominant subjective task strategy types were identified from subjects post hoc introspective reports relevant to the solution of 1 element - 1 transformation Nondegenerate items. The first of these strategy types was coded 1.1d, the second 1.2d, the third 1.3d, and the fourth 1.4d.

Variations of these four dominant strategy types were identified. Variations were apparently due to the subjects adopting either a self terminating or exhaustive application procedure. However, since these did not involve additional performance components or changes in the order of component execution, they were not coded as separate strategy types but, rather, were recorded as variations of the respective dominant strategy types.

In terms of performance components adopted, order of component execution, nature (i.e., self terminating or exhaustive) of the inference, mapping and application procedures, and mode (sequential or alternate) of option scanning, four dominant subjective task strategy types were identified from the subjective strategy reports relevant to the solution of 3 element - 3 transformation Nondegenerate items. The first of the strategy types was coded 3.1d, the second 3.2d, the third 3.3d, and the fourth 3.4d. Variations of these four dominant strategy types were not identified.

A detailed discussion of the dominant subjective task strategy types, as well as their respective variations (in the case of 1 element - 1 transformation Nondegenerate items only), identified from the subjective strategy reports relevant to the solution of 1 element - 1 transformation and 3 element - 3 transformation Nondegenerate items, together with a summary of the numbers of subjects who reportedly adopted each of the strategy types, is presented in the section Task Structure and Solution Strategy in Chapter 6 (Discussion).

Of significance for present research purposes is the degree of association between subjective complexity judgements and subjects' self-reported subjective task strategies pertaining to both examples of each analogy type.

An investigation of this association was conducted separately for both examples of each analogy type.

Degenerate Forced Choice Geometric Analogy Items

As discussed previously in this section, two dominant subjective task strategy types were identified from the subjective strategy reports relevant to 1 element Degenerate items. These were coded 1.1a and 1.2a.

These strategy codes are presented together with the subjective complexity codes in Table 5.5a. Again it must be emphasised that neither these codes nor those presented further on in the discussion of the association between subjective complexity judgements and subjects' self-reported subjective task strategies, represent any "real" arithmetic or mathematical properties of scale. These are purely nominal categories and any ordering of them is arbitrary.

ITEM COMPLEXITY CODE	FREQUENCY OF ITEM COMPLEXITY FOR EACH TASK STRATEGY			BIXONIAL (χ^2)
	1.1a	1.2a	(TOTAL)	
1	18	13	(31)	$p > .10$
2	6	12	(18)	$p > .10$
3	6	0	(6)	$p < .05^*$
4	1	3	(4)	$p > .10$
5	0	1	(1)	$p > .10$
(TOTAL)	(31)	(29)	(60)	$p > .10$

TABLE 5.5a TASK STRATEGY AND ITEM COMPLEXITY FOR 1 ELEMENT DEGENERATE GEOMETRIC ANALOGY ITEMS.

ITEM COMPLEXITY CODE	FREQUENCY OF ITEM COMPLEXITY FOR EACH TASK STRATEGY			BINOMIAL (χ^2)
	3.1a	3.2a	(TOTAL)	
1	17	14	(31)	$p > .10$
2	8	10	(18)	$p > .10$
3	6	0	(6)	$p < .05^*$
4	1	3	(4)	$p > .10$
5	1	0	(1)	$p > .10$
(TOTAL)	(33)	(27)	(60)	$p > .10$

TABLE 5.5b TASK STRATEGY AND ITEM COMPLEXITY FOR 3 ELEMENT DEGENERATE
GEOMETRIC ANALOGY ITEMS.

Given the nominal nature of the two-category rows in Table 5.5a, together with the fact that no predictions were made concerning the association between subjective task strategy and subjective complexity judgement, a binomial two-tail test was employed for data analysis purposes.

Findings set out in Table 5.5a suggest that the association between subjective task strategy (Codes 1.1a and 1.2a) and subjective complexity code 3 is strong and statistically significant ($p < .05$).

Significant levels for other codes are less compelling, whilst any semantic interpretation of codes 4 and 5 will be undermined because of the small number of total observations for each code.

As discussed previously in this section, two dominant subjective task strategy types were identified from the subjective strategy reports relevant to 3 element Degenerate items. These were coded 3.1a and 3.2a.

These strategy codes are presented together with the subjective complexity codes in Table 5.5b.

Given the nominal nature of the two-category rows in Table 5.5b, together with the fact that no predictions were made concerning the association between subjective task strategy and subjective complexity judgement, a binomial two-tail test was employed for data analysis purposes.

Findings set out in Table 5.5b suggest that the association between subjective task strategy (Codes 3.1a and 3.2a) and subjective complexity code 3 is strong and statistically significant ($p < .05$).

Significant levels for other codes are less compelling whilst any semantic interpretation of codes 4 and 5 will be

undermined because of the small number of total observations for each cell.

Semidegenerate (A:A::B:B) Forced Choice Geometric Analogy Items

As discussed previously in this section, two dominant subjective task strategy types were identified from the subjective strategy reports relevant to 1 element - 1 transformation Semidegenerate items of the format A:A::B:B. These were coded 1.1b and 1.2b.

These strategy codes are presented together with the subjective complexity codes in Table 5.5c.

Given the nature of the two-category rows in Table 5.5c, it was noted that no predictions were made concerning the relationship between subjective task strategy and subjective complexity. Hence, a binomial two-tail test was employed for all statistical purposes.

Findings set out in Table 5.5c suggest that the association between subjective task strategy (Codes 1.1b and 1.2b) and subjective complexity codes 1 ($p<.05$) and 2 ($p<.01$) is strong and statistically significant.

Significant levels for other codes are less compelling whilst any semantic interpretation of codes 4 and 5 will be undermined because of the small number of total observations for each cell.

As discussed previously in this section, two dominant subjective task strategy types were identified from the subjective strategy reports relevant to 3 element - 3 transformation Semidegenerate items of the format A:A::B:B. These were coded 3.1b and 3.2b.

ITEM COMPLEXITY CODE	FREQUENCY OF ITEM COMPLEXITY FOR EACH TASK STRATEGY			BINOMIAL (X ²)
	1.1b	1.2b	(TOTAL)	
1	22	9	(31)	p < .05*
2	17	1	(18)	p < .01**
3	6	1	(8)	p > .10
4	3	1	(4)	p > .10
5	1	0	(1)	p > .10
(TOTAL)	(48)	(12)	(60)	p < .05*

TABLE 5.5c TASK STRATEGY AND ITEM COMPLEXITY FOR 1 ELEMENT - 1 TRANSFORMATION
 SEMIOBGENERATE (A:A::B:B) GEOMETRIC ANALOGY ITEMS.

These strategy codes are presented together with the subjective complexity codes in Table 5.5d.

Given the nominal nature of the two-category rows in Table 5.5d, together with the fact that no predictions were made concerning the association between subjective task strategy and subjective complexity judgement, a binomial two-tail test was employed for data analysis purposes.

Findings set out in Table 5.5d suggest that the association between subjective task strategy (Codes 3.1b and 3.2b) and subjective complexity codes 1 ($p < .05$), 2 ($p < .01$) and 3 ($p < .05$) is strong and statistically significant.

Any semantic interpretation of codes 4 and 5 will be undermined because of the small number of total observations for each cell.

Semidegenerate (A:B::A:B) Forced Choice Geometric Analogy Items

As discussed previously in this section, two dominant subjective task strategy types were identified from the subjective strategy reports relevant to 1 element - 1 transformation Semidegenerate items of the format A:B::A:B. These were coded 1.1c and 1.2c.

These strategy codes are presented together with the subjective complexity codes in Table 5.5e.

Given the nominal nature of the two-category rows in Table 5.5e, together with the fact that no predictions were made concerning the association between subjective task strategy and subjective complexity judgement, a binomial two-tail test was employed for data analysis purposes.

Findings set out in Table 5.5e suggest that the association

ITEM COMPLEXITY CODE	FREQUENCY OF ITEM COMPLEXITY FOR EACH TASK STRATEGY			BINOMIAL (χ^2)
	3.1b	3.2b	(TOTAL)	
1	24	7	(31)	$p < .05^*$
2	18	0	(18)	$p < .01^{**}$
3	6	0	(6)	$p < .05^*$
4	3	1	(4)	$p > .10$
5	1	0	(1)	$p > .10$
(TOTAL)	(52)	(8)	(60)	$p < .05^*$

TABLE 5.5d TASK STRATEGY AND ITEM COMPLEXITY FOR 3 ELEMENT - 3 TRANSFORMATION
SENDEGENERATE (A:A::B:B) GEOMETRIC ANALOGY ITEMS.

ITEM COMPLEXITY CODE	FREQUENCY OF ITEM COMPLEXITY FOR EACH TASK STRATEGY			BINOMIAL (χ^2)
	1.1.a	1.2.a	(TOTAL)	
1	12	19	(31)	p > .10
2	11	7	(18)	p > .10
3	6	0	(6)	p < .05*
4	0	4	(4)	p > .10
5	0	1	(1)	p > .10
(TOTAL)	(29)	(31)	(60)	p > .10

TABLE 5.5c TASK STRATEGY AND ITEM COMPLEXITY FOR 1 ELEMENT - 1 TRANSFORMATION
SEMIDEGENERATE (A:B::A:B) GEOMETRIC ANALOGY ITEMS.

between subjective task strategy (Codes 1.1c and 1.2c) and subjective complexity Code 3 is strong and statistically significant ($p < .05$).

Significant levels for other codes are less compelling whilst any semantic interpretation of Codes 4 and 5 will be undermined because of the small number of total observations for each cell.

As discussed previously in this section, two dominant subjective task strategy types were identified from the subjective strategy reports relevant to 3 element - 3 transformation Semidegenerate items of the format A:B::A:B. These were coded 3.1c and 3.2c.

These strategy codes are presented together with the subjective complexity codes in Table 5.5f.

Given the nominal nature of the two-category rows in Table 5.5f, together with the fact that no predictions were made concerning the association between subjective task strategy and subjective complexity judgement, a binomial two-tail test was employed for data analysis purposes.

Findings set out in Table 5.5f fail to suggest any statistically significant association between subjective task strategy (Codes 3.1c and 3.2c) and subjective complexity code.

Nondegenerate Forced Choice Geometric Analogy Items

As discussed previously in this section, four dominant subjective task strategy types were identified from the subjective strategy reports relevant to 1 element - 1 transformation Nondegenerate items. These were coded 1.1d, 1.2d, 1.3d and 1.4d.

ITEM COMPLEXITY CODE	FREQUENCY OF ITEM COMPLEXITY FOR EACH TASK STRATEGY			BINOMIAL (χ^2)
	3.1c	3.2g	(TOTAL)	
1	10	21	(31)	$p > .10$
2	10	8	(18)	$p > .10$
3	3	3	(6)	$p > .10$
4	0	4	(4)	$p > .10$
5	0	1	(1)	$p > .10$
(TOTAL)	(23)	(37)	(60)	$p < .10$

TABLE S.5F TASK STRATEGY AND ITEM COMPLEXITY FOR 3 ELEMENT - 3 TRANSFORMATION
SEMIDEGENERATE (A:B::A:B) GEOMETRIC ANALOGY ITEMS.

ITEM COMPLEXITY CODE	FREQUENCY OF ITEM COMPLEXITY FOR EACH TASK STRATEGY				(TOTAL)	CHI- SQUARED (df = 3)
	1.1d	1.2d	1.3d	1.4d		
1	15	5	9	2	(31)	$p < .01^{**}$
2	15	1	1	1	(18)	$p < .01^{**}$
3	5	1	0	0	(6)	$p < .05^{*}$
4	1	1	2	0	(4)	$p > .10$
5	1	0	0	0	(1)	$p > .10$
(TOTAL)	(37)	(8)	(12)	(3)	(60)	$p < .01^{**}$

TABLE 5.5g TASK STRATEGY AND ITEM COMPLEXITY FOR 1 ELEMENT - 1 TRANSFORMATION
NONDEGENERATE GEOMETRIC ANALOGY ITEMS.

These strategy codes are presented together with the subjective complexity codes in Table 5.5g.

Given the nominal nature of each row in Table 5.5g, a chi-squared one-sample test (Siegel, 1956) were employed for data analysis purposes.

Findings set out in Table 5.5g suggest that the association between subjective task strategy (Codes 1.1d, 1.2d, 1.3d and 1.4d) and subjective complexity codes 1 ($p < .01$), 2 ($p < .01$) and 3 ($p < .05$) is strong and statistically significant.

Any semantic interpretation of codes 4 and 5 will be undermined because of the small number of total observations for each code.

As discussed previously in this section, four dominant subjective task strategy types were identified from the subjective strategy reports relevant to 3 element - 3 transformation Nondegenerate items. These were coded 3.1d, 3.2d, 3.3d and 3.4d.

These strategy codes are presented together with the subjective complexity codes in Table 5.5h.

Given the nominal nature of each row in Table 5.5h, a chi-squared one-sample test (Siegel, 1956) was employed for data analysis purposes.

Findings set out in Table 5.5h suggest that the association between subjective task strategy (Codes 3.1d, 3.2d, 3.3d and 3.4d) and subjective complexity code 2 ($p < .05$) is strong and statistically significant.

Significant levels for other codes are less compelling whilst any semantic interpretation of codes 4 and 5 will be

ITEM COMPLEXITY CODE	FREQUENCY OF ITEM COMPLEXITY FOR EACH TASK STRATEGY					CHI- SQUARED (df = 3)
	3.1d	3.2d	3.3d	3.4d	(TOTAL)	
1	11	7	11	2	(31)	p < .10
2	10	1	5	2	(18)	p < .05*
3	1	1	4	0	(6)	p > .10
4	0	2	2	0	(4)	p > .10
5	0	0	1	0	(1)	p > .10
(TOTAL)	(22)	(11)	(23)	(4)	(60)	p < .01**

TABLE 5.5h TASK STRATEGY AND ITEM COMPLEXITY FOR 3 ELEMENT - 3 TRANSFORMATION
NONDEGENERATE GEOMETRIC ANALOGY ITEMS.

undermined because of the small number of total observations for each cell.

Further subdivision and analysis of complexity code and task strategy association (e.g., investigation by gender and handedness, etc.), was decided against on the same joint grounds advanced for not subdividing the complexity code data (see the section Stimulus Complexity and the Experimental Task earlier in this chapter), viz:- that the lack of theoretical justification for such analysis would most likely yield equivocal findings, and that further cross tabulations would serve to over fragment each of the four separate data sets (arising from the subdivision of the total data set on the basis of geometric analogy type), thereby reducing the number of observations per cell (a procedure cautioned against in the section The Research Design of Sternberg's (1977) Geometric Analogy Experiment and its Relevance to the Present Research, in Chapter 4).

The Cluster Analysis Procedures

The final issue to be investigated through analysis of the experimental data is the extent to which clustering techniques can provide insight into the nature of the processes that underlie performance on forced choice geometric analogy tasks, how individuals differ in terms of these processes, and the task characteristics that influence these performance differences.

As noted in the section Conceptual and Methodological Issues in Chapter 4, analogy solution and geometric analogy solution in particular, despite its long history of inclusion on tests of aptitude and intelligence, has only recently become the subject of empirical and theoretical analysis. Most existing research in the area hinges upon Sternberg's (1977) attempt to specify the components of

performance in this task, and has been constrained by theoretical and methodological weaknesses. In particular the repetitive use of a small number of hypothesis sophisticated subjects, the reduction of subjects' solution strategies into a componential model on the basis of a priori assumptions of sequential component execution, etc.

It has been the contention of the present research that theories of cognitive individual differences based upon observations of small numbers of subjects in repeated measures designs involving large numbers of independent variables are not particularly compelling, and an alternative approach has been proposed through the application of an appropriate clustering algorithm on a substantial data set.

Despite the practical and mathematical uncertainties of cluster analysis discussed in detail in the section Individual Differences and Cluster Analysis in Chapter 4, it is contended that its use in the investigation of mental workload influences on geometric analogy task performance could prove most insightful.

At this juncture it is appropriate to offer a full explanation of the clustering technique applied in the present data analysis.

Of primary concern is the nature of the data to be analysed. Clustering is only applicable to the four response time data matrices arising from the subdivision of the total response time data matrix produced by the experimental procedure. Application to larger data sets including, for example, coding for gender, handedness, subjective strategy, subjective item complexity, etc., will yield cluster solutions based on those variables; they are likely to impose an immediate and strong basis to form a perfect partitioning of the data sets. Rather, the

clustering should be confined to the response time data alone.

The Degenerate response time data matrix comprises 9 factorial design conditions, each of the Semidegenerate reaction time data matrices comprises 18 factorial design conditions, and the Nondegenerate analogy reaction time data matrix comprises 54 factorial design conditions. Each factorial design condition has 60 response times. The objects to be clustered (subjects) thus each have a perfectly standardised profile of ratio data observations for each analogy type. No manipulation of this data is necessary to form standardised profiles and this will eliminate the occurrence of distortions that may alter the true relationship between variables and within each data matrix.

It is, of course, an assumption that variation, structure and pattern exist within each data matrix, and that this variation, structure and pattern is the result of the experimental independent variables. Furthermore, it is assumed that this structure and pattern will emerge within each of the clustering solutions.

A related assumption is that there is interindividual difference contained within the structure and pattern of each data set but, because the factorial design of the matrices facilitates the detection of strategy change within an object's profile, no undetectable intraindividual differences will occur. In practice this assumption means that no individual will acquire membership of more than one cluster or group, per cluster solution.

A final observation about each of the four data matrices is that they consist exclusively of ratio data. This, coupled with the standardised nature of scores for each variable, would warrant the selection of a clustering algorithm with

minimal similarity transformations, such as a distance metric.

All clustering techniques involve a large amount of arithmetic, and would not be feasible without the aid of modern statistical computational packages. The SAS statistical computational package (SAS Institute, 1985) contains fourteen different clustering methods, eleven of these being contained as options within a CLUSTER procedure. Each of these fourteen methods differs in its approach to providing a cluster solution, and each has theoretical and methodological weaknesses and advantages.

Additional to the CLUSTER procedure, the procedure FASTCLUS, also contained in the SAS package (SAS Institute, 1985), identifies disjoint clusters (a cluster solution in which each object is assigned to only one cluster) using MacQueen's (1967) k-means algorithm applied to co-ordinate data. FASTCLUS is especially suitable for large data sets.

The FASTCLUS procedure selects a set of data points r cluster seeds, one from each variable for each cluster. These are used as "best initial guess" to the means of the clusters (Anderberg, 1973). Each observation in turn is then assigned to the nearest seed to form temporary clusters; the seeds being replaced by the cluster mean. This is repeated until no further changes occur in the clusters.

This initialisation method guarantees that if clusters exist in the data set such that all distances between observations in the same cluster are less than all distances between observations in different clusters, and if the correct number of clusters is specified, the FASTCLUS algorithm always finds the clustering solution without iterating (Milligan, 1980).

When executing the FASTCLUS procedure it is necessary to specify the maximum number of clusters (and thereby the maximum number of initial cluster seeds) using a MAXC = n option. In contrast to the CLUSTER procedure, where only one run produces results from one cluster to n clusters, the FASTCLUS procedure requires a separate run for each number of clusters.

The FASTCLUS procedure option, DRIFT, ensures that after initial seed selection each observation is assigned to the cluster with the nearest seed. Thereafter, the seed of the cluster to which an observation is assigned is recalculated as the mean of the observations currently assigned to that cluster, so that the cluster seeds "drift" about within the cluster rather than remain fixed for the duration of a pass through the data.

A further option, DELETE = n, deletes cluster seeds to which n or fewer observations have been assigned after the DRIFT option has been completed, and after each iteration. However, cluster seeds are not deleted after final cluster assignment of observation, thus facilitating the emergence of clusters with less than n should the dissimilarity from other clusters be strong enough.

To establish suitable stringency levels the FASTCLUS procedure was run on all four response time data sets, with MAXC = n, DRIFT and DELETE = 5 options, specifying each maximum cluster from two through to twelve. All other FASTCLUS procedure options were left at default including the default of one iteration. This produced a one cluster solution for each of the four data sets.

These specifications were apparently too stringent, since a one cluster solution fails to reveal variation, structure or pattern within each data matrix. Identical one cluster solutions were produced when setting DELETE = 4 and

specifying each maximum cluster from two through to fifteen. These specifications also appeared to be too stringent.

For present research purposes then, DELETE = 3 was specified, thus ensuring stringent cluster assignment whilst still allowing formation of small groupings (3 = 5% of N = 60) should a cluster solution not heavily influenced by single response outliers influence their generation.

As an additional step for stringency the maximum number of iterations was set at ten to more fully allow the DRIFT and DELETE = options to function.

Using these specifications, the clustering procedure was run on each reaction time data set, for maximum initial seed selections for each value between two and twenty.

The resultant cluster solutions will be considered separately for each analogy type.

Degenerate Forced Choice Geometric Analogy Cluster Solutions

The specification of two through to eight maximum clusters produced a one cluster solution. However, the specification of nine and fifteen maximum clusters proved to be critical. Table 5.6a shows brief summary statistics for the cluster solutions generated at these specified maximum cluster levels. For each cluster solution the table shows the FASTCLUS MAXC = level responsible for generating the solution, the frequency of observations within the cluster, the root mean square across variables of the cluster standard deviations (equal to the root mean square distance between observations in the cluster), the maximum distance from the final cluster seed to any observation within the cluster, the number of the cluster

CRITICAL CLUSTER SOLUTIONS WITH DELETE- OPTION FOR DEGENERATE GEOMETRIC ANALOGY ITEMS							
MAXC =	CLUSTER NUMBER	FREQUENCY	RMS STD DEVIATION	MAXIMUM DISTANCE FROM SEED TO OBSERVATION	NEAREST CLUSTER	CENTROID DISTANCE	PSEUDO F
9	1	3	677.0	4826.0	2	6107.0	108.48
	2	57	208.9	1600.0	1	6107.0	
15	1	4	1122.8	5044.0	2	4946.6	181.68
	2	58	281.7	1513.3	1	4946.5	

TABLE 5.6a SUMMARY OF CRITICAL CLUSTER SOLUTIONS WITH DELETE- OPTION FOR DEGENERATE GEOMETRIC ANALOGY ITEMS.

with mean closest to the mean of the current cluster, the distance between the centroids (means) of the current cluster and the nearest other cluster, and the pseudo F statistic for the cluster solution.

The pseudo F statistic $(Rsq/(c-1))/(1 - Rsq/(n-c))$ (where Rsq is the observed Rsquared, c is the number of clusters, and n is the number of observations) was suggested by Calinski and Harabasz (1974) as a measure that may be applied in estimation of the number of clusters. A basic use is to look for peaks in the pseudo F statistic as an indication of best estimates emerging from a solution (Cooper and Milligan, 1984).

The two cluster solution produced by specifying a maximum of nine initial cluster seeds is exactly replicated when a maximum of ten through to fourteen initial cluster seeds are specified. However the specification of a maximum of fifteen initial cluster seeds produces a new two cluster solution through an adjustment of cluster membership with one member moving from cluster 2 to cluster 1 and the effects of marginally reducing the centroid distance (closeness of the two cluster centres), modifying variable means (general lowering) and standard deviations (generally bigger). This cluster solution is exactly retained for each specified maximum cluster level from sixteen to twenty, thus demonstrating the stability of the two cluster solution.

Having accepted a two cluster solution, it is necessary to establish which two cluster solution is the most representative of the solutions provided by the clustering algorithm; that generated by specifying a maximum of nine initial cluster seeds or that generated by specifying fifteen. There is strong similarity between these two solutions. The solution emerging from the specification of a maximum of fifteen initial cluster seeds is, of course,

the more robust and enduring solution, given the additional initial cluster seeds. A more objective criterion for deciding the "best" solution lies with the magnitude of the pseudo F statistic (Calinski and Harabasz, 1974) which Milligan and Cooper (1983) and Cooper and Milligan (1984) report to be one of the three criteria out of a potential thirty, for estimating the number of population clusters using four hierarchical clustering methods in simulation studies. It must be emphasised that this criterion will be adopted when deciding on the "best" cluster solutions for both of the Semidegenerate, and the Nondegenerate data sets.

In the present case the two cluster solution provided by specifying a maximum of fifteen initial cluster seeds yields the higher pseudo F statistic (181.68) and on this basis will be adopted as the most representative of the solutions provided by the clustering algorithm.

Upon closer examination of this solution it becomes apparent that a clear basis exists for a two cluster structure within the data set. Table 5.6b sets out cluster means and cluster solutions for both clusters. Most noticeable is the fundamental distinction between a "fast" cluster (2) and a "slow" cluster (1). Albeit that the partition is exploit, there is little other evidence to indicate whether the two clusters are merely a reflection of differences in the speed of task performance alone or whether there are other, deeper, information processing constraints supporting this cluster structure. Since the basis of the clustering algorithm is to maximise separation between clusters, the assumptions of usual significance tests, such as analysis of variance F tests, are drastically violated (Aldenderfer and Blashfield, 1984). Consequently, they are not valid for testing between clusters to establish a basis for differences.

CLUSTER 1 MEANS AND STANDARD DEVIATIONS BY CONDITION

KEY: OS0 = DEGENERATE SOLUTION TIME ZERO CUE CONDITION.
 OS2 = DEGENERATE SOLUTION TIME TWO CUE CONDITION.
 1E = 1 ELEMENT, 2E = 2 ELEMENTS, 3E = 3 ELEMENTS.

CONDITION	OS01E	OS02E	OS03E	OS21E	OS22E	OS23E
MEAN	3437.50	4168.75	4806.25	2796.50	3194.25	3547.00
SD	1070.87	1306.71	925.81	988.93	1208.73	1188.87

TABLE 5.6b CLUSTER MEANS AND STANDARD DEVIATIONS FOR EACH DEGENERATE FACTORIAL CONDITION.

CLUSTER 2 MEANS AND STANDARD DEVIATIONS BY CONDITION

KEY: DS0 = DEGENERATE SOLUTION TIME ZERO CUE CONDITION.

DS2 = DEGENERATE SOLUTION TIME TWO CUE CONDITION.

1E = 1 ELEMENT, 2E = 2 ELEMENTS, 3E = 3 ELEMENTS.

CONDITION	DS01E	DS02E	DS03E	DS21E	DS22E	DS23E
MEAN	1604.41	1803.91	2329.95	1195.82	1400.29	1648.80
SD	256.31	267.14	383.17	204.54	276.77	271.65

TABLE 5.66 CLUSTER MEANS AND STANDARD DEVIATIONS FOR EACH DEGENERATE FACTORIAL CONDITION.

However, as noted previously in this section, it is an essential assumption that emergent cluster solutions are directly related to the pattern or structure inherent in each data set, which in turn is the direct result of the effects of the independent variables inherent in the factorial experimental designs. On these grounds it follows that any attempts to inspect for an information processing foundation to the cluster solution should appeal to these experimental variables; their main effects, contrasts, and interactions.

Given that direct between group analysis with conventional statistics is inadvisable and invalid, for the reasons previously outlined, the problem of examining and interpreting the foundation for cluster difference appears to be less quantitative and more qualitative on a within cluster basis. Consequently analysis is restricted to qualitative within cluster analysis using the same ANOVA, CONTRAST and PROFILE procedures employed in the analysis of each of the four reaction time data sets (see the section The Main Experiment earlier in this chapter for a discussion of these procedures).

Accordingly, Table 5.6c sets out a summary of ANOVA profiles for the two cluster membership defined by the adopted cluster solution, whilst Table 5.6d contains a summary of CONTRAST and PROFILE outputs for each cluster.

Interpretation of the statistics in Tables 5.6a - 5.6d will be returned in the section Clusters and Stimulus Properties in Chapter 6 (Discussion).

Semidegenerate (A:A::B:B) Forced Choice Geometric Analogy Cluster Solutions

The specification of two through to eleven maximum clusters produced a one cluster solution. However, the

ANOVA PROFILES BY CLUSTER FOR DEGENERATE GEOMETRIC ANALOGY DATA SET

KEY: CUE = CUE CONDITION (ZERO vs TWO) (N = 2)
 ELEMENT = NUMBER OF ELEMENTS PER TERM (N = 3)

LEVELS OF SIGNIFICANCE: $p \leq .05^*$; $p \leq .01^{**}$;
 $p \leq .001^{***}$; $p \leq .0001^{****}$

SOURCE OF VARIANCE	CLUSTER 1 (p)	CLUSTER 2 (p)
CUE	0.0028**	0.0001****
ELEMENT	0.0007***	0.0001****
CUE * ELEMENT	0.0230*	0.0001****

TABLE 5.6a ANOVA PROFILES BY CLUSTERS FOR DEGENERATE GEOMETRIC ANALOGY RESPONSE TIME DATA SET.

ANOVA CONTRAST BY CLUSTER FOR DEGENERATE GEOMETRIC ANALOGY DATA SET

KEY: CUE = CUE CONDITION (ZERO vs TWO) (N = 2)

ELEMENT = NUMBER OF ELEMENTS PER TERM (N = 3)

LEVELS OF SIGNIFICANCE: $p \leq .05^*$; $p \leq .01^{**}$;

$p \leq .001^{***}$; $p \leq .0001^{****}$

SOURCE OF VARIANCE	CLUSTER 1 (p)	CLUSTER 2 (p)
CUE 1 vs CUE 2	0.0028**	0.0001****
ELEMENT 1 vs ELEMENT 2	0.0476*	0.0001****
ELEMENT 1 vs ELEMENT 3	0.0032**	0.0001****
ELEMENT 2 vs ELEMENT 3	0.0177*	0.0001****

TABLE 5.6d ANOVA CONTRASTS BY CLUSTERS FOR DEGENERATE GEOMETRIC ANALOGY RESPONSE
TIME DATA SET.

specification of twelve maximum clusters proved to be critical. Table 5.7a shows brief summary statistics for the cluster solutions generated at this specified maximum cluster level.

The three cluster solution produced by specifying a maximum of twelve initial cluster seeds is exactly retained for each specified maximum cluster level from thirteen to twenty.

The three cluster solution produced by specifying a maximum of twelve initial cluster seeds yields a pseudo F statistics of 153.16.

Table 5.7b sets out cluster means and cluster solutions for each of the three clusters. Again, albeit that the partitioning is explicit, there is little other evidence to indicate whether the three clusters are merely a reflection of the differences in the speed of task performance alone, or whether there are other, deeper information processing constraints supporting this cluster structure.

Given that direct between group analysis with conventional statistics is inadvisable and invalid, for the reasons previously outlined, once again the problem of examining and interpreting the foundation for cluster difference appears to be less quantitative and more qualitative on a within cluster basis. Consequently analysis is restricted to qualitative within cluster analysis using the same ANOVA, CONTRAST and PROFILE procedures as those applied to the Degenerate data set partitioned on cluster membership of each object (subject).

Accordingly, Table 5.7c sets out a summary of ANOVA profiles for the three cluster membership defined by the adopted cluster solution, whilst Table 5.7d contains a summary of CONTRAST and PROFILE outputs for each cluster.

CRITICAL CLUSTER SOLUTIONS WITH DELETE= OPTION FOR SEMIDEGENERATE (A:A::B:B) GEOMETRIC ANALOGY ITEMS							
MAXC =	CLUSTER NUMBER	FREQUENCY	RMS STD DEVIATION	MAXIMUM DISTANCE FROM SEED TO OBSERVATION	NEAREST CLUSTER	CENTROID DISTANCE	PSEUDO F
12	1	33	438.5	3878.5	2	1992.3	153.16
	2	24	287.8	1731.2	1	1992.3	
	3	3	771.0	6457.1	1	12966.3	

TABLE 5-7a SUMMARY OF CRITICAL CLUSTER SOLUTIONS WITH DELETE= OPTION FOR SEMIDEGENERATE (A:A::B:B) GEOMETRIC ANALOGY ITEMS.

CLUSTER 1 MEANS AND STANDARD DEVIATIONS BY CONDITION

KEY: SA0 = SEMIDEGENERATE (A:A::B:B) SOLUTION TIME ZERO CUE CONDITION.
 SA2 = SEMIDEGENERATE (A:A::B:B) SOLUTION TIME TWO CUE CONDITION.
 1E = 1 ELEMENT, 2E = ELEMENTS, 3E = ELEMENTS.
 1T = 1 TRANSFORMATION, 3T = 3 TRANSFORMATIONS.

CONDITION	SA01E1T	SA01E3T	SA02E1T	SA02E3T	SA03E1T	SA03E3T
MEAN	2194.21	2771.36	2572.27	3150.79	3032.09	3634.21
SD	322.80	453.16	387.69	519.53	499.02	511.03
CONDITION	SA21E1T	SA21E3T	SA22E1T	SA22E3T	SA23E1T	SA23E3T
MEAN	1556.27	1697.61	1817.97	2279.27	2097.09	2730.56
SD	313.43	434.42	332.54	364.19	442.67	479.06

TABLE 5.7b CLUSTER MEANS AND STANDARD DEVIATIONS FOR EACH SEMIDEGENERATE (A:A::B:B) FACTORIAL CONDITION.

CLUSTER 2 MEANS AND STANDARDS DEVIATIONS BY CONDITION

KFY: SA0 - SEMIDEGENERATE (A:A::B:B) SOLUTION TIME ZERO CUE CONDITION.

SA2 - SEMIDEGENERATE (A:A::B:B) SOLUTION TIME TWO CUE CONDITION.

1E = 1 ELEMENT, 2E = ELEMENTS, 3E = ELEMENTS.

1T = 1 TRANSFORMATION, 3T = 3 TRANSFORMATIONS.

CONDITION	SA01E1T	SA01E3T	SA02E1T	SA02E3T	SA03E1T	SA03E3T
MEAN	1785.63	2203.96	2039.75	2450.21	2236.25	2960.33
SD	279.58	365.43	215.33	280.31	342.73	345.35
CONDITION	SA21E1T	SA21E3T	SA22E1T	SA22E3T	SA23E1T	SA23E3T
MEAN	1230.50	1490.67	1411.50	1872.38	1557.08	2086.50
SD	257.99	300.56	264.45	167.66	310.25	260.69

TABLE 5.7b CLUSTER MEANS AND STANDARD DEVIATIONS FOR EACH SEMIDEGENERATE (A:A::B:B) FACTORIAL CONDITION.

CLUSTER 3 MEANS AND STANDARD DEVIATIONS BY CONDITION

KEY: SA0 = SEMIDEGENERATE (A:A::B:B) SOLUTION TIME ZERO CUE CONDITION.
 SA2 = SEMIDEGENERATE (A:A::B:B) SOLUTION TIME TWO CUE CONDITION.
 1E = 1 ELEMENT, 2E = ELEMENTS, 3E = ELEMENTS.
 1T = 1 TRANSFORMATION, 3T = 3 TRANSFORMATIONS.

CONDITION	SA01E1T	SA01E3T	SA02E1T	SA02E3T	SA03E1T	SA03E3T
MEAN	4504.00	6116.33	5896.33	8554.67	5873.33	9289.00
SD	705.89	653.49	784.40	505.11	536.37	704.22

CONDITION	SA21E1T	SA21E3T	SA22E1T	SA22E3T	SA23E1T	SA23E3T
MEAN	3394.00	4824.33	4450.00	6928.67	6081.33	7642.67
SD	571.34	1476.01	788.77	602.53	722.33	735.96

TABLE 5.7b CLUSTER MEANS AND STANDARD DEVIATIONS FOR EACH SEMIDEGENERATE (A:A::B:B)
 FACTORIAL CONDITION.

Interpretation of the statistics in Tables 5.7a - 5.7d will be returned in the section Clusters and Stimulus Properties in Chapter 6 (Discussion).

Semidegenerate (A:B::A:B) Forced Choice Geometric Analogy Cluster Solutions

The specification of two through to seven maximum clusters produced a one cluster solution. However, the specification of eight and ten maximum clusters proved to be critical. Table 5.8a shows brief summary statistics for the cluster solutions generated at these specified maximum cluster levels.

The two cluster solution generated by specifying a maximum of eight initial cluster seeds is exactly replicated when a maximum of nine initial seeds is specified. However, the specification of a maximum of ten initial cluster seeds leads to a new two cluster solution through an adjustment of cluster membership, with three members moving from cluster 1 to cluster 2 and the effects of increasing the centroid distance, modifying variable means (general increase) and standard deviations (generally smaller). The two cluster solution generated by specifying a maximum of ten initial cluster seeds is exactly replicated through the two solutions obtained by specifying a maximum of eleven and twelve initial cluster seeds.

However, with the specification of a further, thirteenth, maximum initial cluster seed, the algorithm re-establishes exactly the same solution obtained by the specification of a maximum of eight initial cluster seeds, and this is retained exactly through the solutions produced by specifying a maximum of fourteen through to twenty initial cluster seeds.

ANOVA PROFILES BY CLUSTER FOR SEMIDEGENERATE (A:A::B:B) GEOMETRIC ANALOGY DATA SET

KEY: CUE = CUE CONDITION (ZERO vs TWO) (N = 2)

ELEMENT = NUMBER OF ELEMENTS PER TERM (N = 3)

TRANS = NUMBER OF TRANSFORMATIONS BETWEEN TERMS (N = 2)

LEVELS OF SIGNIFICANCE: $p \leq .05^*$; $p \leq .01^{**}$;
 $p \leq .001^{***}$; $p \leq .0001^{****}$

SOURCE OF VARIANCE	CLUSTER 1 (p)	CLUSTER 2 (p)	CLUSTER 3 (p)
CUE	0.0001****	0.0001****	0.0178*
ELEMENT	0.0001****	0.0001****	0.0089**
TRANS	0.0001****	0.0001****	0.0039**
CUE * ELEMENT	0.0026**	0.0237*	0.3349
CUE * TRANS	0.0138*	0.1288	0.0701
ELEMENT * TRANS	0.0025**	0.0004***	0.0233*
CUE * ELEMENT * TRANS	0.8920	0.1114	0.6024

TABLE 5.7c ANOVA PROFILES BY CLUSTERS FOR SEMIDEGENERATE (A:A::B:B) GEOMETRIC ANALOGY
 RESPONSE TIME DATA SET.

ANOVA CONTRAST BY CLUSTER FOR SEMIDEGENERATE (A:B::B:B) GEOMETRIC ANALOGY
DATA SET

KEY ; CUE = CUE CONDITION (ZERO vs TWO) (N = 2)

ELEMENT = NUMBER OF ELEMENTS PER TERM (N = 3)

TRANS = NUMBER OF TRANSFORMATIONS BETWEEN TERMS (N = 2)

LEVELS OF SIGNIFICANCE; $p < .05^*$; $p < .01^{**}$;

$p < .001^{***}$; $p < .0001^{****}$

SOURCE OF VARIANCE	CLUSTER 1 (p)	CLUSTER 2 (p)	CLUSTER 3 (p)
CUE 1 vs CUE 2	0.0001****	0.0001****	0.0175*
ELEMENT 1 vs ELEMENT 2	0.0001****	0.0001****	0.0793
ELEMENT 1 vs ELEMENT 3	0.0001****	0.0001****	0.0045**
ELEMENT 2 vs ELEMENT 3	0.0001****	0.0001****	0.3050
TRANS 1 vs TRANS 2	0.0001****	0.0001****	0.0039**

TABLE 5.7d ANOVA CONTRASTS BY CLUSTERS FOR SEMIDEGENERATE (A:A::B:B) GEOMETRIC ANALOGY
RESPONSE TIME DATA SET.

CRITICAL CLUSTER SOLUTIONS WITH DELETE= .OPTION FOR SEXIDEGENERATE (A:B::A:B) GEOMETRIC ANALOGY ITEMS							
MAXC =	CLUSTER NUMBER	FREQUENCY	RMS STD DEVIATION	MAXIMUM DISTANCE FROM SEED TO OBSERVATION	NEAREST CLUSTER	CENTROID DISTANCE	PSEUDO F
8	1	5	1819.2	9865.7	2	9077.3	59.68
	2	54	606.4	4231.5	1	9077.3	
10	1	3	1133.9	7379.8	2	13696.2	90.46
	2	57	681.3	5787.5	1	13696.2	

TABLE 5.8a SUMMARY OF CRITICAL CLUSTER SOLUTIONS WITH DELETE= OPTION FOR SEXIDEGENERATE (A:B::A:B) GEOMETRIC ANALOGY ITEMS.

Having accepted a two cluster solution, it was once again necessary to establish which two cluster solution was the most representative of the solutions produced by the clustering algorithm; that generated by specifying a maximum of eight initial cluster seeds or that generated by specifying ten. Again, magnitude of the pseudo F statistic was adopted as the criterion for deciding the "best" solution. In this case the two cluster solution provided by specifying a maximum of ten initial cluster seeds yields the higher pseudo F statistic (90.46) and on this basis will be adopted as the most representative of the solutions provided by the clustering algorithm.

Upon closer examination of this solution it becomes apparent that a very clear basis exists for a two cluster structure within the data set. Table 5.8b sets out cluster means and cluster solutions for both clusters. As was the case when examining the Degenerate forced choice geometric analogy two cluster solution, most noticeable is the fundamental distinction between a "fast" cluster (2) and a "slow" cluster (1). Again, albeit that the partition is explicit, there is little other evidence to indicate whether the two clusters are merely a reflection of differences in the speed of task performance alone, or whether there are other, deeper information processing constraints supporting this cluster structure.

Given that direct between group analysis with conventional statistics is inadvisable and invalid, for the reasons previously outlined, once again, the problem of examining and interpreting the foundation for cluster solution appears to be less quantitative and more qualitative on a within cluster basis. Consequently, the strategy will again be to restrict analysis to qualitative within cluster analysis using the same ANOVA, CONTRAST and PROFILE procedures as those applied to the Degenerate and Semidegenerate (A:A::B:B) reaction time data sets partitioned on cluster membership of each object (subject).

CLUSTER MEANS AND STANDARD DEVIATIONS BY CONDITION

KEY: SB0 = SENIOEGENERATE (A:B::A:B) SOLUTION TIME ZERO CUE CONDITION.
 SB2 = SENIOEGENERATE (A:B::A:B) SOLUTION TIME TWO CUE CONDITION.
 1E = 1 ELEMENT, 2E = 2 ELEMENTS, 3E = 3 ELEMENTS.
 1T = 1 TRANSFORMATION, 3T = 3 TRANSFORMATIONS.

CONDITION	SB01E1T	SB01E3T	SB02E1T	SB02E3T	SB03E1T	SB03E3T
MEAN	5837.3	6866.7	6212.7	8752.3	9067.7	10236.0
SD	1072.08	1518.29	815.46	485.55	1043.72	1114.06
CONDITION	SB21E1T	SB21E3T	SB22E1T	SB22E3T	SB23E1T	SB23E3T
MEAN	4801.0	5707.7	5115.0	7186.0	6098.3	8135.0
SD	779.49	1470.38	1212.44	1586.38	557.75	1293.12

TABLE 5.6b CLUSTER MEANS AND STANDARD DEVIATIONS FOR EACH SENIOEGENERATE (A:B::A:B) FACTORIAL CONDITION.

CLUSTER 2 MEANS AND STANDARD DEVIATIONS BY CONDITION

KEY: SB0 = SEMIOEGENERATE (A:B::A:B) SOLUTION TIME ZERO CUE CONDITION.
 SB2 = SEMIOEGENERATE (A:B::A:B) SOLUTION TIME TWO CUE CONDITION.
 1E = 1 ELEMENT, 2E = 2 ELEMENTS 3E = 3 ELEMENTS.
 1T = 1 TRANSFORMATION, 3 T = 3 TRANSFORMATIONS.

CONDITION	SB01E1T	SB01E3T	SB02E1T	SB02E3T	SB03E1T	SB03E3T
MEAN	2637.2	3190.9	2847.3	4007.7	3484.8	5188.5
SD	483.17	627.60	533.70	940.90	666.45	1285.60

CONDITION	SB21E1T	SB21E3T	SB22E1T	SB22E3T	SB23E1T	SB23E3T
MEAN	1806.1	2415.6	2301.1	3341.7	2625.0	3707.5
SD	403.61	643.36	431.71	587.87	512.89	647.53

TABLE 5.8b CLUSTER MEANS AND STANDARD DEVIATIONS FOR EACH SEMIOEGENERATE (A:B::A:B)
 FACTORIAL CONDITION.

Accordingly, Table 5.8c sets out a summary of ANOVA profiles for the two cluster membership defined by the adopted cluster solution, whilst Table 5.8d contains a summary of CONTRAST and PROFILE outputs for each cluster.

Interpretation of the statistics in Tables 5.8a - 5.8d will be returned to in the section Clusters and Stimulus Properties in Chapter 6 (Discussion).

Nondegenerate Forced Choice Geometric analogy Cluster Solutions

The specification of two and three maximum clusters produced a one cluster solution. However, the specification of four and eight maximum clusters proved to be critical. Table 5.9a shows brief summary statistics for the cluster solutions generated at these specified maximum cluster levels.

The three cluster solution generated by specifying a maximum of four initial cluster seeds is exactly replicated when a maximum of five through to seven initial seeds is specified. However, the specification of a maximum of eight initial cluster seeds leads to a new three cluster solution through an adjustment of cluster membership, with one member moving from cluster 2 to cluster 1 and the effects of marginally increasing the centroid distance, modifying variable means (general lowering) and standard deviations (generally bigger). The three cluster solution generated by specifying a maximum of eight initial cluster seeds is exactly replicated in the two cluster solutions obtained by specifying a maximum of nine, and ten initial cluster seeds.

However, with the specification of a further, eleventh, maximum initial cluster seed, the algorithm re-establishes exactly the same solution obtained by the specification

ANOVA PROFILES BY CLUSTER FOR SEMIDEGENERATE (A:B::A:B) GEOMETRIC ANALOGY DATA SET

KEY: CUE = CUE CONDITION (ZERO vs TWO) (N = 2)

ELEMENT = NUMBER OF ELEMENTS PER TERM (N=3)

TRANS = NUMBER OF TRANSFORMATIONS BETWEEN TERMS (N = 2)

LEVELS OF SIGNIFICANCE: $p < .05^*$; $p < .01^{**}$; $p < .001^{***}$; $p < .0001^{****}$

SOURCE OF VARIANCE	CLUSTER 1 (p)	CLUSTER 2 (p)
CUE	0.0038**	0.0001****
ELEMENT	0.0171*	0.0001****
TRANS	0.0235*	0.0001****
CUE * ELEMENT	0.0861	0.0001****
CUE * TRANS	0.7431	0.0006***
ELEMENT * TRANS	0.0520	0.0001****
CUE * ELEMENT * TRANS	0.1209	0.0001****

TABLE 5.Bc ANOVA PROFILES BY CLUSTERS FOR SEMIDEGENERATE (A:B::A:B) GEOMETRIC ANALOGY
RESPONSE TIME DATA.

ANOVA CONTRAST BY CLUSTER FOR SEMIDEGENERATE (A:B::A:B) GEOMETRIC ANALOGY DATA SET

KEY; CUE = CUE CONDITION (ZERO vs TWO) (N = 2)
 ELEMENT = NUMBER OF ELEMENTS PER TERM (N=3)
 TRANS = NUMBER OF TRANSFORMATIONS BETWEEN TERMS (N = 2)

LEVELS OF SIGNIFICANCE; $p \leq .05^*$; $p \leq .01^{**}$;
 $p \leq .001^{***}$; $p \leq .0001^{****}$

SOURCE OF VARIANCE	CLUSTER 1 (p)	CLUSTER 2 (p)
CUE 1 vs CUE 2	0.0035**	0.0001****
ELEMENT 1 vs ELEMENT 2	0.0504	0.0001****
ELEMENT 1 vs ELEMENT 3	0.0664	0.0001****
ELEMENT 2 vs ELEMENT 3	0.0778	0.0001****
TRANS 1 vs TRANS 2	0.0235*	0.0001****

TABLE 5.9d ANOVA CONTRASTS BY CLUSTERS FOR SEMIDEGENERATE (A:B::A:B) GEOMETRIC ANALOGY
 RESPONSE TIME DATA SET.

CRITICAL CLUSTER SOLUTIONS WITH DELETE= OPTION FOR NONDEGENERATE GEOMETRIC ANALOGY ITEMS							
MAXC =	CLUSTER NUMBER	FREQUENCY	RMS STD DEVIATION	MAXIMUM DISTANCE FROM SEED TO OBSERVATION	NEAREST CLUSTER	CENTROID DISTANCE	PSEUDO F
4	1	31	814.1	7945.8	2	8304.7	65.40
	2	24	847.7	5018.2	1	8304.7	
	3	5	1600.5	11866.3	1	19328.7	
8	1	32	815.5	8386.4	2	8341.0	65.27
	2	23	838.5	5140.2	1	8341.0	
	3	5	1600.5	12509.4	1	19446.3	

TABLE 5.9a SUMMARY OF CRITICAL CLUSTER SOLUTIONS WITH DELETE= OPTION FOR NONDEGENERATE
GEOMETRIC ANALOGY ITEMS.

of a maximum of four initial cluster seeds, and this is retained exactly through the solutions produced by specifying a maximum of twelve through to twenty initial cluster seeds.

Having accepted a three cluster solution, it was once again necessary to establish which three cluster solution was the most representative of the solutions provided by the clustering algorithm; that generated by specifying a maximum of four initial cluster seeds or that generated by specifying eight. Again, the magnitude of the pseudo F statistic was adopted as the criterion for deciding the "best" solution. In the present case the three cluster solution provided by specifying a maximum of four initial cluster seeds yields a slightly higher pseudo F statistic (65.40) and on this basis will be adopted as the most representative of the solutions provided by the clustering algorithm.

Table 5.9b sets out cluster means and cluster solutions for each of the three clusters. Again, albeit that the partitioning is explicit, there is little other evidence to indicate whether the three clusters are merely a reflection of differences in the speed of task performance alone, or whether there are other, deeper information processing constraints supporting this cluster structure.

Given that direct between group analysis with conventional statistics is inadvisable and invalid, for the reasons previously outlined, once again, the problem of examining and interpreting the foundation for cluster solution appears to be less quantitative and more qualitative on a within cluster basis. Consequently, analysis is restricted to qualitative within cluster analysis using the same ANOVA, CONTRAST and PROFILE procedures as those applied to the Degenerate, and both Semidegenerate reaction time data sets partitioned on cluster membership of each object (subject).

CLUSTER 1 MEANS AND STANDARD DEVIATIONS BY CONDITION

KEY: NS0 - NONDEGENERATE SOLUTION TIME ZERO CUE CONDITION.
 NS2 - NONDEGENERATE SOLUTION TIME TWO CUE CONDITION.
 1E - 1 ELEMENT, 2E - 2 ELEMENTS, 3E - 3 ELEMENTS.
 1T - 1 TRANSFORMATION, 2T - 2 TRANSFORMATIONS, 3T - 3 TRANSFORMATIONS.

CONDITION	NS01E1T	NS01E2T	NS01E3T	NS02E1T	NS02E2T	NS02E3T
MEAN	3401.7	4540.3	5454.7	4149.9	6241.15	7905.2
SD	702.86	609.51	841.20	600.74	845.06	964.02
CONDITION	NS03E1T	NS03E2T	NS03E3T	NS21E1T	NS21E2T	NS21E3T
MEAN	5097.8	8029.25	9180.95	2163.2	2851.75	3697.3
SD	714.83	1159.60	761.46	614.66	626.3	782.55
CONDITION	NS22E1T	NS22E2T	NS22E3T	NS23E1T	NS23E2T	NS23E3T
MEAN	3064.0	3628.4	4847.65	3413.55	4407.85	5529.3
SD	491.88	896.66	993.39	728.06	1120.21	824.68

TABLE 5.9b CLUSTER MEANS AND STANDARD DEVIATIONS FOR EACH NONDEGENERATE FACTORIAL CONDITION.

CLUSTER 2 MEANS AND STANDARD DEVIATIONS BY CONDITION

KEY : NS0 = NONDEGENERATE SOLUTION TIME ZERO CUE CONDITION.

NS2 = NONDEGENERATE SOLUTION TIME TWO CUE CONDITION.

1E = 1 ELEMENT, 2E = 2 ELEMENTS, 3E = 3 ELEMENTS.

1T = 1 TRANSFORMATION, 2T = 2 TRANSFORMATIONS, 3T = 3 TRANSFORMATIONS.

CONDITION	NS01E1T	NS01E2T	NS01E3T	NS02E1T	NS02E2T	NS02E3T
MEAN	2881.5	3033.55	4091.25	3282.85	4769.0	5448.7
SD	411.24	414.86	676.75	446.87	669.77	1007.39

CONDITION	NS03E1T	NS03E2T	NS03E3T	NS21E1T	NS21E2T	NS21E3T
MEAN	4068.85	5112.4	6342.15	1868.15	2152.05	3153.0
SD	699.14	771.28	1297.04	498.39	540.95	451.32

CONDITION	NS22E1T	NS22E2T	NS22E3T	NS23E1T	NS23E2T	NS23E3T
MEAN	2567.25	3030.0	3620.45	2835.1	3597.65	4297.35
SD	386.41	516.44	589.91	525.73	584.18	502.95

TABLE 5.9b CLUSTER MEANS AND STANDARD DEVIATIONS FOR EACH NONDEGENERATE FACTORIAL CONDITION.

CLUSTER 3 MEANS AND STANDARD DEVIATIONS BY CONDITION

KEY: NS0 = NONDEGENERATE SOLUTION TIME ZERO CUE CONDITION.

NS2 = NONDEGENERATE SOLUTION TIME TWO CUE CONDITION.

1E = 1 ELEMENT, 2E = 2 ELEMENTS, 3E = 3 ELEMENTS.

1T = 1 TRANSFORMATION, 2T = 2 TRANSFORMATIONS, 3T = 3 TRANSFORMATIONS.

CONDITION	NS01E1T	NS01E2T	NS01E3T	NS02E1T	NS02E2T	NS02E3T
MEAN	6027.8	7124.1	5136.1	6877.4	9226.4	12031.3
SD	1982.9	2208.93	1768.91	1,25.11	1358.7	1132.56
CONDITION	NS03E1T	NS03E2T	NS03E3T	NS21E1T	NS21E2T	NS21E3T
MEAN	9583.1	11871.1	13646.3	4515.6	5355.5	5932.1
SD	2806.06	596.61	1445.27	1552.27	1888.77	1791.71
CONDITION	NS22E1T	NS22E2T	NS22E3T	NS23E1T	NS23E2T	NS23E3T
MEAN	5073.4	5705.9	5405.4	6122.0	7834.8	9276.4
SD	1754.60	1328.45	762.77	2070.00	1294.23	1120.65

TABLE 5.9b CLUSTER MEANS AND STANDARD DEVIATIONS FOR EACH NONDEGENERATE FACTORIAL CONDITION.

Accordingly, Table 5.9c sets out a summary of ANOVA profiles for the three cluster membership defined by the adopted cluster solution, whilst Table 5.9d contains a summary of CONTRAST and PROFILE outputs for each cluster.

Interpretation of the statistics in Tables 5.9a - 5.9d will be returned to in the section, Clusters and Stimulus Properties in Chapter 6 (Discussion).

Finally, in order to establish whether the basis for clustering has been prescribed by some other partitioning variable, a series of nonparametric tests were conducted on the cluster solutions pertaining to each of the analogy types.

At this point, however, it should be mentioned that there was no attempt to investigate the relationship between cluster membership and handedness, given that only (three) subjects (5%) were recorded as being predominantly left handed.

For the Degenerate two cluster solution, a binomial two tail test failed to reveal a significant association between cluster membership and gender at the 10 per cent significance level. Furthermore, neither the association between cluster membership and subjective item complexity code group, nor that between cluster membership and subjective strategy code group was shown to be significant at the 10 per cent significance level, employing the same nonparametric test.

For the Semidegenerate (A:A::B:B) three cluster solution, a binomial two tail test failed to reveal a significant association between cluster membership and gender at the 10 per cent significance level. Similarly, chi-squared tests of association between cluster membership and subjective item complexity code group, and between cluster membership

ANOVA PROFILES BY CLUSTER FOR NONDEGENERATE GEOMETRIC ANALOGY DATA SET

KEY: CUE = CUE CONDITION (ZERO vs TWO) (N=2)

ELEMENT = NUMBER OF ELEMENTS PER TERM (N = 3)

TRANS = NUMBER OF TRANSFORMATIONS BETWEEN TERMS (N=3)

LEVELS OF SIGNIFICANCE; $p < .05^*$; $p < .01^{**}$;

$p < .001^{***}$; $p < .0001^{****}$

SOURCE OF VARIANCE	CLUSTER 1 (p)	CLUSTER 2 (p)	CLUSTER 3 (p)
CUE	0.0001****	0.0001****	0.0001****
ELEMENT	0.0001****	0.0001****	0.0001****
TRANS	0.0001****	0.0001****	0.0001****
CUE * ELEMENT	0.0001****	0.0002***	0.0011**
CUE * TRANS	0.0001****	0.0001****	0.0022**
ELEMENT * TRANS	0.0001****	0.0001****	0.0388*
CUE * ELEMENT * TRANS	0.0001****	0.0001****	0.1365

TABLE 5.9 ANOVA PROFILES BY CLUSTERS FOR NONDEGENERATE GEOMETRIC ANALOGY RESPONSE
TIME DATA.

ANOVA CONTRAST BY CLUSTER FOR NONDEGENERATE GEOMETRIC ANALOGY DATA SET

KEY: CUE = CUE CONDITION (ZERO vs TWO) (N = 2)

ELEMENT = NUMBER OF ELEMENTS PER TERM (N=3)

TRANS = NUMBER OF TRANSFORMATIONS BETWEEN TERMS (N = 2)

LEVELS OF SIGNIFICANCE: $p < .05^*$; $p < .01^{**}$; $p < .001^{***}$; $p < .0001^{****}$

SOURCE OF VARIANCE	CLUSTER 1 (p)	CLUSTER 2 (p)	CLUSTER 3 (p)
CUE 1 vs CUE 2	0.0001****	0.0001****	0.0001****
ELEMENT 1 vs ELEMENT 2	0.0001****	0.0001****	0.0064**
ELEMENT 1 vs ELEMENT 3	0.0001****	0.0001****	0.0006***
ELEMENT 2 vs ELEMENT 3	0.0001****	0.0001****	0.0007***
TRANS 1 vs TRANS 2	0.0001****	0.0001****	0.0071**
TRANS 1 vs TRANS 3	0.0001****	0.0001****	0.0003***
TRANS 2 vs TRANS 3	0.0001****	0.0001****	0.0004***

TABLE 5.9d ANOVA CONTRASTS BY CLUSTER FOR NONDEGENERATE GEOMETRIC ANALOGY RESPONSE TIME DATA SET.

and subjective strategy code group failed to reveal a significant association at the 10 per cent level; though in view of the number of cells with frequency counts of less than five, this latter analysis may not be truly accurate.

For the Semidegenerate (A:B::A:B) two cluster solution, a binomial two tail test failed to reveal a significant association between cluster membership and gender at the 10 per cent significance level. Furthermore, neither the association between cluster membership and subjective item complexity group, nor that between cluster membership and subjective strategy code group was shown to be significant at the 10 per cent significance level, employing the same nonparametric test.

Lastly, for the Nondegenerate three cluster solution, a binomial two tail test failed to reveal a significant association between cluster membership and gender at the 10 per cent significance level. Similarly, chi-squared tests of association between cluster membership and subjective item complexity code group, and between cluster membership and subjective strategy code group, failed to reveal a significant association at the 10 per cent significance level, though in view of the number of cells with frequency counts of less than five, this latter analysis may not be truly accurate.

In sum the partitioning variables of gender, subjective item complexity and subjective strategy offer no associative explanation for the basis of cluster membership in any of the cluster solutions relevant to the "four" forced choice geometric analogy types.

Chapter Summary

This chapter has presented an explanation of the rationale supporting the chosen data analysis techniques employed in

the present research, together with brief explanations of the techniques themselves and the presentation of results yielded by these techniques.

An interpretation of these results and a discussion of their implications is contained in the following chapter, Chapter 6 (Discussion).

CHAPTER 6

DISCUSSION

Overview to Discussion of Results

In the section Overview to Data Analysis in Chapter 5, it was mentioned that analysis and subsequent results emerging from the present research framework would be presented in accordance with a number of related questions concerning the concepts central to analogical reasoning, componential analysis, mental workload, etc.

Consistent with this general approach to the presentation of research findings, the results of these data analyses will not be discussed in terms of discrete "facts" but, rather, in terms of the related concepts that, throughout the theoretical and conceptual discussions in Chapters 2 and 3, have been suggested as central to the investigation of mental workload influences on (Geometric) analogical problem solving.

The first issue to be considered is that of stimulus complexity. Arguably this issue represents the most essential factor in any experimental paradigm purporting to address the issues of spatial cognition through judgements about objects, shapes and their transformations. It will be recalled from Chapters 2 and 3 that much of the equivocal nature of research findings in the crucial areas of analogical reasoning and mental workload have hinged upon experimenters' a priori assignments of stimulus properties along some arbitrary criterion/dimension of "complexity" (e.g., Sternberg, 1977; Mulholland et al., 1980).

Consequently the analysis of data and the interpretation of effects of stimulus complexity in the design of the present research is dealt with in the section Stimulus Complexity and Task Performance later in this chapter.

Secondly, the issues of experimental design, task demands, individual differences and the opportunity for differential task strategy, are all central to the clarity with which any result may be discussed and generalised. These issues are dealt with in the section Task Structure and Solution Strategy later in this chapter.

The combined outcomes of the discussions of the stimulus complexity and solution strategy issues are central to the overall interpretation of the present experiment's findings on any wider scale; for example, in terms of a theory about analogical reasoning research, individual differences and spatial cognition. This level of interpretation is dealt with in the section "Componential Analysis" Revisited later in the chapter.

The next two issues to be addressed are promoted by the exploratory analysis of individual differences in spatial cognition through the use of a cluster analysis technique. Again, the issue revolves around the concepts of, firstly stimulus properties and the attendant evidence for an interpretation of cluster content on the basis of subjective self-reported solution strategies, and secondly, the extent to which the cluster solution can be used to account for patterns of true individual differences. These two issues will be dealt with respectively in the sections Clusters and Stimulus Properties and Clusters and Individual Differences later in this chapter.

The sixth issue to be considered is that of major concern of the present research; the extent to which findings from an apparently stable cognitive phenomenon, geometric

analogy problem solving, can offer insight into the influence of mental workload variation on cognitive task performance. This issue will be dealt with in the section Componential Analysis and Mental Workload later in this chapter.

The last issue to be considered is perhaps the broadest. It was mentioned in Chapter 2 that although Sternberg's theory of analogical reasoning provided valuable insights into the nature of cognitive abilities, it has until now been excluded from research designed to investigate the concept of mental workload. It is the contention of the present research that Sternberg's research could well provide valuable information regarding the nature of task structure, information which could serve to enhance the present understanding of the concept of mental workload. A development such as this could stand to benefit future research into the development of strategies for reducing the mental workload experienced by operators of automated systems. This speculation is dealt with in the section Prospects for Componential Analysis and Mental Workload Research later in this chapter.

Stimulus Complexity and Task Performance

In Chapters 3 and 4 it was noted that central to the componential interpretation of geometric analogy solution is the hypothesis that a major indicator of item complexity is the total time to solve an item. However, Mulholland et al. (1980) state that their own rational analysis of geometric analogy items found on standardised intelligence tests indicated that item complexity was apparently related to increases in both the number of elements per term and the number of transformations between terms in an item. This observation suggests that item complexity can be expressed in terms of faulty processing that may result from either (or both) 'incorrect element or (and) transformation identification and testing.

Sternberg (1977), in his pioneering investigation of forced choice geometric analogy solution, considered the amount of precueing information supplied to subjects to be a major index of item complexity. It was mentioned in Chapter 4 that the precueing procedure adopted by Sternberg (1977) in his Geometric Analogy Experiment, would be employed in the present research procedure. Furthermore, it was contended that employing such a precueing procedure would facilitate the manipulation of task openness (uncertainty), with the amount of precueing information supplied being indirectly proportional to the degree of task openness, i.e., the more cues provided, the less the degree of task openness.

The present study combines certain assumptions regarding item complexity from the research of both Sternberg (1977) and Mulholland et al. (1980), i.e., that geometric analogy item complexity is related to the amount of precueing information supplied to subjects, the number of elements per term of an item and the number of transformations between terms of an item.

In the present experiment, the definition of item complexity was dependent upon geometric analogy type. For Degenerate items, complexity was defined in terms of cue condition (zero or two) and number of elements per term (one, two or three), for Semidegenerate items (both variations, i.e., $A:A::B:B$ and $A:B::A:B$) and Nondenerate items, complexity was defined in terms of cue condition (zero or two), number of elements per term (one, two or three) and number of transformations between terms (one or three for both variations of the Semidegenerate items, and one, two or three for the Nondenerate items).

The incorporation of these dimensions of item complexity into the design of the present experiment provides one obvious method of assessing the stimulus complexity effect, through direct parametric analysis of response time;

seeking main effects and interactions related to these dimensions. In addition, the inclusion of pre and post-test subjective rankings of item complexity provides an opportunity to further examine the relationship between complexity and other experimental factors, for example, analogy format.

The stimulus complexity issue will be considered separately for each of the geometric analogy types.

Stimulus Complexity and Degenerate Geometric Analogy Task Performance

Table 5.2a (page 186) shows that in analysis of variance for the Degenerate geometric analogy factorial data set, strong, significant main effects for cue condition and number of elements per term are present in the data. Table 5.3a (page 190) shows that single degree of freedom contrasts between cue conditions, and between number of elements per term identify significant differences between each pair comparison, whilst a superficial scan of the means set out in Table 5.1a (page 180) will indicate that, in general, solution times for items presented in the two cue condition were faster than solution times for items presented in the zero cue condition, and that an increase in the number of elements per term was directly proportional to an increase in solution time for items presented in both prequeuing conditions.

These effects correspond with expectations based on the experimental assumptions of item complexity highlighted in the section Overview to Discussion of Results earlier in this chapter. In addition, they are in accordance with the assumptions of both Sternberg (1977) and Mulholland et al., (1980) regarding geometric analogy item complexity.

Moreover, Table 5.2a (page 186) shows a strongly

significant interaction between cue condition and number of elements per term, indicating that the effects of number of elements per term vary significantly across cue condition. In fact, the means presented in Table 5.1a (page 180) indicate that for each element variation solution times for items presented in the two cue condition were faster than solution times for items presented in the zero cue condition. Again, these findings are consistent with the experimental assumptions of item complexity.

Stimulus Complexity and Semidegenerate (A:A::B:B) Geometric Analogy Task Performance

Table 5.2b (page 187) indicates that in analysis of variance for the Semidegenerate (A:A::B:B) geometric analogy factorial data set, strong, significant main effects for cue condition, number of elements per term and number of transformations between terms are present in the data. Table 5.3b (page 191) shows that single degree of freedom contrasts between cue conditions, between number of elements per term, and between number of transformations between terms identify significant differences between each pair comparison, whilst a superficial scan of the means set out in Table 5.1b (page 181) will indicate that in general, solution times for items presented in the two cue condition were faster than solution times for items presented in the zero cue condition. Equally important, is that number of transformations between terms, rather than number of elements per term, appeared to be the major determinant of solution time for items presented in both zero and two cue conditions. While an increase in the number of elements per term resulted in an increase in solution time; a corresponding increase in the number of transformations between terms was shown to have a more significant influence on solution time. Apparently evaluation of transformations took more time than did element analysis and was the primary source of increases in solution time.

This finding is consistent with that of Mulholland et al. (1980).

Moreover, Table 5.2b (page 187) shows a strongly significant interaction between cue condition and number of elements per term, cue condition and number of transformations between terms, and between number of elements per term and number of transformations between terms. In fact, the means presented in Table 5.1b (page 181, indicate that for each element variation and simultaneous transformation variation, solution times for items presented in the two cue condition were consistently faster than solution times for items presented in the zero cue condition. In addition, it is evident that the effects of number of elements per term vary significantly across variations in number of transformations between terms. These findings are consistent with those of both Sternberg (1977) and Mulholland et al. (1980).

Stimulus Complexity and Semidegenerate (A:B::A:B) Geometric Analogy Task Performance

Table 5.2c (page 188) indicates that in analysis of variance for the Semidegenerate (A:B::A:B) geometric analogy factorial data set, strong, significant main effects for cue condition, number of elements per term and number of transformations between terms are present in the data. Table 5.3c (page 192) shows that single degree of freedom contrasts between cue conditions, between number of elements per term, and between number of transformations between terms identify significant differences between each pair comparison, whilst a superficial scan of the means set out in Table 5.1c (page 182) will indicate that in general, solution times for items presented in the two cue condition were faster than solution times for items presented in the zero cue condition. Furthermore, as was found when examining the mean solution times for the Semidegenerate

(A:A::B:B) items, number of transformations between terms, rather than number of element per term, appeared to be the major determinant of solution time for items presented in both cue conditions. While an increase in the number of elements per term resulted in an increase in solution time, a corresponding increase in the number of transformations between term was found to have a more significant influence on solution time. As was the case with the Semidegenerate (A:A::B:B) items, evaluation of transformations appeared to take more time than did element analysis and was the primary source of increases in solution times. Again, this finding is consistent with that of Mulholland et al. (1980).

Moreover, Table 5.2c (page 188) shows a strongly significant interaction between cue condition and number of elements per term, cue condition and number of transformations between terms, number of elements per term and number of transformations between terms. In fact, the means presented in Table 5.1c (page 182) indicate that for each element variation and corresponding transformation variation solution times for items presented in the two cue condition were consistently faster than for items presented in the zero cue condition. In addition, as was the case with the Semidegenerate (A:A::B:B) items, the effects of number of elements per term changed significantly across variations in the number of transformations between terms. Again, these findings are consistent with those of both Sternberg (1977) and Mulholland et al. (1980).

Stimulus Complexity and Nondegenerate Geometric Analogy Task Performance

Table 5.2d (page 189) indicates that in analysis of variance for the Nondegenerate geometric analogy factorial data set, strong, significant main effects for cue condition, number of elements per term and number of

transformations between terms are present in the data. Table 5.3d (page 193) shows that single degree of freedom contrasts between cue conditions, between number of elements per term, and between number of transformations between terms identify significant differences between each pair comparison, whilst a superficial scan of the means set out in Table 5.1d (page 183) will indicate that in general, solution times for items presented in the two cue condition were faster than solution times for items presented in the zero cue condition. Moreover, as was found when examining the mean solution times for the Semidegenerate (A:A::B:B) and Semidegenerate (A:B::A:B) items, number of transformations between terms, rather than number of elements per term, appear to be the major determinant of solution time for items presented in both cue conditions. While an increase in the number of elements per term resulted in an increase in solution time, a corresponding increase in the number of transformations between terms was found to have a more significant influence on solution time. As was the case with both Semidegenerate (A:A::B:B) and Semidegenerate (A:B::A:B) analogy types, evaluation of transformations appeared to take more time than did element analysis and was the primary source of increases in solution time. Again, this finding is consistent with that of Mulholland et al., (1980).

In addition, Table 5.2d (page 189) shows a strongly significant interaction between cue condition and number of elements per term, cue condition and number of transformations between terms, number of elements per term and number of transformations between terms and between cue condition, number of elements per term and number of transformations between terms. In fact, the means presented in Table 5.1d (page 183) indicate that for each element variation and corresponding transformation variation, solution times for items presented in the two cue condition were consistently faster than solution times

for items presented in the zero cue condition. Moreover, as was the case with both Semidegenerate (A:A::B:B) and Semidegenerate (A:B::A:B) analogy types, the effects of number of elements per term changed significantly across variations in the number of transformations between terms. Again, these findings are consistent with those of both Sternberg (1977) and Mulholland et al. (1980).

The presence of the effects identified for each analogy type does not, however, act as their full explanation. An appropriate supplementary question should be the extent to which these effects are consistent for all subjects in this comparatively large group. Indeed is stimulus complexity a uniform phenomenon?

It will be recalled from Chapters 4 and 5 that subjects were asked to subjectively rank four forced choice geometric analogy items (one from each of the experimental analogy types) in order of complexity, firstly before exposure to the experimental task and again thereafter. The assumption is that any changes in complexity ranking will reflect a modification of the subject's complexity criteria because of their exposure to the experimental task demands. This procedure provides the opportunity to investigate the previously unexplored relationship between item complexity and item format.

Table 5.4 (page 195) sets out subjects' rankings. Of particular significance is that only five of the potential sixty-four orderings emerge and the frequency of their distribution is non random. An examination of the means set out in Tables 5.1a through to 5.1d indicates that in general subjects responded to Semidegenerate items (both A:A::B:B and A:B::A:B variations) faster than Nondegenerate items, and to Degenerate items faster than both variations of Semidegenerate items. In addition, response times for Semidegenerate (A:A::B:B) items were faster than those for

Semidegenerate (A:B::A:B) items. Of interest also is the coincidence of ranked complexity judgements with the effect noted above. In Table 5.4, with rank 1 given to the least complex item and rank 4 to the most complex, a Degenerate (B) < Semidegenerate (A:A::B:B) (D) < Semidegenerate (A:B::A:B) (A) < Nondegenerate (C) order would be equal to an A = 3, B = 1, C = 4, D = 2 ranking set (3142). This order is strongest in Group 1 (N = 31) but only as a post-test judgement.

The 3142 order features in each of the remaining code groups, Group 2 (N = 18), however, adopt this subjective criteria pretest and maintain it. In groups 3 (N = 6), 4 (N = 4) and 5 (N = 1) the 3142 order is present only as a post-test judgement, indicating an adjustment in judgement criteria following exposure to task demands.

On an interpretive level, complexity code groups 1 (N = 31), 3 (N = 6), 4 (N = 4) and 5 (N = 1) indicate a shift from pre to post-test judgements. As noted previously in this section, this indicates an adjustment in judgement criteria for each of these code groups, following exposure to task demands. It is, however, beyond the scope of this study, given its exploratory nature, to attempt an explanation of these adjustments.

Of significance to the present research is the presence of the 3142 order as a post-test judgement for each of the complexity code groups. On this basis, the least complex post-test item is judged as the Degenerate item. In addition, the consistent 3142 post-test order indicates that subjects ranked the Semidegenerate (A:A::B:B) item as being less complex than the Semidegenerate (A:B::A:B) item. Since both Semidegenerate items comprised the same number of elements per term, and transformations between terms, and given that they were both presented in the zero cue condition, it is contended that in addition to the

individual and combined influence of number of elements per term and number transformations between terms on subjective item complexity, item format is a further contributing factor.

Finally, the most complex item is judged as the Nondegenerate item. Given that the Nondegenerate item comprised the same number of elements and transformations, and was presented in the same precueing condition as the two Semidegenerate items, provides support to the contention that item format is a contributing factor to subjective item complexity.

In general, the stimulus complexity effects identified for each of the forced choice geometric analogy types support the experimental assumptions that forced choice geometric analogy item complexity is related to the amount of precueing information supplied, the number of elements per term and the number of transformations between terms. Moreover, the subjective complexity ranking procedure adopted, and the results thereof, presented in Table 5.4 (page 195), suggest that irrespective of their approach to task demands, the basis for subjective item complexity is uniform for all subjects.

It was contended in the section Solution Strategy and the Experimental Task in Chapter 5, that appealing to subjects' post hoc introspective reports could provide valuable insights into the extent to which variations in the demand characteristics of forced choice geometric analogy items influence the order of execution of the cognitive components (proposed by Sternberg (1977) as being) involved in geometric analogy solution. Complexity code groups were thus used as the basis for post hoc partitioning of the performance data pertaining to each of the geometric analogy types, to determine whether subjective complexity code offered an associative

explanation for the basis of subjective strategy code group membership. This issue will be reviewed in the section Task Structure and Solution Strategy next in this chapter.

Task Structure and Solution Strategy

It was noted in the section Aims of the Present research in Chapter 4 that this study aims to investigate the generalisability of the cognitive component models described by Sternberg (1977) as best defining task performance on Nondegenerate forced choice geometric analogy items. Specifically, whether meaningful intra and interindividual variations are apparent, and if so whether such variations could be attributed to the level of mental workload experienced by subjects as a result of their exposure to experimental task demands. Also, it was stated that the present study would propose models, based on subjects' post hoc introspective reports, that best define task performance on Degenerate and Semidegenerate forced choice geometric analogies, in terms of performance components adopted and order of component execution.

Attempting to address these experimental aims is of central importance to the design of the present research procedure. The rationale here is straightforward.

Firstly, it will be recalled from the section Procedure in Chapter 4 that all subjects taking part in the present research were initially screened to ensure that they had no knowledge of cognitive psychology experimental procedures in general and cognitive components theory and mental workload research in particular. All subjects may thus be thought of as being research naive in the true sense. This is in contrast to much of the published research into analogical problem solving, which often presents data derived from small numbers of hypothesis sophisticated or

research sophisticated subjects (see the section The Research Design of Sternberg's (1977) Geometric Analogy Experiment, and its Relevance to the Present Research in Chapter 4).

Secondly, controlled use of experimental instructions was enforced to avoid implicit strategy priming and response set creation. This is a major contrast to the research procedures of Sternberg (1977) and Mulholland et al. (1980), where prior exposure to, and performance on geometric analogy items were criteria for subject selection.

The potential implications of such procedures are discussed in the section Subjective Influence on Cognitive Task Performance in Chapter 4. Given the present research procedure it was assumed that results yielded were attributable to variations in item complexity, and that subjective task performance was representative of the range of reasoning strategies available for geometric analogy solution, as well as being highly responsive to workload influences.

Thirdly, following completion of the experimental trials, subjects were asked whether they had developed any conscious strategy for handling the items from each of the geometric analogy types. Adopting the procedure outlined in the section Procedure in Chapter 4, each strategy explanation was carefully analysed, recorded and coded.

Having employed the procedure described in the section Solution Strategy and the Experimental Task in Chapter 5, it was assumed that variations in subjective strategy reports for the two examples of each analogy type would reflect a modification in the subjects' solution strategy brought about by the variation in task demand characteristics between the two examples.

It was noted in the section Solution Strategy and the Experimental Task in Chapter 5 that, for each geometric analogy type, a detailed discussion of the dominant subjective task strategy types, as well as their respective variations, identified from subjects' post hoc introspective reports, would be presented in this section.

A separate discussion will be presented for each geometric analogy type.

Degenerate Geometric Analogy Subjective Task Strategy Types

It will be recalled from the section Degenerate Forced Choice Geometric Analogy Subjective Task Strategy Types in Chapter 5, that in terms of performance components adopted and order of component execution, two dominant subjective task strategy types were identified from subjects' post hoc introspective reports relating to the 1 element Degenerate geometric analogy example.

Thirty-one subjects (52%) reported employing the first of these dominant strategy types. A schematic flow chart of this strategy type is presented in Figure 6.1a.

Subjects employing this strategy type begin analogy solution by encoding the first analogy term, and then the second analogy term. Next they infer the relation between the first and second analogy terms. The subjects then encode the third analogy term. Following this, the two answer options are encoded. Having encoded the two answer options, subjects attempt to apply from the third term to each answer option a relation analogous to the one inferred between the first and second terms. Subjects would then respond with either "A" or "B" thereby completing the analogy problems.

Three variations of this strategy type were identified. These variations apparently arose as a result of subjects employing either a self-terminating or exhaustive application process. The first of these variations, employed by fourteen subjects (23%), involved a self-terminating application process, irrespective of which answer option was the correct one. The second variation, employed by six subjects (10%), involved an exhaustive application process, irrespective of which answer option was the correct one. The third variation, employed by eleven subjects (18%), involved a self-terminating application process if the first answer option was correct, and an exhaustive application process if the second answer option was correct.

Twenty-nine subjects (48%) reported employing the second dominant strategy type. A schematic flow chart of this strategy type is presented in Figure 6.1b.

Subjects employing this strategy type begin analogy solution by encoding the first analogy term, and then the second analogy term. Next they infer the relation between the first and second analogy terms. The subjects then encode the third analogy term and, following this, map the relation between the first and third terms. The two answer options are then encoded. Subjects then attempt to apply from the third term to each answer option a relation analogous to the one inferred between the first and second terms. Subjects would then respond with either "A" or "B", thereby completing the analogy problem.

Two variations of this strategy type were identified. Again, these variations arose as a result of subjects employing either a self-terminating or exhaustive application process. The first of these variations, employed by eleven subjects (18%), involved a self-terminating application process, irrespective of which

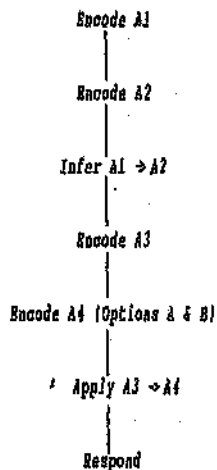


Figure 6.1 (a)

Schematic flow chart of the first dominant subjective task strategy type applicable to Degenerate item solution.

* Variations of this process

- As applicable to 1 element Degenerate item solution.
 - (1) Self-terminating, irrespective of which answer option was correct - employed by fourteen subjects (23%).
 - (2) Exhaustive, irrespective of which answer option was correct - employed by six subjects (10%).
 - (3) Self-terminating if first answer option correct, and exhaustive if second answer option correct - employed by eleven subjects (18%).
- As applicable to 3 element Degenerate item solution.
 - (1) Exhaustive, irrespective of which answer option was correct - employed by nine subjects (15%).
 - (2) Self-terminating if first answer option correct, and exhaustive if second answer option correct - employed by twenty four subjects (40%).

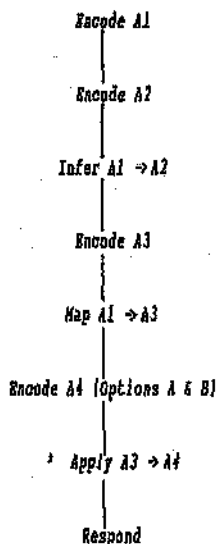


Figure 6.1 (b)

Schematic flow chart of the second dominant subjective task strategy type applicable to Degenerate item solution.

Variations of this process

- As applicable to 1 element Degenerate item solution.
 - (1) Self-terminating, irrespective of which answer option was correct - employed by eleven subjects (18%).
 - (2) Self-terminating if first answer option correct, and exhaustive if second answer option correct - employed by eighteen subjects (30%).
- As applicable to 3 element Degenerate item solution.
 - (1) Self-terminating, irrespective of which answer option was correct - employed by one subject (1.6%).
 - (2) Self-terminating if first answer option correct, and exhaustive if second answer option correct - employed by twenty-six subjects (43%).

answer option was the correct one. The second variation, employed by eighteen subjects (30%), involved a self-terminating application process if the first answer option was correct, and an exhaustive application process if the second answer option was correct.

It was noted previously, in the section Degenerate Forced Choice Geometric Analogy Subjective Task Strategy Types in Chapter 5, that in terms of performance components adopted and order of component execution, two dominant subjective task strategy types, identical to those identified for the 1 element Degenerate geometric analogy example, were identified from subjects' post hoc introspective reports relating to the 3 element Degenerate geometric analogy example.

Thirty-three subjects (55%) reported employing the first of these dominant strategy types. See figure 6.1a earlier in the section for the schematic flow chart of this strategy type. Since this strategy type has been described previously, in the discussion pertaining to the 1 element Degenerate geometric analogy example, there is no need to describe it again. It is, however, important to note that two variations of this strategy type were identified. Again, these variations could apparently be attributed to subjects employing either a self-terminating or exhaustive application process. The first of these variations, employed by nine subjects (15%), involved an exhaustive application process, irrespective of which answer option was the correct one. The second variation, employed by twenty-four subjects (40%) involved a self-terminating application process if the first answer option was correct, and an exhaustive application process if the second answer option was correct.

Twenty-seven subjects (45%) reported employing the second dominant strategy type. See Figure 6.1b earlier in this

section for the schematic flow chart of this strategy type. Since this strategy type was described previously, in the discussion pertaining to the 1 element Degenerate geometric analogy example, it will not be described again here. It is, however, important to note that two variations of this strategy type were identified. Again, these variations could apparently be attributed to subjects employing either a self-terminating or exhaustive application process. The first of these variations, employed by one subject (2%), involved a self-terminating application process, irrespective of which answer option was the correct one. The second variation, employed by twenty-six subjects (43%), involved a self-terminating application process if the first answer option was correct, and an exhaustive application process if the second answer option was correct.

The preceding discussion of subjects' post hoc introspective reports relating to the solution of both 1 element and 3 element Degenerate forced choice geometric analogy items reveals a number of issues relevant to the present research. Firstly, in terms of the performance components adopted and order of component execution, subjective strategy reports suggest the dominance of two strategy types for both 1 element and 3 element Degenerate items (see Figures 6.1a and 6.1b earlier in this section for schematic flow charts of these dominant strategy types). The major difference between these strategy types being that the one presented in Figure 6.1b incorporates a mapping process, between the first and third analogy terms, immediately after the encoding of the third term, whereas the strategy type presented in Figure 6.1a does not incorporate this process. Secondly, and probably more significant to the present study, is that the two dominant subjective task strategy types identified from subjects' post hoc introspective reports relating to 3 element Degenerate items were identical to those reported for the

solution of 1 element Degenerate items. Thus, it would appear that despite the increase in number of elements per analogy term, from 1 to 3, the identical two strategy types dominate. This would appear to suggest that on the basis of subjective strategy reports, for Degenerates forced choice geometric analogy items, increases in item complexity fail to elicit meaningful strategic variations on the part of subjects in order to overcome increases in mental workload and maintain high standards of task performance. Thirdly, on the basis of subjective task strategy reports, variations of the dominant strategy types reported for both 1 and 3 element Degenerate item solution were identified. As noted earlier in this section these variations apparently arose as a result of subjects employing either a self-terminating or exhaustive application process. The nature of these variations, for both 1 and 3 element Degenerate items, was discussed previously in this section. It is, however, important to note that subject's post hoc introspective reports pertaining to both 1 and 3 element Degenerate item solution revealed that the order of answer options, i.e., which answer option was correct (A or B), was a determining factor of the type of application process employed. Where the first answer option was correct, subjects reporting this strategy type variation would employ a self-terminating application process, and where the second answer option was correct, these subjects would employ an exhaustive application process. This pattern was found to be more prominent in the subjective task strategy reports relevant to the solution of 3 element Degenerate items than in those relating to the solution of 1 element Degenerate items.

Therefore, based on subjects' post hoc introspective reports regarding performance components adopted and order of component execution, it would appear that the process models presented in Figures 6.1a and 6.1b best define task performance on Degenerate forced choice geometric analogy items.

Semidegenerate (A:A::B:B) Geometric Analogy Subjective
Task Strategy Types

It will be recalled from the section Semidegenerate Forced Choice Geometric Analogy (Format A:A::B:B) Subject Task Strategy Types in Chapter 5, that in terms of performance components adopted and order of component execution, two dominant subjective task strategy types were identified from subjects' post hoc introspective reports relating to the 1 element - 1 transformation Semidegenerate (A:A::B:B) geometric analogy example.

Forty-eight subjects (80%) reported employing the first of these dominant strategy types. A schematic flow chart of this strategy type is presented in Figure 6.2a.

Subjects employing this strategy type begin analogy solution by encoding the first analogy term, and then the second analogy term. Next they infer the relation between the first two analogy terms. The subjects then encode the third analogy term and, following this, map the relation between the first and third terms. The two answer options are then encoded. Subjects then attempt to apply from the third term to each answer option a relation analogous to the one inferred between the first and second terms. Subjects would then respond with either "A" or "B", thereby completing the analogy problem.

Three variations of this strategy type were identified. These variations apparently arose as a result of subjects employing either a self-terminating or exhaustive application process. The first of these variations, employed by nineteen subjects (32%), involved a self-terminating application process, irrespective of which answer option was the correct one. The second variation, employed by five subjects (8%), involved an exhaustive application process, irrespective of which answer option

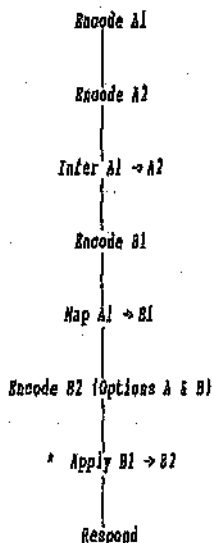


Figure 6.2 (a)

Schematic flow chart of the First dominant subjective task strategy type applicable to Semidegenerate (A:A::B:B) item solution.

Variations of this process

- As applicable to 1 element - 1 transformation Semidegenerate (A:A::B:B) item solution.
 - (1) Self-terminating, irrespective of which answer option was correct - employed by nineteen subjects (32%).
 - (2) Exhaustive, irrespective of which answer option was correct - employed by five subjects (8%).
 - (3) Self-terminating if first answer option correct, and exhaustive if second answer option correct - employed by 24 subjects (40%).

- As applicable to 3 element - 3 transformation Semidegenerate (A:A::B:B) item solution.
 - (1) Exhaustive, irrespective of which answer option was correct - employed by thirty-nine subjects (65%).
 - (2) Self-terminating if first answer option correct, and exhaustive if second answer option correct - employed by thirteen subjects (22%).

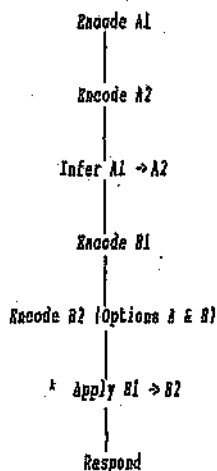


Figure 6.2 (b) Schematic flow chart of the second dominant subjective task strategy type applicable to Semidegenerate (A:A::B:B) item solution.

Variations of this process

- As applicable to 1 element - 1 transformation Semidegenerate (A:A::B:B) item solution.
 - (1) Self-terminating, irrespective of which answer option was correct - employed by five subjects (8%).
 - (2) Exhaustive, irrespective of which answer option was correct - employed by two subjects (4%).
 - (3) Self-terminating if first answer option correct, and exhaustive if second answer option correct - employed by five subjects (8%).
- As applicable to 3 element - 3 transformation Semidegenerate (A:A::B:B) item solution.
 - (1) Exhaustive, irrespective of which answer option was correct - employed by eight subjects (13%).

was the correct one. The third variation, employed by twenty-four subjects (40%), involved a self-terminating application process if the first answer option was correct, and an exhaustive application process if the second answer option was correct.

Twelve subjects (20%) reported employing the second dominant strategy type. A schematic flow chart of this strategy type is presented in Figure 6.2b.

Subjects employing this strategy type begin analogy solution by encoding the first analogy term, and then the second analogy term. Next they infer the relation between the first two analogy terms. The subjects then encode the third analogy term. Following this, the two answer options are encoded. Having encoded the two answer options, subjects attempt to apply from the third term to each answer option a relation analogous to the one inferred between the first two terms. Subjects would then respond with either "A" or "B", thereby completing the analogy problem.

Three variations of this strategy type were identified. Again, these variations apparently arose as a result of subjects employing either a self-terminating or exhaustive application process. The first of these variations, employed by five subjects (8%), involved a self-terminating application process, irrespective of which answer option was the correct one. The second variation, employed by two subjects (4%), involved an exhaustive application process, irrespective of which answer option was the correct one. The third variation, employed by five subjects (8%), involved a self-terminating application process if the first answer option was correct, and an exhaustive application process if the second answer option was correct.

It was noted previously, in the section Semidegenerate Forced Choice Geometric Analogy (Format A:A::B:B) Subjective Task Strategy Types in Chapter 5, that in terms of performance components adopted and order of component execution, two dominant subjective task strategy types, identical to those identified for the 1 element - 1 transformation Semidegenerate (A:A::B:B) geometric analogy example, were identified from subjects' post hoc introspective reports relating to the 3 element - 3 transformation Semidegenerate (A:A::B:B) geometric analogy example.

Fifty-two subjects (87%) reported employing the first of these dominant strategy types. See Figure 6.2a earlier in this section for the schematic flow chart of this strategy type. Since this strategy type has been described previously, in the discussion relating to the 1 element - 1 transformation Semidegenerate (A:A::B:B) geometric analogy example, it will not be described again. It is, however, important to note that two variations of this strategy type were identified. These variations could apparently be attributed to subjects employing either a self-terminating or exhaustive application process. The first of these variations, employed by thirty-nine subjects (65%), involved an exhaustive application process, irrespective of which answer option was the correct one. The second variation, employed by thirteen subjects (22%), involved a self-terminating application process if the first answer option was correct, and an exhaustive application process if the second answer option was correct.

Eight subjects (13%) reported employing the second dominant strategy type. See Figure 6.2b earlier in this section for the schematic flow chart of this strategy type. Since this strategy type has been described previously, in the discussion pertaining to the 1 element - 1 transformation Semidegenerate (A:A::B:B) geometric analogy example, it

will not be described again. It is, however, important to note that only one variation of this strategy type was identified. This variation, employed by eight subjects (13%), involved an exhaustive application process, irrespective of which answer option was the correct one.

The preceding discussion of subjects' post hoc introspective reports relating to the solution of both 1 element - 1 transformation and 3 elements - 3 transformation Semidegenerate (A:A::B:B) forced choice geometric analogy items reveals a number of issues relevant to the present research. Firstly, in terms of performance components adopted and order of component execution, subjective strategy reports suggest the dominance of two strategy types for both 1 element - 1 transformation and 3 element - 3 transformation Semidegenerate (A:A::B:B) items (see Figures 6.2a and 6.2b earlier in this section for schematic flow charts of these strategy types). The major difference between these strategy types being that the one presented in Figure 6.2a incorporates a mapping process, between the first and third analogy terms, immediately after the encoding of the third term, whereas the strategy type presented in Figure 6.2b does not incorporate this process. Secondly, and possibly more significant to the present study, is that the two dominant subjective task strategy types identified from subjects' post hoc introspective reports relating to 3 element - 3 transformation Semidegenerate (A:A::B:B) items were identical to those reported for the solution of 1 element - 1 transformation Semidegenerate (A:A::B:B) items. Thus, it would appear that despite the increase in both the number of elements per analogy term as well as the number of transformations between analogy terms, the identical two strategy types dominate. This would appear to suggest that, on the basis of subjective strategy reports, for Semidegenerate (A:A::B:B) forced choice geometric analogy items, increases in item complexity fail to elicit

meaningful strategic variations on the part of subjects in order to overcome increases in mental workload and maintain high standards of task performance. Thirdly, on the basis of subjective task strategy reports, variations of the dominant strategy types for both 1 element - 1 transformation and 3 element - 3 transformation Semidegenerate (A:A::B:B) item solution were identified. As noted earlier in this section these variations apparently arose as a result of subjects employing either a self-terminating or exhaustive application process. Three variations of both dominant strategy types pertaining to 1 element - 1 transformation Semidegenerate (A:A::B:B) item solution were identified. These variations were discussed previously in this section. Similarly, variations of both dominant strategy types pertaining to 3 element - 3 transformation Semidegenerate (A:A::B:B) item solution were identified. Again, these variations were discussed previously in this section. It is, however, important to note that subjects' post hoc introspective reports pertaining to both 1 element - 1 transformation and 3 element - 3 transformation Semidegenerate (A:A::B:B) item solution revealed that order of answer options, i.e., which answer option was correct (A or B), was a determining factor of the type of application process employed. Where the first answer option was correct, subjects reporting this strategy type variation would employ a self-terminating application process, and where the second answer option was correct, these subjects would report an exhaustive application process. This pattern was found to be more prominent in the subjective task strategy reports relating to the solution of 1 element - 1 transformation Semidegenerate (A:A::B:B) items than those pertaining to the solution of 3 element - 3 transformation Semidegenerate (A:A::B:B) items.

Therefore, based on subjects' post hoc introspective reports regarding performance components adopted and order