AN EXPERIMENTAL STUDY OF HUMAN REASONING AND CONCEPTUAL BEHAVIOUR

JOHN E. TAPLIN B.Sc.(Hons.)

Department of Psychology,
University of Adelaide.
1971

CONTENTS

SUMMARY

The thesis begins by defining the meaning of concept and concept learning. A taxonomy of concepts is suggested, from which the family of nominal concepts is chosen for detailed investigation, and in particular those nominal concepts based on the rules of conjunction, inclusive disjunction, conditional, and biconditional. Two aspects of concept learning are distinguished: attribute identification in which the subject is given the rule relevant to the concept and required to find the attribute values which are combined by this rule; and rule learning in which the subject is given the relevant attribute values and set to determine the rule. An investigation of the effect of the rule on the difficulty of concept learning in each of these two tasks forms the major part of the thesis. This variable is considered alongside a number of other variables which have been found to influence performance in conceptual tasks.

A review of experiments conducted by other workers and an analysis of the results of experiments reported in this dissertation indicates that the effect of conceptual rule in attribute identification interacts with the type of instances encountered by subjects during learning. In particular, the type of initial instance seems to play

an important role in determining attribute-identification difficulty of different types of concepts.

A study of the process of concept identification is made beginning with the work of Bruner, Goodnow, and Austin (1956) on strategies. A critique of this work is put forward and alternative approaches considered. It is concluded that identification of strategies in concept attainment is not possible without recourse to subjective reports. Using this method an attempt to classify strategies following extensive training on attribute identification for four conceptual rules is made. The implications of the type of strategies employed for the effect of aids to memory on performance are considered in two experiments in which the form of the memory aid is varied.

Hypothesis-testing or concept-centred approaches to concept attainment are discussed in two parts - sampling and evaluation of hypotheses. Most research in the literature on hypothesis testing deals with sampling and a review of this work is made. On evaluation, the commonly accepted information-processing model is criticised and rejected and an alternative statistical decision model is postulated. Experiments investigating the number and type of instances selected in evaluating an hypothesis support the model proposed. The role of memory in hypothesis testing is considered and an experiment shows differential recall of various rules and also of positive and negative instances. These findings are interpreted as being due to differences in the ease of classifying instances as positive or negative for the different rules.

In rule-learning tasks a review of the literature suggests a stable order of rule difficulty except for one study which differs from other studies primarily in that the selection procedure is employed. An experiment using the selection procedure is designed to replicate this finding, but fails to obtain an order of rule difficulty which is any different from that typically found in other studies. Possible accounts for these rule differences are considered; the most probable account seems to be in terms of a strategic tendency towards conjunction. Implications of rule-learning and attribute-identification studies for complete concept learning, in which both rule and attribute values are unknown, are assessed. The usefulness of the findings from concept learning studies for explicating problem solving is also discussed, with special reference to the solving of mathematical groups (Jeeves, 1968).

This work has essentially been concerned with aspects of inductive reasoning. Finally, a look at deductive reasoning is taken and the applicability of propositional logic as a model of human performance considered. Several discrepancies between the logical model and performance are noted which suggest modifications required in the model. The problem generally with such a model seems to be to find linguistic expressions which are equivalent in meaning to the logical operations.

The thesis concludes by considering some general directions which future research may follow. The implications discussed are taken from a consideration of the role of rules in accounts of reasoning and conceptual behaviour.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any University, and to the best of my knowledge and belief, it contains no material previously published or written by another person, except when due reference is made in the text.

(Signed)

ACKNOWLEDGEMENTS

The research reported in this thesis was carried out in the Department of Psychology at the University of Adelaide from 1968 to 1971 under a Commonwealth Postgraduate Award. During this time I have had three supervisors whose guidance and encouragement is gratefully acknowledged: Professor M. A. Jeeves, from January, 1968 to August, 1969, who initially aroused my interest in the area; Dr. D. McNicol, from August, 1969 to February, 1971, who provided valuable assistance in the use of the Department's PDP 8/L computer for analyzing the data; and Dr. I. D. John, from February, 1971 to June, 1971, who had the unenviable task of reading the initial draft of the thesis. In addition to these people I also wish to thank: Bill Gould and Inara Proske for their comments on the last two chapters; Sue Thom and Christine Symonds for drawing the figures; Sue Williams for writing the analysis of variance program for the CDC 6400 computer; Gerard Barrio for constructing the apparatus for Experiments XIII and XIV; Peter Young for his assistance in conducting Experiments XIII and XIV; and Catherine Burden for typing the thesis. To these people and to the many people who acted as pilot or experimental subjects I am indebted.

From the research described here two papers have been accepted for publication. These articles are "Effect of initial instance on the attribute identification of concepts using the selection procedure", which is to appear in the *Journal of Experimental Psychology*, and "Reasoning with conditional sentences", which will be published by the *Journal of Verbal Learning and Verbal Behaviour*. Several helpful comments on the experiments reported in these papers were made by the Consulting Editors of these Journals.

CHAPTER 1

AN INTRODUCTION TO A STUDY OF CONCEPTUAL BEHAVIOUR

Concepts and concept learning

In experimental psychology the word "concept" has a somewhat different meaning from its everyday connotation. Ordinarily we may use "concept" to refer to an idea, a mental image, or an abstract state of affairs. In psychology we say that "a concept exists whenever two or more distinguishable objects or events have been grouped or classified together and set apart from other objects on the basis of some common feature or property characteristic of each" (Bourne, 1966). For example, the concept "Ph.D student" includes those people enrolled at a university, who are candidates for a postgraduate degree where that degree is the degree of Doctor of Philosophy. It therefore excludes all people not enrolled at a university, all university students not taking postgraduate studies, and all postgraduate students not candidates for the degree of Doctor of Philosophy. A concept is therefore a basis for grouping or classifying objects or events together.

The ability to conceptualize confers several advantages on an organism (Bruner, Goodnow, and Austin, 1956). Firstly and most importantly by categorizing as equivalent discriminably different events, the individual is able to reduce the complexity of his environment. Workers in a number of different areas of psychology have pointed to a principle of economy in much of human performance. For example, Attneave (1954) has suggested that"a major function of the perceptual machinery is to strip away some of the redundancy of stimulation, to describe or encode incoming information in a form more economical than that in which it impinges on the receptors." Likewise, Oldfield (1954) and Miller (1956) have observed that many messages would be beyond our capacity to receive, process, and remember if they were not recoded into a more economical form. It should be noted that the act of categorizing is not the only method available to the organism for reducing the complexity of its environment. Welford (1968), Berlyne (1965), and Posner (1964) have all referred to two processes which have this effective outcome: one of these is conceptualizing or coding or abstraction; the other is selective attention. Welford puts it this way: "The process of organizing and grouping incoming data may be thought of as the abstraction of constants from the total mass of data presented in space and over time, together with the selection of some data as dominant and important while the rest are relegated to the background and more or less neglected."

Other benefits conferred on an organism by the ability to form concepts follow from this principal advantage. Thus, a second advantage is that it provides the means by which the objects of the world about us are identified. Thirdly, the establishment of a concept based on a set of defining attribute values reduces the necessity of constant learning. Fourthly, it provides the direction for instrumental activity. For example, to know that a substance is "poison" is to know in advance about appropriate and inappropriate actions to be taken. Finally, categorizing permits the opportunity for the ordering and relating of classes of events due to overlapping between classes as, for example, between "smoke" and "fire".

From the foregoing discussion, it should already be apparent that an understanding of the way in which organisms form and use concepts will have application to a wide range of behaviour. The particular organism selected for study in this thesis is man. This is not meant to indicate that lower organisms are unable to form concepts, even though their ability to do so has been doubted by some writers (Osgood, 1953; Hunt, 1962). Indeed, the existence of a strict discontinuity between man and other animals has not been satisfactorily demonstrated. Most studies in which animal subjects have been used to categorize a set of stimuli have been placed under the heading of discrimination learning. Kintsch (1970) has pointed out that the operational definitions of stimulus generalization in discrimination learning experiments and of concept in concept learning experiments are identical, namely, the same response is made to a set of discriminably different stimuli. On this definition, animals can certainly learn concepts. One early attempt to compare conceptual behaviour as we ascend the phylogenetic scale was made by Kendler and Kendler (1962). Generally, however, little comparative research has been carried out in this area. Therefore, whilst recognizing that corresponding behaviour may be found in lower animals, this thesis will be devoted entirely to the study of conceptual processes in humans.

One aspect of human behaviour with which conceptual behaviour is frequently associated is thought. The fact that there is no generally accepted definition of thinking makes it rather difficult, if not impossible, to determine the precise nature of the relationship between the two behaviours. Nevertheless, three kinds of connections have been postulated. The first is Harre's (1966) assertion that "concepts are the vehicles of thought". That is to say, thinking consists of the linking together of a number of concepts. Presumably, an understanding of the psychological characteristics involved in acquiring and using concepts should therefore contribute towards an understanding of thinking. To date, this kind of notion has stimulated very little research. This does not necessarily mean that an approach based on it might not throw important light on the nature of thinking. Rather it could be taken to mean that present knowledge of conceptual processes is inadequate. A second connection which has received considerably more attention postulates that the process of attaining a concept involves thought. Thus, by analyzing the pattern of responses made by a subject seeking to attain a concept, inferences about the nature of thinking in this situation may be drawn. Regularities observed in responding are called strategies and some considerable effort has gone into identifying them (e.g., Bruner, Goodnow, and Austin, 1956). A third proposal comes from Bourne (1969) who argues that "any particular behaviour ... is recognizably consistent with and instantiates a rule". Bourne suggests that the highly organised nature of behaviour is an indication that such behaviour is rule-following, irrespective of whether or not the individual is aware of the rule. Moreover, he suggests that the study of conceptual behaviour presents a fruitful method of discovering some of these rules and the difficulty subjects have in learning and using them. Since many of these same rules have been identified in human thought by logicians, the findings from concept experiments may prove useful in accounting for thinking behaviour. This proposal is offered by Bourne as an alternative to underlying process descriptions of behaviour, which form the second postulated connection between conceptual and thinking behaviour. No attempt will be made here to make a decision as to which alternative is likely to prove the most fruitful. Discussion of both of these approaches will be contained in later chapters and it will become evident then which is preferred.

A taxonomy of conceptual groupings

Just as the number of objects and events which may be grouped to form concepts is infinite, so therefore is the number of concepts able to be formed also infinite. Before any feasible approach to an understanding of conceptual processes becomes possible, it is first of all necessary to attempt some classification of concepts themselves.

According to Harre (1966), philosophers have employed two systems for classifying concepts - categorial and hierarchical. The categorial taxonomy originated with Aristotle, who assigned concepts to various categories such as Substance, Attribute, Quantity, Quality, and Relation, depending on the kinds of questions which might properly be asked about anything. Kant devised a different, but still categorial form of classification. In his version, it is the kinds of propositional forms or judgments that are possible on which the taxonomy ofconcepts is based. The hierarchical classification, on the other hand, seeks to arrange concepts into different levels depending on their explanatory power. Thus, higher levels consist of concepts which are based on other concepts from a lower or more fundamental level.

The most appropriate method for classifying concepts in conceptual tasks seems to involve a conbination of these two systems. Essentially, this approach distinguishes between the relevant stimulus dimensions and the rule relating these stimulus dimensions, and attempts to devise some scheme for placing different kinds of rules into different categories. Within each of these categories, a number of rules appear and these may be found to be related in an hierarchical manner.

Adams (1953) has initiated such a taxonomy with his suggestion that concepts may be classified in a similar manner to the way in which Stevens (1951) classified measurement scales. Depending on the relationship between values on the relevant stimulus dimensions, concepts may be nominal, ordinal, interval, or ratio. Thus, nominal concepts are formed according to a rule for arranging observations into equivalence classes, such that observations falling into the same class are thought of as qualitatively the same and those in different classes as qualitatively different in some respect or respects. Nominal concepts may be simple, i.e., having only one feature in common, or complex, i.e., having more than one feature in common. Thus, for example, "triangle" is a simple nominal concept defined by all 3-sided shapes. On the other hand, "red triangle" is a complex concept specifying the joint presence of two characteristics, the colour red and the number of sides three. Ordinal concepts are defined by ordered relationships between the stimulus attributes. For example, an "isosceles triangle" is an ordinal concept in which two of the sides of the triangle have the same length. Similarly, "an obtuse triangle" is defined by the fact that one angle is greater than 90^{o}. Interval and ratio concepts consist of mathematical or functional expressions which prescribe some numerical response from a given set or sets of numerical stimulus values. An interval concept differs from a ratio concept according to the particular scale of measurement on which its stimulus and response values can be located (see Stevens (1951) for the distinction between interval and ratio scales).

The present investigation, like most studies of human conceptual behaviour, will be concerned with nominal concepts. Relational concepts have received some attention in recent years in the work of Hunter (1957), Huttenlocher (1968), DeSoto, London, and Handel (1965), and Clark (1969) on three-term series problems. Mathematical concepts have been less systematically considered but preliminary studies have been conducted by Uhl (1963), Restle (1970), and Scandura (1970).

Nominal conceptual rules

Within the family of nominal concepts further classification is possible. For concepts having up to two relevant attribute values, ten different logical rules for grouping stimuli may be identified (Hunt, 1962; Neisser and Weene, 1962). To illustrate these rules, consider a well-defined, multi-dimensional stimulus population in the form of an array of geometrical designs characterized by variations in shape (square, circle, or cross), colour (red, green, or black), number of figures (one, two, or three), and number of borders (none, one, or two). Suppose two of these attribute values, say, redness and squareness, are arbitrarily designated as relevant to the concept. For nominal concepts, stimuli are divided into two, those which are members of the concept, called positive instances, and those which are not members of the concept, called negative instances, i.e., there is a binary response system.

Sixteen binary partitions of this stimulus population are possible. Of these, two are trivial because they place the entire population into either the positive or the negative category. In addition, there are four pairs in which the partitioning of the stimulus population for each pair follows the same rule, with only a change in the attribute values distinguishing each member of the pair. Ten different, non-trivial rules remain and these fall into five complementary pairs having the property that any instance which is positive for one member of the pair is negative for the other. A summary of these rules is given in Table 1 in the form of a truth table, which shows the state of each rule on each of the four possible stimulus contingencies. The truth (T) or falsity (F) of a rule signifies a positive or negative instance respectively; the truth or falsity of an attribute value signifies the presence or absence of that attribute value in the stimulus.

It will be noticed in Table 1 that each of the ten rules has been given a symbolic description. These descriptions are derived from symbolic logic in which they represent connectives having the same truth-table function. In logic, these connectives have been assigned names and conventional verbal expressions have been adopted to describe their operation. These same descriptions have been employed in studies of concept learning and a summary in terms of our two relevant attribute values, redness and squareness, may be found in Table 2.

Also in Table 2, alternative logically equivalent expressions have been provided for some of the rules with the aid of DeMorgan's laws (Jeffrey, 1967). Using these laws, Neisser and Weene (1962) have suggested that rules may be hierarchically classified such that single attribute value concepts make up the lowest level, with higher levels being formed firstly, from the supposedly fundamental operations of conjunction, inclusive disjunction, and negation, and secondly, combinations of these fundamental rules. It should be understood here that these same laws make it possible also to form different hierarchies. For example, it is possible to adopt only conjunction and negation as fundamental operations so that higher levels are composed of combinations and multiplications of these rules. Whichever hierarchy should be accepted is an empirical question and more will be said on this later. For the present, it will suffice to observe that concepts may be hierarchically organised within a category.

State of attributes		State of conceptual rule									
R	S	Basic rule					Complementary rule				
		R	$R \wedge S$	$R \cup S$	$R \to S$	$R \leftrightarrow S$	$\overline{R}$	R/S	$R \downarrow S$	$R \wedge \overline{S}$	$R \overline{\cup} S$
T	T	T	T	T	T	T	F	F	F	F	F
T	F	T	F	T	F	F	F	T	F	T	T
F	T	F	F	T	T	F	T	T	F	F	T
F	F	F	F	F	T	T	T	T	T	F	F

TABLE 1 Truth table for ten different rules.

This dissertation aims to investigate some of the psychological differences between these rules when subjects are required to learn concepts involving them. Before such an investigation can be properly undertaken, however, an appreciation is necessary of some of the variables present in a conceptual task. These will be discussed in Chapter 2.

BASIC RULE			COMPLEMENTARY RULE		
Name	Symbolic Description [1]	Verbal Description	Name	Symbolic Description	Verbal Description
Affirmation	R	All red patterns are examples of the concept	Negation	$\overline{R}$	All patterns which are not red are examples
Conjunction	$R \wedge S$	All patterns which are red and square are examples	Alternative Denial or Disjunctive absence	$R \mid S$ $[\overline{R} \cup \overline{S}]$	All patterns which are either not red or not square are examples
Inclusive disjunction	$R \cup S$	All patterns which are red or square or both are examples	Joint Denial or Conjunctive absence	$R \downarrow S$ $[\overline{R} \wedge \overline{S}]$	All patterns which are neither red nor square are examples
Conditional	$R \rightarrow S$ $[\overline{R} \cup S]$	If a pattern is red, then it must be a square to be an example	Exclusion	$R \wedge \overline{S}$	All patterns which are red and not square are examples
Biconditional	$R \leftrightarrow S$ $[(R \wedge S) \cup (\overline{R} \wedge \overline{S})]$	Red patterns are examples if and only if they are squares	Exclusive disjunction	$R \overline{\cup} S$ $[(R \wedge \overline{S}) \cup (\overline{R} \wedge S)]$	All patterns which are red or square but not both are examples

1 R and S stand for red and square (relevant attributes), respectively. Symbolic descriptions using only three basic operations, $\wedge$, $\cup$, and negation, are given in brackets:

TABLE 2 Conceptual rules with their symbolic and verbal descriptions (From Haygood and Bourne, 1965).

CHAPTER 2

VARIABLES IN CONCEPTUAL TASKS

Essentially, the experimental paradigm for the study of conceptual behaviour involves the presentation of a set of discriminably different stimuli to the subject with the requirement that he make one of a fixed number of discrete classificatory responses for each stimulus in the set. Since the number of responses available is always less than the number of stimuli in the set, the subject must respond to different stimuli in the same manner. The task is to discover the rule according to which responses are assigned to stimuli. A number of variables have been found to influence the subject's performance on this task.

Stimulus material

Probably the most commonly used stimulus materials in conceptual tasks are geometrical designs, typically printed or drawn on cards or photographed. Ordinarily, these vary along well-defined dimensions such as colour and shape, each dimension consisting of a fixed number of levels or values. The prevalence of their use in experimental studies may be due to their simplicity, familiarity, and highly dimensionalized nature (Bourne, 1966), but no study has systematically compared them with other kinds of materials less frequently employed.

Among other stimulus materials used are strings of letters (e.g., Neisser and Weene, 1962) in which each position in the string may have any one out of a specified pool of consonants.

Other studies have utilized less abstract, more concrete materials. One such instance comes from the experiments of Underwood and Richardson (1956 a; 1956 b) who required subjects to classify the names of common objects by some adjective, descriptive of sensory impressions aroused by the nouns. For example, "snow", "tooth", "chalk", and "milk" may be grouped together by the adjectival concept, "white". Using verbal materials introduces an additional variable, namely, the strength of the association between the stimulus noun and the adjective response. Underwood and Richardson have calibrated a restricted list of nouns and associated adjectives in terms of this variable which they have called response dominance. This calibrated material was then used to explore a possible relationship between dominance and speed of concept learning. As one would expect, it was found that the higher the dominance, the greater was the degree of learning. Coleman (1964) has since replicated this finding. It would seem that response dominance provides a measure of the strength of preestablished conceptual relationships.

A comparison between abstract and thematic stimulus materials has been made by Bruner, Goodnow, and Austin (1956).

The nature of the thematic material was such that it could be easily dimensionalized and that it should be suggestive with meaning, "evocative of a little story or theme". The result of using this form of stimuli was seen in that the subject tended to adopt hypotheses about the possible groupings which seemed reasonable in the light of the evoked theme. Moreover, with thematic material more than with abstract material, certain attributes tended to take on nonrational importance so that the subject's hypotheses seemed to be formulated around them.

Stimulus presentation procedures

Two different procedures exist for presenting stimuli in conceptual tasks: a reception procedure and a selection procedure. With the reception procedure, instances of the concept are usually presented successively to the subject in an order which has been predetermined by the experimenter. After each instance is presented, the subject is required to categorise it as positive (i.e., an instance included by the definition of the concept) or negative. Informative feedback is then provided by the experimenter on whether the subject's classificatory response was correct or not. Then the stimulus is removed, a new one presented, and the same process repeated. Some variations from this basic method are sometimes used. One is to employ a simultaneous presentation method in which subjects are permitted to see all stimulus patterns at once instead of successively. This has the effect of simplifying the task, because information from previous trials does not have to be retained in memory (Hovland and Weiss, 1953). Another is to require the subject to make an additional response to categorizing the presented stimulus. In this case the subject is asked to state a reason or hypothesis for the particular category chosen on each trial. This has the advantage of yielding more information on the subject's behaviour during the task as well as providing a second measure of overall performance. Some of the relationships between category responses and hypotheses have been indicated by Bourne (1965).

With the selection procedure, all instances are usually displayed simultaneously. Typically the task begins by the experimenter designating one member of the stimulus population as a positive instance of the concept which must be discovered. From this point all instances are selected by the subject himself so that the sequence of instances is under his control. As each instance is chosen, the experimenter indicates whether it is positive or negative. When the subject considers that he has identified the concept, he verbalizes it to the experimenter. If his hypothesis is correct, the problem is considered solved; if incorrect, the procedure is continued until the concept has been verbalized. In a number of experiments employing a selection procedure, the subject is required to make an hypothesis after each instance. In this way more information may be gained concerning the subject's behaviour during the task. Requiring subjects to hypothesize on every trial has been shown by Byers and Davidson (1967) to facilitate concept identification, however. Unless data on hypothesizing behaviour is specifically needed, it would seem inadvisable to introduce this additional variable when using a selection procedure.

Several studies have sought to compare efficiency of concept identification using the two procedures yoked so that each subject on the reception procedure receives the same set of stimuli in the same order as selected by another subject on the selection procedure. Unfortunately, no definite conclusions are possible about the relative difficulty of concept attainment using these procedures. Two studies (Huttenlocher, 1962a; Murray and Gregg, 1969) favour better performance on the reception procedure, one (Hunt, 1965) favours better performance on the selection procedure, and two suggest interaction effects between the type of procedure employed and conceptual rule (Schwartz, 1966), and between the procedure and amount of stimulus variability (Laughlin, 1969b).

Schwartz (1966) reported that the selection paradigm gave superior performance for conjunctive concepts, while disjunctive concepts were solved more easily using the reception paradigm. His conclusions seem open to objection for at least four reasons. Firstly, Laughlin (1969b) did not obtain any significant Paradigm X Rule interaction. Secondly, since no post-hoc analysis is given, it is impossible to know that the interaction obtained by Schwartz is due to the yoked selection and reception procedures. Rather, it seems more likely that the effect is attributable to the relative difficulty of the random condition which was also conducted for each rule. Thirdly, the time to attainment measure for disjunction is in the direction of easier concept attainment with a reception procedure, but a trials to attainment measure on this same rule is in the reverse direction. Fourthly, the response measures for disjunction seem unreliable, anyway, since they are based only on those subjects who actually attained the concept. More than half the subjects failed to achieve this objective within the limit of one hour or 64 trials.

In Laughlin's (1969b) experiment stimulus variability was manipulated using a six two-valued attribute display and a four three-valued attribute display. Laughlin found that the reception paradigm was more difficult than the selection paradigm using a 4 - 3 stimulus universe, but less difficult using a 6 - 2 universe. Unfortunately, the two stimulus populations confound interdimensional and intradimensional variability, so we are unable to know whether either one or both of these sources of irrelevant information interacts with the procedure. Other experiments have utilized different amounts of both sources of variability. Thus Huttenlocher (1962a) used two (or three) attributes each with two values, Hunt (1965) four attributes with four values each, and Murray and Gregg (1969) two four-valued attributes. If both inter- and intra-dimensional variability are important, then this would suggest that the total number of stimuli is the factor influencing the relative difficulty of the selection and reception paradigms. This hypothesis does divide the experimental findings into two groups. For small numbers of stimuli, e.g., 8 (Huttenlocher, 1962a), 16 (Murray and Gregg, 1969), and 64 (Laughlin, 1969b), easier concept identification is obtained using a reception procedure rather than a selection procedure. For larger numbers of stimuli, e.g., 81 (Laughlin, 1969b), and 256 (Hunt, 1965), easier concept identification is obtained using a selection procedure rather than a reception procedure. Further work is necessary to confirm this hypothesis.

Alternative hypotheses may suggest a role for the way in which stimuli are presented, that is, either successively or simultaneously. However, there is no indication that this factor can be used to discriminate between discrepant results. Moreover, Laughlin's (1969b) failure to find an interaction between paradigm and a memory aid condition tends not to support such an hypothesis, since memory load has been cited as an essential difference between simultaneous and successive presentation methods (Hovland and Weiss, 1953).

Stimulus variability

The stimulus population in most concept problems is characterized by a set of dimensions, each of which has two or more values. Some of these dimensions are relevant to problem solution in the sense that they define the concept to be identified. Other dimensions are irrelevant and can be ignored. Moreover, within dimensions values may also be relevant or irrelevant depending on whether they may be used to specify the concept. A number of experiments have sought to examine the effects of the amount of relevant and irrelevant information on performance.

1. Amount of irrelevant information

a. Due to interdimensional variability

Using visually presented stimuli, many workers have shown that task difficulty measured by trials, time, and errors to solution increases linearly as the number of irrelevant dimensions increases (Brown and Archer, 1956; Bourne, 1957; Bourne and Pendleton, 1958; Battig and Bourne, 1961; Lordahl, 1961; Walker and Bourne, 1961; Bourne, Guy, Dodd, and Justesen, 1965; Haygood and Stevenson, 1967). The only experiments to report a nonlinear relationship were by Archer, Bourne, and Brown (1955), who obtained both significant linear and quadratic trends for increasing difficulty with increasing amount of irrelevant information, and Byers and Davidson (1968), who were unable to find a significant linear trend although the means for the three levels of irrelevant information were ordered in the expected direction.

With aurally presented stimuli, Bulgarella and Archer (1962) obtained a linear relationship similar to that found using visual stimulus material. Lordahl (1961), on the other hand, found no effect on performance with increasing amounts of irrelevant auditory information. Keele and Archer (1967) have attributed Lordahl's failure to find any effect either to the type of auditory material (pure tones) he used or to the fact that less than half the subjects identified the concept within the fixed number of trials (192 trials) presented. It seems doubtful that the former explanation can account for Lordahl's results since Bulgarella and Archer also utilized pure tones as auditory material. Discarding those subjects who had not reached the criterion for attaining the concept after 384 trials, and using sentences as auditory material, Keele and Archer were able to demonstrate a clear linear correspondence between difficulty and the number of irrelevant dimensions.

b. Due to intradimensional variability.

Battig and Bourne (1961) were the first to systematically explore the effects of intra-dimensional variability, using stimulus material having 2, 4, or 6 values per dimension. The solution in each case involved two relevant dimensions. With two-valued dimensions the subject was required to combine one value from each relevant dimension to solve the problem. With six-valued dimensions three values from each relevant dimension had to be combined. Using this procedure, Battig and Bourne found a linear decrement in performance with increasing number of values on each dimension. On the basis of this result and also the observation of a linear decrement in performance with an increasing number of irrelevant dimensions, Battig and Bourne suggested tentatively that both types of irrelevant information, inter- and intra-dimensional, might be considered in the same way theoretically.

The procedure employed by Battig and Bourne necessitates some caution in accepting this possibility. In increasing the number of values on each dimension, these workers have increased the number of relevant values per relevant dimension as well as the number of irrelevant values on each dimension. In other words, not only had the amount of irrelevant information been increased, but also the amount of relevant information.

Haygood, Harbert, and Omlor (1970) have followed a more suitable procedure for assessing the effects of irrelevant intradimensional variability independent of relevant intradimensional variability. Their method was simply to select only one value in each case from the particular dimension chosen to define the concept. They found that, instead of an increase in difficulty with increasing numbers of values on each dimension, there was a decrease. This result suggests that the two types of irrelevant information are not alike in their effects on performance, in contrast to Battig and Bourne's suggestion. The relationship obtained by Haygood et al. between improvement in performance and increasing irrelevant intradimensional variability was positively decelerating, such that changes in performance as the number of values per dimension became larger seemed less likely to be significant.

2. <u>Amount of relevant information</u>

The evidence pertaining to the effect on problem difficulty of the amount of relevant information is complicated by confounding factors which have been present in some experiments. The basic finding is that difficulty increases linearly with the number of relevant dimensions. This result has been obtained by Byers and Davidson (1968) and Looney and Haygood (1968) using visually presented stimuli, and by Bulgarella and Archer (1962) using abstract auditory signals. In an experiment in which the number of relevant variables was confounded with the number of response categories, Walker and Bourne (1961) reported an accelerating increase in difficulty with increases in the number of relevant dimensions. Bourne (1966) has attributed the exponential relationship to the joint effects of the two variables confounded. Another study (Glanzer, Huttenlocher, and Clark, 1963) confounded the numbers of relevant and irrelevant dimensions in such a way that as the number of relevant dimensions increased, the number of irrelevant dimensions decreased. From the results of their experiment, Glanzer et al. concluded that the most difficult concept was the one requiring an equal division of dimensions into relevant and irrelevant. Laughlin (1968a) also confounded the number of relevant and irrelevant dimensions in an experiment investigating conditional concepts. Like Glanzer et al., he reported a curvilinear effect on performance as the number of relevant dimensions increased, but in his experiment the ratio of the number of relevant dimensions to the total number of stimulus dimensions for the most difficult concept was slightly more than a half.

3. Stimulus redundancy

Redundancy exists in a stimulus population whenever the actual number of stimulus patterns is less than the number that could be generated from the given stimulus variables, that is, the number of dimensions and the number of values on each dimension. For example, suppose all large patterns are red and all small patterns are green. It is clear in this case that the actual number of patterns will be less than is theoretically possible since there are no large green and no small red patterns. We may therefore say that this population contains some redundancy. In this example the dimensions of colour and size are perfectly correlated. Thus, the information contained in these two dimensions is completely redundant. By varying the number of dimensions correlated in a stimulus population, the amount of redundancy in the population is changed.

Bourne and Haygood (1959, 1961) have conducted several experiments investigating the effects of amount of stimulus redundancy on concept identification. When redundancy was limited to relevant stimulus dimensions, performance improved as redundancy increased. This effect became more pronounced as the number of nonredundant irrelevant dimensions became larger. When the redundancy was limited to irrelevant dimensions, a decrement in performance resulted with increasing redundancy. However, this decrement in performance from adding redundant irrelevant dimensions was not as great as that obtained by adding an equivalent number of nonredundant irrelevant dimensions.

Garner (1962) has suggested that form as well as amount of redundancy may be an important variable. By form of redundancy he is referring to the particular stimulus patterns in a population, emphasizing that it matters not only how many possible patterns have been excluded but also which ones. The form of redundancy utilized by Bourne and Haygood (1959, 1961) consisted of a perfect correlation between the levels of two or more stimulus dimensions. Haygood and Bourne (1964) sought to compare this form of redundancy (Form A) with a second form (Form B) in which the levels of one dimension were contingent upon a combination of levels within two or more other dimensions. For problems with the same amount of relevant redundancy, Haygood and Bourne found performance using Form A redundancy superior to Form B, confirming Garner's expectation of the importance of this variable. Both experimental groups made fewer errors than two control groups in which the redundant dimension was either eliminated or converted to nonredundant irrelevance. No interaction effect on performance was evident between the amount and the form of redundancy. Furthermore, the interaction between form of redundancy and the number of irrelevant dimensions was also nonsignificant, in contrast to the earlier finding of Bourne and Haygood for amount of redundancy.

The above studies on the effects of redundancy have utilized visual stimuli in the form of geometrical designs. A recent study by Keele and Archer (1967) also using visual material has replicated the findings on the effects of amount of redundant relevant information, except for the significant interaction between amount of redundancy and amount of irrelevant information reported by Bourne and Haygood. With auditory stimuli a surprise result was the failure to find an effect of redundant relevant information on performance in one experiment, although in a second experiment some improvement was observed.

Salience of stimulus information

Several studies have demonstrated a role for the salience or attention value of stimulus information in determining the difficulty of a conceptual task. Mention has already been made of the importance of concrete or thematic stimulus materials on the hypotheses likely to be entertained by the subject. Evidence can also be drawn from studies employing abstract geometrical patterns that some dimensions are more salient than others. For example, Brown and Archer (1956) found concepts having spatial location as the relevant dimension to be more difficult than others involving form or colour. The effect of the salience of stimulus information has been shown by Archer (1962) to depend on the relevance or irrelevance of that information. Thus, performance is facilitated when relevant stimulus information is more salient, and impeded when irrelevant information is more salient.

A number of methods exist for manipulating this variable experimentally. One method has already been referred to in the earlier discussion of meaningful or concrete stimulus material in concept identification. Another approach is to emphasize the relevant dimension by the addition of a redundant relevant dimension. For example, Hull (1920) attempted to increase the attention value of the relevant attribute in the Chinese characters used as stimuli by colouring it. A third way of controlling salience concerns the discriminability of relevant and irrelevant information. Baum (1954) showed that when several concepts are to be learned concurrently, those for which the instances are least discriminable from each other and most discriminable from instances of other concepts are the easiest to learn; conversely, concepts containing highly discriminable instances which bear some similarity to other concepts are most difficult to learn. Support for this finding also comes from Shepp and Zeaman (1966) and Jacoby and Radtke (1969). Another factor affecting the salience of stimulus information is the organization of the stimulus display. Two examples will be given. The first comes from an experiment by Bruner, Goodnow, and Austin (1956) comparing random stimulus arrays with systematically arranged stimulus arrays in which adjacent stimuli differed by only one attribute value. They found performance to be significantly better with the orderly array than with a random array. The second concerns the embodiment of all stimulus attributes in each stimulus pattern or over several spatially separated patterns. An experiment by Shepard, Hovland, and Jenkins (1961) revealed better performance on compact than on distributed stimulus displays. Increased salience of stimulus dimensions may also be induced through preliminary instructions or pretraining. In a review of studies of shift performance in human concept identification, Wolff (1967) showed that intradimensional shifts are easier generally than extradimensional shifts, and this difference becomes more marked with overtraining on the preshift problem. Thus, it would appear that any pretraining which involved the relevant dimension is likely to have a facilitative effect on subsequent performance, presumably because the subject is then attending to the relevant dimension. A slightly different type of pretraining having a similar effect was used by Gelfand (1958), who found that preliminary experience with words associated with relevant stimulus attributes facilitated performance relative to neutral or irrelevant pretraining.

Stimulus sequence

An hypothesis proposed by Underwood (1952) has provided the stimulus for most experiments in recent years dealing with the effects of differing sequences of instances on concept learning. Underwood's postulate was that, in general, in order for relationships between stimuli to be perceived and acquired, responses to those stimuli must occur contiguously. Specifically, if instances of other concepts are interpolated between instances of the same concept, then contiguity is reduced between instances of the particular concept concerned and so acquisition of the concept will be more difficult.

This prediction has received support from a number of studies. Kurtz and Hovland (1956) required subjects to learn four concepts using either a mixed sequence of instances of each of the four concepts or an unmixed sequence. Performance on concept identification was better for the unmixed than the mixed order. Newman (1956) compared two mixed orders of presentation which differed in the average number of interpolated stimuli shown between instances of the same concept. Significantly better performance was obtained when instances of the same concept were closer together. Three levels of instance contiguity were manipulated by Schulz, Miller, and Radtke (1963) using six verbal concepts and the results indicated concept identification difficulty to be directly related to level of contiguity. Bourne and Jennings (1963) investigated four degrees of contiguity among instances of four different concepts. They observed a linear improvement in performance with increased contiguity between instances of each concept. Similar evidence in support of the hypothesis comes from Peterson (1962).

A somewhat different experimental situation involving a series of positive and/or negative instances of a single concept may also be regarded as being relevant to Underwood's hypothesis of instance contiguity. A sequence of all positive instances is one of maximum contiguity, whereas a mixed sequence of positive and negative instances has less contiguity. Experiments by Hovland and Weiss (1953) and Whitman and Garner (1963) have demonstrated better concept learning using all positive instead of mixed positive and negative instances. A recent study by Bourne and Guy (1968b), however, showed that when there are more positive instances than negative instances in the stimulus population as for the conditional rule, less difficulty is found with a mixed sequence than an all positive sequence. In this case, though, a sequence composed of all negative instances proved easier than both mixed and all positive sequences. An exception to these findings comes from an experiment by Huttenlocher (1962b) who found a mixed series of positive and negative instances was easier than both all positive and all negative series for one dimensional concepts. This result may be related to the fact that the series used contained only two instances, suggesting that longer sequences are necessary for contiguity to have an effect.

Not only may contiguity of instances be varied by means of the positive - negative sequence, but also by presenting a series of positive instances of one type and then another. This method has been used by Haygood, Sandlin, Yoder, and Dodd (1969) with disjunctive concepts utilizing positive instances having only one relevant attribute value. Three different sequences were tested, each with the same order of positives and negatives, but varying in the number of positive instances of a particular type presented before alternating to the other type of positive instance. The results once again supported the contiguity hypothesis in that fewest errors were made at the highest contiguity level, that is, when the number of instances of the same type presented consecutively was greatest.

An alternative account of stimulus sequence effects has been provided by Detambel and Stolurow (1956) in terms of the effect of sequential manipulations of relevant and irrelevant stimulus dimensions. They proposed that concept identification is most efficient when (a) the same relevant cue is presented on adjacent trials and irrelevant stimulus components changed, or (b) the relevant cue is varied and irrelevant stimulus components held constant. Using these rules, Anderson and Guthrie (1966) obtained results which indicated that a special type of mixed series of instances, namely, an alternating series, produced fewer errors during training than a constant series, contrary to the expectation of Underwood's instance contiguity hypothesis. An interaction effect on both training trials and test trials between the type of series (alternating or constant) and the number of stimulus dimensions changing from trial to trial (1 or 3) was interpreted as providing support for Detambel and Stolurow's rules for a "good" stimulus sequence. For the alternating series, the group in which 1 stimulus dimension changed between each pair of adjacent training trials showed fewer errors than the group in which 3 dimensions changed. For the constant series, on the other hand, fewer errors were made by the group in which 3 dimensions changed compared with the group in which 1 dimension changed.

Peterson (1968) and Dominowski (1968) have independently made the observation that in many investigations of stimulus sequence effects the order of stimulus presentations has been confounded by testing order. This fact makes it possible that the findings of these experiments may be due simply to response bias effects. That is, subjects may learn to make few errors by learning the sequential constraints in the stimuli presented instead of learning the concept. Where test trials have been used not confounding order, the evidence in favour of contiguity is not so strong. For example, Kurtz and Hovland found no significant difference between the number of correctly categorized test stimuli following mixed and unmixed order of presentation. Their evidence supporting contiguity rests on more correct verbal descriptions of the concepts being made after an unmixed series compared with a mixed series. Similarly, the evidence against contiguity is also weakened, since Anderson and Guthrie's finding of better performance for alternating series than for constant series was obtained only on training trials. No difference between the two types of series was found on test trials.

Peterson and Dominowski have conducted experiments which enable stimulus sequence effects to be assessed in the absence of responding cues. Both have obtained results confirming the contiguity hypothesis. However, an attempt by Peterson to replicate Anderson and Guthrie's results was unsuccessful. Further work is necessary to determine the conditions, if any, under which Detambel and Stolurow's rules for stimulus sequences apply. Other aspects of stimulus sequences not considered in previous investigations also require examination. For example, the suggestion is contained in the work of Bruner, Goodnow, and Austin (1956) that the first instance presented in a conceptual task may have an important effect on subsequent behaviour on the task. Similar suggestions have also been made by Huttenlocher (1962b) and Glanzer, Huttenlocher, and

Clark (1963). In another context, the concept formed of a stimulus person has been shown to be strongly influenced by the kind of information first encountered (Luchins, 1957).

Positive and negative instances

Smoke (1933) was the first to suggest that the efficiency of concept attainment might be affected by presenting instances of what a concept is not rather than instances of what it is. He did not, however, evaluate the amounts of information logically conveyed by these negative and positive instances. It is not clear, therefore, from his data whether the relatively inefficient use of negative instances was due to the relatively small amount of information transmitted by them relevant to the concept or to difficulty experienced by subjects in assimilating negative information.

A more suitable method for comparing positive and negative instances was proposed by Hovland (1952). If the number of possible concepts of a particular type is specified by the stimulus population employed and these concepts can be listed, then the amount of information transmitted by a positive or a negative instance can be determined by the number of possible concepts logically eliminated by its presentation. Using this procedure for equating the informational content of positive and negative instances, Hovland and Weiss (1953) found that problems defined by all positive instances were easier to solve than those defined by all negative instances, with a mixture of positive and negative instances having intermediate difficulty. In a slightly different but related experimental task, Wason (1959, 1961) likewise indicated that negative statements took longer to process than positive statements, even when the amount of information conveyed by both statements was equated.

Later experiments have shown that the amount of information transmitted by positive and negative instances is also an important variable. In the experiments so far described in this section, only conjunctive or unidimensional problems were used, and in these cases the informational value per instance of positive instances is typically greater than of negative instances. With rules like the conditional, the reverse is true. Bourne and Guy (1968b) have demonstrated when conditional concepts are required to be identified that negative instances produce easier concept identification than either mixed positive and negative instances or all positive instances. For the four three-valued attribute stimulus population utilized by Bourne and Guy, an inclusive disjunctive concept yields an informational value per instance for negative instances which is slightly greater than for positive instances. On this type of problem, however, subjects required more trials to solution and made more errors with an all negative sequence than either a mixed or an all positive sequence, contrary to the expectations of the informational analysis. Apparently, the disparity in informational value between positive and negative instances in this case is insufficient to be able to compensate for ineffective handling of negative information.

Several accounts have been offered why negative information is found more difficult to process than positive. Freibergs and Tulving (1961) suggested that the greater difficulty of negative instances may be a reflection of the observation that positive instances are more frequently handled in everyday circumstances than are negative instances. Given sufficient practice at utilizing negative information to form concepts, they suggested that the differences between positives and negatives should disappear. An experiment examining this hypothesis indicated virtually no difference in difficulty in concept attainment between all positive and all negative instance sequences after practice on 20 problems.

Wason and Jones (1963) suggested that the connotations of the negative affect the times taken to process negative information more than its function as a logical constant. Using neutral signs (nonsense syllables) in place of positive and negative statements, the difference between the response times to the two types of sentences was significantly reduced. Some residual difference still remained, due perhaps to a tendency to translate negative information into positive form. Related to this is Eifermann's (1961) finding that the Hebrew word for "not" which is normally used in all contexts required longer processing time than the other Hebrew word for "not" which is normally used in all contexts except the prohibitive. Both negative statements, however, had longer response latencies than a positive statement. The importance of the connotation of the negative has also been demonstrated in a typical concept learning task by Seggie (1969) who found the difference between positive and negative instances disappeared when neutral symbols were used instead.

Perhaps a third account of the difficulty obtained when processing negative information comes from an assertion by Greene (1970) that the natural semantic function of the negative is to change the meaning of a prior affirmative statement, not to express the equivalent of an affirmative statement. To support this assertion Greene showed that negative statements which changed the meaning of an affirmative statement took less time to process than other negative statements which expressed the same meaning as an affirmative. A related finding is that when stimuli are described in terms of an exceptional item and a residual class, the response to negative statements is facilitated if these statements deny that the exceptional item possesses the property of the residual class

(Wason, 1965). On Greene's assertion, therefore, Wason's (1959, 1961) finding of longer responses to negative information than to positive information was because the negative was used in an unnatural way, to state the equivalent of the positive sentence. Applying Green's suggestion to conceptual problems, requiring subjects to learn concepts entirely from negative instances is also using the negative in an unaccustomed manner. A more appropriate use would consist possibly of a mixture of positive and negative information. Relevant to this notion is Huttenlocher's (1962b) result where a mixed sequence of positive and negative instances yielded better learning than both all negative and all positive instances. Other evidence contrary to Huttenlocher's (e.g., Hovland and Weiss, 1953; Whitman and Garner, 1963), however, should be cited here. Clearly further work is required to resolve this discrepancy. It would seem that contiguity effects may be confounded with the effects of positive and negative instances. Perhaps trial-by-trial response measures need to be recorded in addition to overall performance indices such as trials, errors, and time to solution.

Another reason for the difficulty of negative instances compared with positive instances was put forward by Hovland and Weiss (1953). They hypothesized that with positive instances the relevant attribute values are directly perceptible.

With negative instances, however, the required characteristics are not in direct view so that the subject must infer the proper combination from the fact that these characteristics never appear together. Hovland and Weiss proposed this explanation for conjunctive concept learning. An extension to nonconjunctive concepts has been suggested more recently by Haygood and Devine (1967). For these rules, these authors reported that increasing the proportion of instances embodying both relevant attribute values with the proportion of positive instances held constant improved performance. That is, the most efficiently processed positive instance is the one having both relevant attribute values (TT). As for conjunction, Haygood and Devine suggested that identification of the concept from TT instances involves little more than a perceptual task. To identify the concept from other instances, on the other hand, is essentially an inferential problem in which information must be transformed "in-the-head". Two components are involved here. Firstly, for instances other than TT instances, the concept must be derived using an indirect procedure which has been shown by Campbell (1965), for example, to be a more difficult method of problem-solving than a direct procedure. Secondly, the information must be transformed "in-the-head" for all instances except TT instances where information processing receives perceptual support, that is, non-TT instances impose a greater memory load than TT instances. More will be said later in this chapter on the importance of memory in concept learning. For the present we shall note Bruner, Goodnow, and Austin's (1956) observation that concept problems performed "in-the-head" are more difficult than problems in which an orderly array of instances is continually available to the subject.

Informative feedback

1. Informativeness of feedback

The role of the amount of information supplied to the subject following a category response has been investigated in several studies. One such study was conducted by Bourne and Pendleton (1958) using a four-category conjunctive task. Informativeness was varied in two ways: (1) the information specified in feedback, and (2) the probability that feedback would be omitted on some trials. The effect of amount of information specification was considered using two conditions. In one condition the subject was told on each trial whether his response was correct or incorrect; in the other condition the subject was shown, in addition, the correct response whenever he made an error. Performance was significantly better when more information was specified during feedback. The effect of probability of informative feedback was studied under conditions in which feedback was omitted on 0%, 10%, 20%, and 30% of trials. As expected, more trials to solution were required as the percentage of trials without feedback increased. Moreover, the increased number of trials to solution required was greater than the number of trials without feedback, suggesting that the no-feedback instances had an interfering effect on retention of information from feedback instances.

A recent study by Laughlin (1969a) has confirmed the finding that performance is improved by increasing the amount of information provided during feedback. In this experiment information specification for conditional concepts was manipulated using 11 feedback conditions. Four conditions specified one of the four possible cases of concept instances (if factor present, then factor present; if present, then absent; if absent, then present; if absent, then absent), leaving the other three cases ambiguous; 6 conditions specified two cases, leaving two cases ambiguous; and 1 condition specified each of the four cases. The selection procedure was employed so that feedback was supplied following the selection of each instance. Performance was found to be a direct function of the number of cases of concept instances specified.

Another experiment by Buss and Buss (1956) was concerned with a different aspect of the informativeness of feedback.

This study sought to compare the effect of giving the feedback "right" after correct responses and the feedback "wrong" after incorrect responses. Three conditions were used: (1) "right" for a correct response, nothing for an incorrect response; (2) nothing for a correct response, and "wrong" for an incorrect response; and (3) "right" for a correct response, and "wrong" for an incorrect response. The results indicated better learning on the "right" - "wrong" and nothing - "wrong" conditions than on the "right"- nothing condition. Similar results were reported by Buss, Braden, Orgel, and Buss (1956). These findings are rather unexpected since more information was supplied on the "right" - wrong condition than on either of the other conditions so that it might have been predicted that fastest learning should occur on this condition. The explanation proposed by Buss to fit this data was given in terms of reinforcement theory and may be divided into three assumptions: (1) the absence of feedback is a non-reinforcer; (2) "right" is a weak positive reinforcer; and (3) "wrong" is a relatively strong negative reinforcer. It follows that "right" - "wrong" and nothing - "wrong" are almost equivalent in reinforcement value and both provide greater reinforcement than "right" - nothing. Hence any program of feedback which uses "wrong" should produce more rapid learning than programs without "wrong". A variation on this explanation was suggested by Buchwald (1959) who assumed that, over a series of trials, a blank trial (one with no feedback) acquires reinforcement value opposite in sign to the overt feedback with which it is paired.

It has been pointed out by Bourne, Guy, and Wadsworth (1967) that in the Buss studies there were more ways of being wrong than of being right, so that there were fewer feedback trials on the "right" - nothing condition than the nothing - "wrong" and "right" - "wrong" conditions. Thus, another explanation of Buss's data may be given in terms of the total number of trials accompanied by feedback. No assumptions need be made concerning differential reinforcement values of "right" and "wrong". In an experiment in which the proportion of responses accompanied by confirming ("right") or infirming ("wrong") feedback was varied, Bourne et al. found the relative frequency of feedback to be able to account for most of the effect on performance.

2. Temporal factors

Several distinct temporal characteristics of a trial in a conceptual task may be identified (Bourne and Battig, 1966). These include the duration of stimulus presentation, the stimulus - response interval, the delay of informative feedback, the postfeedback interval, and the intertrial interval. Although experimental studies of simple learning in animal and human subjects have investigated all of these in some detail, only the last three mentioned have received attention in concept identification.

The delay of informative feedback is the time between a subject's response on any trial and the presentation of feedback. The effect of delay of reinforcement (which may be regarded as synonymous with informative feedback) has been studied in a variety of tasks other than concept identification. For example, a review of the effects of delay of reward in instrumental conditioning with animals (D'Amato, 1969) indicated that asymptotic performance is inversely related to the duration of reward delay, the longer the delay, the lower the asymptote. D'Amato also noted that, unlike reward magnitude and drive level, the performance differences to which different delays of reward give rise emerge rather slowly. In studies with human subjects, informative feedback or reinforcement has also been termed knowledge of results. It comes as something of a surprise to find that delay of knowledge of results in a motor learning task has no effect on the performance of human subjects, unless some intervening responses occur between the chosen response and feedback (Adams, 1969).

Bourne (1957) conducted a concept identification experiment with human subjects and obtained results which suggested that, as the delay of informative feedback increased, performance became worse. That is, it seems as if human subjects on conceptual problems behave more like animals on an instrumental learning task than humans on a motor learning task. However, in Bourne's study delay of informative feedback was confounded with another temporal variable, the postfeedback interval, that is, the time between informative feedback and the presentation of the next stimulus. The possibility that this latter variable may have had an effect was suggested by findings in motor learning that the critical temporal variable was the intertrial interval which is essentially the sum of the delay and postfeedback intervals. A subsequent experiment by Bourne and Bunderson (1963) confirmed this possibility. They discovered no effect on performance of delay of informative feedback, but showed that performance improved with increasing postfeedback interval. Thus, Bourne's earlier result was seen to be an artifact of failing to control the duration of the postfeedback interval. There would appear to be, therefore, some generality between the results for motor learning and concept learning. Longer postfeedback intervals have been used by Bourne, Guy, Dodd, and Justesen (1965) who observed increasing duration of the interval to give rise to facilitation of performance up to an optimum interval, beyond which performance became slightly worse. The optimum interval obtained was longer as the number of irrelevant stimulus dimensions was increased.

These effects seem to indicate that some kind of task-relevant activity takes place during the post feedback interval. Presumably, subjects may use the interval to rehearse, memorize, and process the information conveyed by the stimulus and its accompanying feedback. When long intervals are used, some forgetting of information may take place and so have a retarding effect on performance.

Availability of previous instances

In many concept learning tasks the stimuli to be categorized by the subject are displayed successively, with only one stimulus available for inspection on any trial. In other tasks the stimuli to be categorized are displayed simultaneously and remain available until the end of the problem. Hovland and Weiss (1953) found that when only negative instances were presented, more subjects identified the concept using simultaneous presentation than successive presentation. For positive instances only, the number of subjects attaining the concept was so high in both cases that type of presentation had no effect.

Simultaneous and successive methods of presentation were also compared by Cahill and Hovland (1960), again using only negative instances. In this experiment, a slightly different method of simultaneous presentation was employed in that stimuli were presented successively but, once exposed, remained in view for later inspection. Again, more correct responses were made under the simultaneous condition that the successive condition. In the successive condition, memory errors, consisting of hypotheses proposed which were incompatible with previously seen but no longer available instances, were found to increase progressively with the numbers of intervening instances.

Bourne, Goldstein, and Link (1964) observed that simultaneous and successive presentation methods as used by Cahill and Hovland are the end points of a continuum of stimulus availability. Availability may be defined as the number of previously exposed stimuli which a subject may inspect on any trial. In the successive condition no previously presented stimuli are available; in the simultaneous condition all previously presented stimuli are available.

Bourne, Goldstein, and Link conducted three experiments to test the effect of availability of previous stimuli. In the first experiment from 0 to 5 stimuli were exposed on any trial and performance was shown to improve as the number of available stimuli increased. This effect was more marked in problems with a larger number of relevant attributes. The outcome of the second experiment in which 0 to 10 stimuli were exposed was similar except that performance worsened slightly when more than 5 stimuli were exposed on each trial. The final experiment suggested that the worsening effect on performance of a large amount of available information may be due to the limited time (15 seconds) the subjects were allowed in which to make their responses. When this limitation was removed performance continued to improve when up to 8 previous stimuli were available for inspection.

A series of studies by Pishkin and Wolfgang (1965), Pishkin, Wolfgang, and Rasmussen (1967) and Pishkin (1967) examined the effect on performance of not only the number but also the type of instances available. Only the availability of correctly sorted instances in a four-choice concept learning task facilitated performance; availability of incorrectly sorted instances did not affect performance. When corrective feedback was provided on incorrectly sorted instances, the level of performance was markedly improved, but again no facilitation was found for increasing numbers of this type of instance available. These findings may be interpreted as indicating that both the stimulus and its correct category must be given in order for the availability of past instance to facilitate concept learning.

The experiments discussed thus far have been conducted only with conjunctive problems. The effects of availability of positive and negative instances for nonconjunctive rules were considered by Bourne, Ekstrand, and Montgomery (1969). This experiment differed from earlier experiments also in that a selection procedure was used instead of the usual reception paradigm. The results confirmed the importance of the availability of previous stimuli and their respective categories as a variable affecting concept learning. The relative effectiveness of retained instances differed, however, with the rule. For example, positive instances were more informative for conjunctive and negative instances for conditional concepts. On disjunctive and biconditional problems, subjects appeared to benefit equally from the availability of either instance type. However, for the biconditional, availability of the most frequently chosen instance type facilitated performance.

In the above studies the retention of instances has been under the control of the experimenter. Two studies (Trabasso and Bower, 1964b; Bourne and O'Banion, 1969) have attempted to determine the number of instances subjects are able to recall themselves during concept identification. Trabasso and Bower tested subjects for their recall of six stimulus cards and their appropriate categories. Overall the probability of correct recall of an instance was only slightly above chance, even though subjects were instructed beforehand that their recall of instances presented would be tested as well as their performance at concept identification. A serial position effect of the presentation order of instances on recall was apparent from the data. In particular, the results showed clear evidence of a primacy effect and also a smaller recency effect in recall.

Bourne and O'Banion (1969) tested recall of instances in concept identification without any prior warning before the task. Six instances were again presented successively to the subject who then stated an hypothesis about the solution. Six problems were presented and after some problems the subject was asked to freely recall each instance and its category. The later in the problem series recall was required, the poorer recall was, suggesting that subjects learn that retention of the stimulus pattern and its category is unnecessary. Like Trabasso and Bower, Bourne and O'Banion noted that only a minimal amount of information was retained. This observation was in contrast to large percentage of problems solved, 76%. Also, like Trabasso and Bower, Bourne and O'Banion obtained evidence for both primacy and recency effects in recall as a function of presentation order.

It would seem from this evidence that subjects tend not to retain information about the concept in the form of the stimuli encountered and their associated categories. Rather, information that is retained would appear to be derived or abstracted from each instance presented. It may be in the form of a working hypothesis of the concept or it may represent a summary of the supposed relevance of particular attributes. Investigations of the strategies utilized in concept learning have indicated the existence of both of these forms (Bruner, Goodnow, and Austin, 1956). Retention of information in one or other of these forms has the advantage of reducing the load on memory. Associated with such an advantage, however, is the disadvantage that some information must be rejected. When information relevant to the concept is rejected, performance is impaired.

When previous instances are available, no information need be rejected. Instead the task becomes one of selecting the appropriate or relevant attributes from the instances displayed. Obviously, if the categories to which instances are assigned are not displayed along with the instance, this information must be stored by the subject. Since such information is highly confusible, memory errors in performance may be expected. If insufficient time is allowed to inspect the available stimuli, again some relevant information will not be selected so that performance will be impaired. Furthermore if the arrangement of available stimuli is such that relevant information becomes difficult to select, performance may also show a decrement. This possibility has been demonstrated by Bruner, Goodnow, and Austin (1956) using ordered and random arrays of stimuli, and may perhaps account for Laughlin's (1968b) failure to find an improvement in performance when subjects were supplied with pencil and paper as a memory aid.

General comments

This review has attempted to survey research carried out on the major task variables in concept learning. Clearly further work is necessary if the role of these variables is to be fully understood. Some suggestions for future experiments have been made during the course of the review. One gap in present knowledge which perhaps needs more emphasizing is the study of interaction effects between variables. In many cases in the preceding discussion, the effect of two variables together was different from the sum of the effects of the two variables examined separately.

What conclusions have been made in this review should be taken into account when considering the experiments to follow in subsequent chapters. Unless information is provided to the contrary, these experiments employ in general a standard procedure. Thus, the stimulus material consists of geometrical designs varying along four dimensions, each dimension having three values. The salience of each of these dimensions no doubt varies, so effects of particular dimensions are minimized through random selection of the relevant attribute values for each subject and problem. Two relevant attribute values are chosen in each case. The selection paradigm is employed so the stimulus sequence presented is under the control of the subject, except for the initial instance which is determined by the experimenter. In most experiments the initial instance is positive. The effect of presenting a negative instance at the start of a problem is examined in Experiment II. The rate at which instances are selected is paced by the subject and feedback is administered immediately following each instance selected. In all experiments the stimulus population is displayed on a table in front of the subject. In some experiments stimuli are arranged in an orderly, systematic array, in others a random array is used. No memory aids are made available to the subject which he may use to assist himself in retaining either previous instances selected and their categories or information derived from these past selections. Exceptions to this generalization are Experiments IV and V in which the use and effectiveness of various artificial memories form the subject of inquiry.

This review should not be regarded as being exhaustive of all variables present in concept learning. Other variables not considered here should also be mentioned, e.g., age, intelligence, motivation, personality, social variables, etc.

For the most part, these are variables connected with the subjects in the experiment. Their omission is not intended to signify that they are unimportant in concept identification. Rather it is an indication generally that little is understood of the role of these variables, mainly because of methodological difficulties associated with their use in controlled experimental studies. No detailed review of such variables is presently available to the knowledge of the writer. Without doubt, this indicates serious neglect by workers interested in concept learning. Nevertheless, a brief introduction to the effects of some of these variables may be found in Bourne (1966) and Jensen (1966). In the experiments to be reported in this dissertation, the subjects are randomly selected from the subject pool of first year students doing psychology at the University of Adelaide. Accordingly their average age lies between 18 and 19 years and they are of above average intelligence. Each subject is run individually and the only other person is the room at the time of the experiment is the experimenter, who gives the initial instructions, provides feedback on each instance selected, and decides when each problem is completed. The motivation of the subject is assumed to be controlled to a large extent by the instructions. It became evident over the course of these experiments that many subjects regarded the task as a test of their intelligence, despite being told otherwise. Generally therefore it may be assumed that subjects are highly motivated to perform well. The experiments are conducted in a room whose surroundings, while pleasant, are not likely to distract the subject from his task.

One other variable has not been considered in this review, namely, the logical relationship between relevant attribute values or conceptual rule. Most of the studies so far discussed have used either affirmational or conjunctive concepts. Occasionally experiments incorporating other conceptual rules have been conducted and in most cases these have been mentioned along with the other studies. The aim of the succeeding chapters is to investigate the effect of this variable on conceptual performance. In conceptual problems three types of instructions to subjects may be used (Haygood and Bourne, 1965): firstly, subjects may be required to learn both the relevant attribute values and the rule relating them in the concept - complete learning; secondly, subjects may be instructed concerning the particular rule involved in the concept and set the task of identifying the relevant attribute values - attribute identification; finally, subjects may be given the relevant attribute values of the concept and required to learn the rule which combines them - rule learning. Virtually all of the studies considered in this chapter have utilized one of the first two instructional conditions. Relatively few studies have investigated rule-learning. In this dissertation experiments on the attribute identification of concepts are described in Chapters III, IV and V; rule-learning is examined in Chapter VI.

CHAPTER 3

ATTRIBUTE IDENTIFICATION OF CONCEPTS EMBODYING DIFFERENT LOGICAL RULES

In Chapter 1 ten different nominal conceptual rules were distinguished for relating two relevant attribute values. In Chapter 2 it was suggested that these rules represented a variable whose effect on performance required experimental investigation in the same way as other task variables. In this chapter this effect will be examined with reference to attribute-identification tasks.

Effect of rules on attribute-identification difficulty

Early studies of conceptual rules were concerned for the most part with conjunction and inclusive disjunction. Probably the first workers to consider in any detail disjunctive concept identification were Bruner, Goodnow, and Austin (1956), who, although they made no direct comparison between conjunctive and disjunctive rules, suggested that disjunctive concepts were more difficult to attain. Later and more systematic investigations have confirmed this suggestion.

Hunt and Hovland (1960) studied the order in which three types of concepts were proposed as a solution when all were consistent with a particular grouping of stimuli. A training series of positive and negative instances was presented to each subject. The instances in this series were chosen in such a way that the positive category could be described logically either by a conjunctive, a disjunctive, or a relational rule. The relational concept need not concern us here since this is formed by an ordinal rule. The training series was followed by a test series of instances in which the subject was required to select those instances which were positive according to the concept he had formed during training. Hunt and Hovland found conjunctive concepts to have been formed with greater frequency than disjunctive concepts. If frequency of formation may be interpreted as an indicator of the relative difficulty of each rule, then it can be concluded that conjunctive concepts are easier than disjunctive concepts. It is apparent, however, that the task used by Hunt and Hovland was not an attribute-identification task, but more like a complete - learning task. A reasonable question to ask concerning these results, therefore, is whether the differences obtained were due to rule - learning or attribute - identification or both.

An answer to the question whether such differences occur on attribute - identification is found in an experiment by Conant and Trabasso (1964). In this study subjects were required to discover the solution to conjunctive and inclusive disjunctive problems using a selection procedure. Problems of both types were presented to all subjects according to a counterbalanced design such that half the subjects took conjunction first, the other half disjunction. The stimulus population contained four dimensions, each with two levels. The results showed that disjunctive concepts necessitated more trials to solution than conjunctive. Similar results have been reported for fifth-grade children (DiVesta and Walls, 1969).

Recent studies have extended the number of rules considered. Haygood and Bourne (1965) compared performance on four different rules - conjunction, inclusive disjunction, conjunctive absence, and conditional. The stimuli were geometrical designs varying on four dimensions each with three levels. These were presented to the subject according to the reception paradigm in a randomized sequence consisting of equal numbers of positive and negative instances. All subjects were given five successive problems involving the same rule on each problem. On the first problem rules were found to differ markedly in difficulty with conjunction having the least errors and trials to solution and conditional the most; no difference was found between inclusive disjunction and its complement, conjunctive absence. The differences between rules diminished with successive problems, such that after five problems there appeared to be no difference between conjunction, inclusive disjunction, and conjunctive absence. More trials to solution were still required for the conditional, however.

Bourne (1967) examined the relative difficulty of the conjunctive, inclusive disjunctive, conditional, and biconditional rules, using a similar procedure to that employed by Haygood and Bourne with two changes. Firstly, stimuli were presented in such a way that, within each set of 40 trials, the numbers of TT, TF, FT, and FF instances were equated. Secondly, subjects were provided with sample patterns illustrating the proper response assignment of each of the four stimulus classes. This latter modification meant that, if the subjects understood the rule, described and demonstrated in the preliminary instructions, and if he had acquired the notion of stimulus classes, and if the four sample patterns were properly chosen, then sufficient information was available at the beginning of a problem to identify the two relevant attribute values of the concept. The results showed that, on the first problem, errors were made on all rules. For this problem rules differed in difficulty with the order from easiest to most difficult being conjunction, inclusive disjunction, conditional, and biconditional. The effect of the two procedural modifications was to reduce the difficulty of attribute - identification, but not sufficiently to overcome rule differences. The relative contributions of each modification to this reduction cannot be ascertained. On later problems, however, the effect of making sample patterns available could be clearly seen by the fact that no errors were made prior to concept identification. This was true for all rules, although the ordinal numbers of the first problem on which no errors occurred for the various rules showed a similar ranked order to the ranking of rules according to their difficulty on the first problem. Thus, errorless performance was reached for the easiest rule, conjunction, sooner (problem two) than the most difficult rule, biconditional (problem six).

Laughlin and Jordan (1967) compared the difficulty of attribute - identification of conjunctive, inclusive disjunctive, and biconditional concepts. The stimulus display employed was a systematically arranged array of cards having six different colour attributes, each having two values - plus or minus. In contrast to the experiments of Haygood and Bourne, Laughlin and Jordan utilized a selection procedure in which subjects were free to choose their own instances following an initial positive instance. Their results showed some departure from the Haygood and Bourne findings, in particular the relative difficulty of the biconditional. They found biconditional concepts to be significantly less difficult than disjunctive concepts for both number of card choices and time to solution; no significant difference was obtained between conjunction and biconditional.

Laughlin (1968b) examined the relative difficulty of eight conceptual rules. Again he used a stimulus universe with six two-valued attributes. The selection paradigm was employed so that the procedure on each trial after the initial positive instance was as follows: (1) the subject selected a card, (2) the experimenter indicated whether the card was positive or negative, (3) the subject made an hypothesis of the concept, (4) this hypothesis was confirmed or disconfirmed by the experimenter. The results reported by Laughlin revealed a significant effect of rules with an order of increasing difficulty of conjunction, conjunctive absence, exclusion, exclusive disjunction, biconditional, disjunctive absence, inclusive disjunction, and conditional.

An identical order of rules was obtained by Giambra (1969) for the number of trials to attribute-identification. In this experiment the selection procedure was again used with the modification that half the subjects were presented with a positive instance as the initial card, and the other half with a negative instance. Two different stimulus populations were utilized in random arrays. Half the subjects were presented with a display having three two-valued attributes, while the other half had a display with four two-valued attributes.

It seems clear from this review that there are reliable discrepancies in findings on the relative difficulty of different conceptual rules. These discrepancies may be divided into two parts. Firstly, Laughlin and Giambra differ from Haygood and Bourne in the suggested order of difficulty of the so-called "basic" rules. Laughlin and Giambra posit this order to be conjunction, biconditional, inclusive disjunction, and conditional. Haygood and Bourne, on the other hand, suggest the order to be conjunction, inclusive disjunction, conditional, and biconditional. Secondly, Laughlin and Giambra's results indicate that the logically complementary rule is not of the same degree of difficulty as the basic rule it complements. Haygood and Bourne, however, found inclusive disjunction and its complement, conjunctive absence, to be similarly difficult. In the discussion which follows we shall concentrate mainly on the first of these discrepancies, that is, the order of difficulty of the four "basic" rules. It is quite likely, nevertheless, that much of what is said is also relevant to the second major discrepancy listed above.

Variables interacting with rules

It is apparent that the effect of rules as a variable in concept identification is interactive with the effects of other task conditions which vary between the two sets of experiments. Several methodological differences can be found. Of these, three can be identified which may possibly account for the discrepant findings.

Firstly, the studies by Haygood and Bourne and Laughlin and Giambra differ in the amount of stimulus variability involved. Haygood and Bourne used a four attribute, three valued (4 - 3) stimulus population. Laughlin's stimulus population had six attributes each with two values (6 - 2). Giambra used two levels of interdimensional variability; his stimulus displays contained either three or four binary attributes (3 - 2 or 4 - 2). The important difference between Haygood and Bourne, on the one hand, and Laughlin and Giambra, on the other, is the level of intradimensional variability employed. Arguing from considerations of the strategies subjects use in solving concepts, Laughlin (1968b) suggested that his display facilitated the solution of biconditional concepts compared with that used by Haygood and Bourne since "biconditional concepts are probably relatively more difficult than conjunctive or disjunctive as the number of values per attribute increases." Unfortunately, no empirical evidence on the effect of varying the level of intradimensional variability is available for the various rules. In the study by Haygood, Harbert, and Omlor (1970) in which performance was found to improve in a positively decelerating fashion with increasing number of values per attribute, only unidimensional or affirmational concepts were used. A recent study by Laughlin (1969b) has sought to compare the effect of a 4 - 3 stimulus universe as used by Haygood and Bourne with a 6 - 2 stimulus universe which he himself used previously. His results showed a significant Rule X Universe interaction. This interaction effect was not seen in the order of "basic" rules, however. For both stimulus universes, the order from easiest to most difficult was conjunction and biconditional, inclusive disjunction, and conditional. The interaction effect was due rather to the relative difficulty of conjunctive absence with the two universes. Unlike other rules, conjunctive absence concepts were not identified in fewer card choices with the 4 - 3 universe than the 6 - 2 universe. Evidently, the amount of stimulus variability is not the factor whose interaction with conceptual rule may be used to account for the discrepant orders of "basic" rule difficulty.

A second procedural difference is the fact that Laughlin and Giambra both employed the selection procedure, while Haygood and Bourne adopted the reception procedure. In the previous chapter evidence reviewed comparing the two procedures yoked so that each reception subject received the same instances in the same order as a selection subject was inconclusive. One reason was a report by Schwartz (1966) of a significant interaction between the type of procedure employed and conceptual rule. Laughlin (1969b), however, was unable to find any such interaction effect, and this result together with several other objections to Schwartz's experiment cited in the previous chapter were used to raise doubts about his conclusions. A recent study by Giambra (1970) appears to confirm Laughlin's finding of no interaction between procedure and rule. In this study the conditional rule was found to be less difficult than the biconditional on both the selection and yoked reception procedures. An interesting point to note here is that the order of rules reported by Giambra (1970) is the same as that reported by Haygood and Bourne, even for the selection procedure. The type of procedure employed, therefore, does not appear to be able to account for the discrepant orders, firstly, because of a lack of conclusive evidence of an interaction between rule and mode of presentation, and secondly, because the discrepant orders occur even using the same mode of presentation. Clearly, some other variable is operating.

A closely related variable methodologically to mode of presentation is the sequence of instances encountered by the subject. In Laughlin's (1968b) and Giambra's (1969) experiments the sequence of instances was selected by the subject himself after the initial instance. In Haygood and Bourne's experiments the sequence of instances was programmed by the experimenter according to predetermined conditions. These conditions were firstly, that the sequence of instances was randomized, and secondly, that the sequence should contain equal numbers of positive and negative instances (Haygood and Bourne, 1965) or equal numbers of TT, TF, FT, and FF instances (Bourne, 1967). These conditions affect instance contiguity in two ways. Firstly, randomizing the instances presented effectively minimizes the amount of instance contiguity, given constraints in the proportion of positive and negative instances. Evidence on human randomness (Baddeley, 1966) indicates that subjects on the selection paradigm are unlikely to select instances completely at random, even if they are instructed to do so. It seems likely, therefore, that the sequences subjects generate themselves will display greater contiguity than a random series. It is doubtful, however, all other things being equal, that the contiguity of sequences generated for some rules should be greater than or less than for other rules. It seems unlikely, therefore, that this difference between Laughlin and Giambra, and Haygood and Bourne can account for their different findings on rule difficulty. Secondly and more importantly, the amount of instance contiguity is a function of the proportion of positive instances in the sequence. The higher is the proportion of positive instances, the greater is the amount of instance contiguity. The percentage of positive instances in the sequences presented in the experiments by Haygood and Bourne is easily determined for each rule from the conditions on which their sequences were based. Thus, in Haygood and Bourne's (1965) study the sequence contained 50% positive instances for all rules. In the experiment reported by Bourne (1967), 25% of the instances for conjunction were positive, 50% for biconditional, and 75% for inclusive disjunction and conditional. Unfortunately, neither Laughlin nor Giambra report the proportion of positive instances selected on any rule. This information is essential for meaningful comparisons to be made between the sequences generated in their experiment and the sequences presented by Haygood and Bourne. If, for example, a greater percentage of positive instances were found to be selected for the biconditional than presented by Haygood and Bourne, then this would mean that instance contiguity on this rule was greater in Laughlin's and Giambra's experiments than in Haygood and Bourne's, which in turn would imply that the biconditional was relatively easier in the former set of experiments than the latter.

Some indication may be gained of the percentage of positive instances selected from the percentage of the stimulus population which were positive instances. This, of course, assumes that subjects choose instances randomly, an assumption which has already been questioned. If this assumption was true, it would mean that the proportion of positive instances selected in Laughlin and Giambra was 25% for conjunction, 50% for biconditional, and 75% for inclusive disjunction and conditional. Comparing this with Haygood and Bourne (1965) suggests that instance contiguity may have been greater for inclusive disjunction and conditional, and less for conjunction in Laughlin's and Giambra's studies. From this it could be concluded that inclusive disjunction and conditional would be relatively easier and conjunction relatively more difficult in Laughlin's and Giambra's experiments than in Haygood and Bourne's. This conclusion, however, is completely contrary to what has been found for inclusive disjunction and conditional.

Perhaps this conflict between prediction and result may be resolved by considering two further observations about positive and negative instances. These are firstly, that positive and negative instances may transmit different amounts of information relevant to the concept. The effect of the nonequivalence of information of positive and negative instances has been illustrated by Bourne and Guy (1968b) who showed that conditional concepts are more easily identified from more informative negative instances than from less informative positive instances. When, however, the informativeness of positive and negative instances is approximately equivalent, as for inclusive disjunction with a 4 - 3 stimulus universe, concept identification is better with positive instances than negative instances. This is the second aspect, namely, that other things being equal, subjects seem to prefer to process positive information than negative information.

These observations suggest the following predictions for rule difficulty in the two sets of experiments. We shall again assume that subjects in Laughlin's and Giambra's experiments selected instances in the same proportion as they occurred in the stimulus population. To begin, the experiment by Haygood and Bourne (1965) will be related to Laughlin and Giambra. For conjunction, positive instances are more informative than negative instances. Thus, the lower proportion of positive instances in Laughlin and Giambra compared with Haygood and Bourne suggests that conjunctive concepts should be relatively more difficult in the former experiments. For inclusive disjunction, the situation is complicated by the fact that, for the stimulus universe employed by Laughlin and Giambra, negatives are more informative than positives, whereas for Haygood and Bourne's universe, positives and negatives convey approximately the same amount of information about the concept. Perhaps a tentative solution can be reached by considering the data provided by Bourne and Guy (1968b). This data may possibly be interpreted to suggest that changes in or disjunctive groups, but many in the conditional and biconditional groups requested further clarification of their conceptual type. When the subject indicated he was ready, the task was begun. The typed instructions sheet was available throughout the task.

2. Results

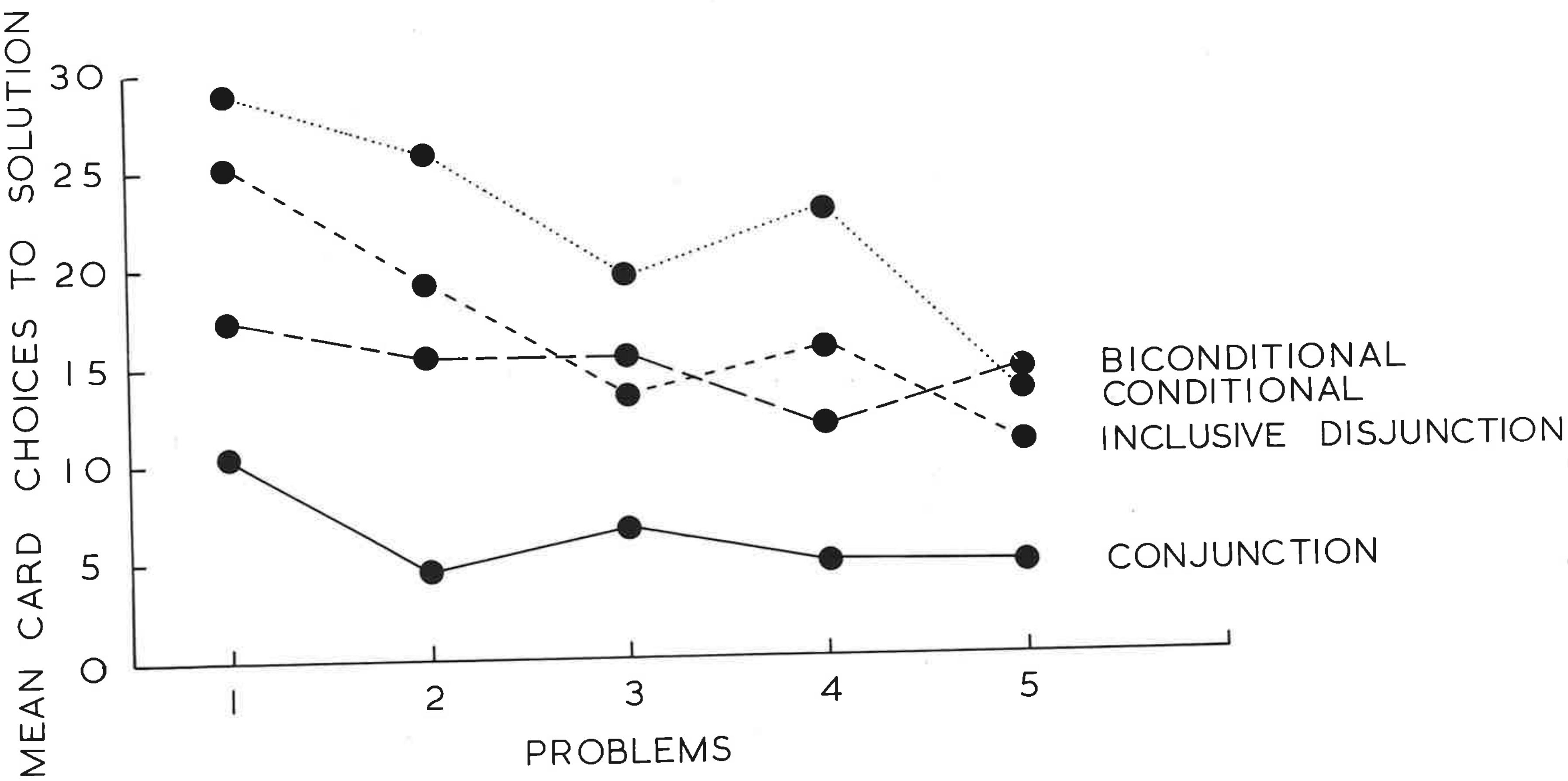
The mean and standard deviation of the number of card choices to solution for each conceptual rule and problem are given in Table 1. An analysis of variance was performed on this data transformed by a square root transformation in order to make the variances more homogeneous. The effect of conceptual rule was significant, $\underline{F}$ (3, 48) = 11.07, $\underline{p} < .001$, and a significant difference was also found between problems, $\underline{F}$ (4,192) = 7.18, $\underline{p} < .001$. No significant interaction between conceptual rules and problems was obtained.

Conceptual Rule		Problem				
		1	2	3	4	5
Conjunction	Mean	10.4	4.5	6.7	4.8	4.7
	S.D.	9.0	2.8	5.5	3.4	2.2
Inclusive Disjunction	Mean	25.2	19.2	13.5	15.8	10.8
	S.D.	18.6	14.4	11.0	11.5	4.4
Conditional	Mean	29.0	25.8	19.6	22.8	13.4
	S.D.	23.6	13.2	14.1	16.0	6.0
Biconditional	Mean	17.3	15.5	15.5	11.8	14.6
	S.D.	10.5	10.2	15.8	16.1	15.1

TABLE 1. Number of card choices to solution.

A summary of this analysis may be seen in Table 2. A post hoc analysis using the Newman-Keuls method for multiple-range comparisons (Winer, 1962) revealed: for rules, significant differences between conjunction and other rules ($p < .01$), and between the biconditional and the conditional ($p < .05$); for problems, more trials were required on Problem 1 than on other problems ($p < .01$), and on Problem 2 than on Problem 5 ($p < .05$). A factorial analysis of trend (Winer, 1962) indicated that the order of rules - conjunction, biconditional, inclusive disjunction, and conditional - showed a highly significant linear trend such that 93.3% of the variation in difficulty for the conceptual rules could be predicted by a linear regression equation. Problems also displayed a significant linear trend for decreasing difficulty with increasing number of problems such that 93.5% of the variation between problems could be predicted from a linear regression equation. A graphical representation of these findings can be found in Figure 1.

FIGURE 1. Mean number of card choices to solution for each rule over the 5 problems



MEAN CARD CHOICES TO SOLUTION
30
25
20
15
10
5
0
1
2
3
4
5
PROBLEMS
BICONDITIONAL
CONDITIONAL
INCLUSIVE DISJUNCTION
CONJUNCTION

Source of variation	S.S.	D.F.	M.S.	F	P
Between subjects	380.45	51			
Rules	155.56	3	51.85	11.07	<.001
Subjects within groups	224.89	48	4.69		
Within subjects	290.28	208			
Problems	35.89	4	8.97	7.18	<.001
Rules X Problems	14.38	12	1.20	<1	N.S.
Problems X Subjects within groups	240.01	192	1.25		

TABLE 2. Summary of analysis of variance on the number of trials to solution transformed using a square-root transformation.

The type of instances chosen were measured in terms of the percentage of negative instances selected on each problem for each conceptual rule. A summary of this analysis is given in Table 3. Of course, the percentage of positive instances is found from the percentage of instances selected which were not negative. The percentages given in Table 3 may be compared with the percentage of negative instances occurring in the stimulus population for each type of concept - conjunction 88.9%, inclusive disjunction 44.4%, conditional 22.2%, and biconditional 44.4%. A sign test indicated that a lower percentage of negative instances were selected for conjunction on all problems ($\underline{p}< .001$), for inclusive disjunction on all problems ($\underline{p}< .001$), and for biconditional on Problems 2 ($p < .05$), 3 ($p < .01$), and 5 ($p < .05$). The same test indicated a higher percentage of negative instances were selected for the conditional on Problems 2, 3 (both $p < .01$), and 4 and 5 (both $p < .001$).

Conceptual Rule	Problem				
	1	2	3	4	5
Conjunction	39.2	27.5	36.4	28.6	40.6
Inclusive disjunction	26.7	23.8	25.8	27.1	26.6
Conditional	25.7	36.2	34.3	42.1	44.0
Biconditional	42.1	40.2	31.6	44.3	34.4

TABLE 3. Mean percentage of negative instances selected.

Spearman rank correlations were computed for each type of concept between the percentage of negative instances chosen and the number of card choices to solution. The correlations obtained were, for conjunction, rho (63) = 0.57, $p < .001$, for inclusive disjunction, rho (63) = 0.29, $p < .05$, for conditional, rho (63) = -0.38, $p < .01$, and for biconditional, rho (63) = 0.20, $p < .20$.

Instances were also analysed into stimulus classes - TT, TF, FT, FF. The mean percentage of each of these four

Conceptual Rule	Stimulus Class			
	TT	TF	FT	FF
Conjunction	65.5	15.9	15.5	3.0
Inclusive disjunction	23.0	21.8	29.2	26.0
Conditional	25.2	36.5	16.5	21.9
Biconditional	38.9	14.8	23.7	22.6

TABLE 4. Mean percentage of four classes of instances selected taken over all problems.

classes of instances selected for each rule calculated over all problems is given in Table 4. These percentages can be compared with the percentage of each type of instance as it occurred in the stimulus population - TT 11.1%, TF 22.2%, FT 22.2%, and FF 44.4%. Sign tests showed that a higher percentage of TT instances ($\underline{p} < .001$) and a lower percentage of FF instances ($\underline{p} < .001$) were selected than were present in the stimulus population for all rules. With TF instances, a higher percentage was selected for conditional ($p < .001$) and lower percentages for conjunction ($p < .006$) and biconditional ($p < .001$). With FT instances, a higher percentage was selected for inclusive disjunction ($p < .002$) and lower percentages for conjunction ($p < .002$) and conditional ($p < .001$).

Since the number of trials to solution was not found to be significantly related to the proportion of positive instances selected on the biconditional, an estimate of the correlation between this performance measure and the percentage of TT and of FF instances was obtained using Spearman's rank correlation technique. It will be realised that TT and FF instances are the two types of positive instances for the biconditional. Significant correlations were found with both types of positive instance, but the signs of the correlations were opposite. For TT instances, rho (63) = - 0.51, $p < .001$; for FF instances, rho (63) = 0.33, $p < .01$. A similar analysis was performed for inclusive disjunction. For this rule, TT instances yielded rho (63) = - 0.46, $p < .001$; for instances with only one relevant attribute value (i.e. TF + FT), rho (63) = 0.17, $p < .20$.

3. Discussion

The results indicate that the ascending order of difficulty of conceptual rules in this experiment is conjunction, biconditional, inclusive disjunction, and conditional. This is the same order as was obtained by Laughlin (1968b) and Giambra (1969), but different from the order reported by Haygood and Bourne (1965) and Bourne (1967). Evidently, amount of stimulus variability is not the factor producing the two different orders, since the stimulus universe employed in the present experiment was the same as that utilized by Haygood and Bourne. Evidence presented earlier in this chapter suggests that the type of procedure, reception versus selection, does not interact with conceptual rule so as to explain the two orders, either.

In the introduction to this experiment an argument was developed proposing that the discrepant orders of rules were due to differences in the sequences of instances encountered by subjects in Haygood and Bourne's experiment and in the experiments by Laughlin and Giambra. Since no data was available from Laughlin or Giambra on the type of instances selected for each rule, it was assumed then that subjects selected positive and negative instances in the proportion in which they occurred in the stimulus population. The present results show that on almost every problem and for all rules, this assumption is false. Rather, they indicate that subjects have a tendency to select more positive instances for conjunction, inclusive disjunction, and biconditional, and more negative instances for conditional. Compared with the 50% positive and 50% negative instances presented for each rule by Haygood and Bourne, this means that subjects selected a higher proportion of positive instances and a correspondingly lower proportion of negative instances for each rule in the present experiment.

A second assumption in the argument set out in the introduction was that performance was improved as the proportion of positive instances was increased for conjunction, inclusive disjunction, and biconditional, but for the conditional better performance was produced by increasing the proportion of negative instances. Support for this assumption comes from the correlations between the percentage of negative instances selected and the number of trials to solution: positive correlations were obtained for conjunction, inclusive disjunction, and biconditional, while the conditional yielded a negative correlation.

From the differences noted between the percentage of positive and negative instances presented by Haygood and Bourne and selected by subjects in the present experiment, it can be deduced, therefore, that performance in the present experiment should be facilitated on conjunction, inclusive disjunction, and biconditional, but retarded on conditional, relative to performance in Haygood and Bourne's experiment. This accounts for the place of conjunctive, inclusive disjunctive, and conditional rules in the orders of difficulty obtained in the two experiments, but some further explanation is necessary to show why the biconditional should be assumed to be the most difficult rule in the Haygood and Bourne studies, when it has been found to be relatively easy in this experiment and in the experiments of Laughlin and Giambra. This is made more obvious when it is recalled that the correlation between the percentage of negative instances and the number of trials to solution for the biconditional was low and non-significant. Thus, the difficulty of the rule is less likely to be affected by a change in the proportion of negative instances encountered.

The evidence for the extreme relative difficulty of the biconditional comes from Bourne (1967). Applying the same argument to his experiment as was used above on Haygood and Bourne (1965), however, still does not fully account for the different orders obtained. In Bourne's experiment 25% positive instances were presented over 40 trials for conjunction, 75% for inclusive disjunction and conditional, and 50% for the biconditional. Comparing these percentages with the percentages chosen by subjects in the present experiment suggests that conjunction, conditional, and biconditional should show some facilitation in the present experiment, with no change for inclusive disjunction. Once again this prediction seems to fit the data for conjunction, inclusive disjunction, and conditional, but does not account for the relative ease of the biconditional.

Possibly the discrepancy is due to the other modification to the procedure introduced by Bourne (1967). This was to provide the subject with sample patterns showing the response assignment of each of the four stimulus classes - TT, TF, FT, and FF. Perhaps considerable difficulty is experienced with the biconditional rule compared with other rules in identifying the relevant attribute values from these sample patterns. Unfortunately, Bourne did not run a control group of subjects who were not presented with sample patterns.

Looking at the proportion of TT, TF, FT, and FF instances selected in the present experiment, it is apparent that instances from these four classes were not selected in the same proportions as occurred in the stimulus population for any of the rules. In particular, subjects displayed a clear tendency to choose more TT instances and less FF instances. The former observation is of interest in view of Haygood and Devine's (1967) demonstration that among the positive instances of inclusive disjunctive and biconditional concepts, those exhibiting both relevant attribute values (i.e., TT instances) produced better performance. Correlations between the proportions of each type of positive instance selected and performance for inclusive disjunction and biconditional in the present experiment confirm Haygood and Devine's finding. Indeed for the biconditional, performance was found to be a direct function of the percentage of TT instances but inversely related to the percentage of FF instances. This seems to imply that the biconditional in the present experiment may have been relatively less difficult than in Bourne's experiment, since in the present experiment the number of TT instances was almost twice the number of FF instances whereas in Bourne's experiment the frequency of positive instances taken from both classes was the same. No effect, however, is expected for inclusive disjunction since the percentages of TT, TF, and FT instances presented in Bourne's experiment and selected in the present experiment are of the same order.

The general finding that subjects did not select instances in the proportion in which they occurred in the stimulus population fails to replicate a recent result by Bourne, Ekstrand, and Montgomery (1969). These workers reported that, in general, subjects' selections from the four truth table classes reflected closely their 1: 2: 2: 4 distribution in the stimulus population. Some deviations from this trend, nevertheless, were noted: for example, more TT instances were selected on conjunction, and more TF instances were selected on conditional. It is also noteworthy that Bourne et al. (1969) found the biconditional to be relatively more difficult than in the present experiment; their results showed the biconditional to be of the same order of difficulty as the conditional, with the disjunctive less difficult and the conjunctive least difficult. (These results were for the no feedback retained condition). When subjects were divided into better performers and poorer performers, it was found that the better subjects selected TT instances with relatively high frequency in biconditional problems compared with poorer subjects, confirming the implication made in the previous paragraph on the relationship between the relative frequency of TT instances and biconditional performance.

Two procedural differences between Bourne et al's experiment and the present experiment may account for the discrepant findings on the distribution of selections from the four stimulus classes. Firstly, Bourne et al. used a stimulus display in which instances were randomly arranged, whereas the instances in the present experiment were systematically arranged. In a comparison between the effects of these two displays on conjunctive concept identification, Bruner, Goodnow, and Austin (1956) observed that the choices made by subjects using an ordered array of instances were more "similar" to the initial positive instance than to any other preceding instance. This observation was less prevalent for subjects with a random array of instances. (Similarity here was measured by which preceding card had the most attribute values in common with the present card. Where two preceding cards had the same number of attribute values in common with the present card, the one encountered earlier was counted as being the more similar). Since the initial positive instance was a TT instance, this means that subsequent instances tended to be chosen mainly from the TT, TF, and FT stimulus classes with an ordered array, with more FF instances selected with a random array. In the present experiment the initial instance was a TT instance for each rule. Extrapolating from Bruner et al's findings with conjunction to other rules suggests that for the same initial instance an ordered array of instances tends to produce more TT and less FF instances than a random array. Further work is necessary to test this hypothesis for nonconjunctive rules.

Secondly, Bourne et al. did not provide subjects with a sample positive instance at the beginning of a problem. In a discussion of disjunctive concept attainment, Bruner et al. (1956) have referred to the effect of the type of illustrative instance presented prior to a problem. They suggested that subjects given an initial negative instance display a greater tendency to select further negative instances than subjects given a positive instance. That is to say, the nature of the initial instance tends to produce a bias in subsequent selections such that more instances of the same type as the initial instance are selected. Whether this is true for all rules is an empirical question whose answer is sought in the next experiment. Thus, in the present experiment the presentation of a positive instance at the beginning of the task may be predicted to have given rise to an increased proportion of positive instances selected if compared with an experiment in which a negative instance was initially given. When no initial instance is presented, as in Bourne et al., two outcomes are possible. Either the type of instances chosen will be a function of the type of instance first encountered, so that the proportion of each type of instance selected will be biased in the direction of whichever is the more frequent in the stimulus population, or, there will be no bias in selections so that instances will be selected in the same proportions as they occur in the population. The finding of an interaction between rule and truth table class by Bourne et al. indicated that the most informative, that is, the least frequent instances of any rule tended to be selected most often. This is clearly contradictory to the first possible outcome above, suggesting that the latter is nearer to the truth. In this case, the proportions of each class of instance selected when no initial instance is given should lie in between the proportions chosen following an initial positive instance and an initial negative instance. Thus, the effect of providing the subject with an initial positive instance in the present experiment may have been to increase the percentage of positive instances selected compared with the Bourne et al. experiment.

Experiment II: Effect of initial negative instance on attribute identification of concepts.

The suggestion was made in the preceding discussion that the type of initial instance may have an important effect on subsequent behaviour on an attribute-identification task. The particular effect proposed was that the nature of the initial instance tends to predispose subjects to select more instances of the same type on subsequent trials.

The type of initial instance may be important in two further ways. Firstly, the amount of information transmitted about the concept as measured by the number of possible concepts logically eliminated by the initial instance is likely to depend on whether this instance is positive or negative. For a stimulus universe having four three-valued attributes, an analysis of the number of possible concepts logically eliminated by an initial instance reveals that, with a positive instance, 6 concepts remain for conjunction, 30 for biconditional and inclusive disjunction, and 84 for conditional; with a negative instance, 24 possible concepts remain for conditional, inclusive disjunction, and biconditional, and 48 for conjunction. From this informational analysis it is expected that performance on all conceptual rules except conjunction should be improved when a negative instance is given initially instead of a positive instance. Secondly, evidence from Trabasso and Bower (1964b) and Bourne and O'Banion (1969) indicates that recall of the first instance presented during concept identification is better than that for any other instance. Thus, if an informational analysis is true, this observation should serve to make more marked the predicted improvement in performance with an initial negative instance compared with an initial positive instance for all rules except conjunction.

Experiment II was designed to investigate the effect of an initial negative instance on the number and type of instances selected during attribute identification of concepts involving four "basic" conceptual rules. In all respects except the initial instance presented, the instructions, procedure, and stimulus display were the same as those utilized in Experiment I. Subjects in this experiment were required to solve three problems. Data taken from the first three problems in Experiment I was used in part of the statistical analysis to directly compare the effects of initial positive and initial negative instances.

1. Method

a. Design and Subjects. - The subjects were 52 undergraduate students randomly chosen from the first-year psychology subject pool at the University of Adelaide. No subject had previously taken part in an experiment on concept attainment.

A 4 x 3 repeated-measures factorial design was used with the variables (a) conceptual rule (conjunction, inclusive disjunction, conditional, and biconditional), and (b) problems (three for each subject). Thirteen subjects were assigned to each of the four rules at random with the limitation that there be approximately equal numbers of each sex on each rule. The three problems involved different attribute values for each subject, these values being randomly chosen from the 54 possible pairs of attribute-value combinations of two attributes for all rules except conditional, where selection was made from the 108 possible combinations.

The initial negative instance for each problem was randomly drawn from the subset of negative instances possessing neither of the two relevant attribute values for conjunction and inclusive disjunction, and one of the relevant attribute values for conditional and biconditional.

b. Stimuli. - The same stimuli were used as in Experiment I.

c. Procedure and Instructions. - The details of the procedure and instructions were identical to those in Experiment I.

2. Results

The mean and standard deviation of the number of card choices to solution for each conceptual rule and problem are given in Table 5.

Conceptual Rule		Problem 1	Problem 2	Problem 3
Conjunction	Mean	26.6	17.7	22.2
	S.D.	14.2	5.5	7.9
Inclusive disjunction	Mean	25.3	22.1	21.5
	S.D.	12.8	15.2	12.1
Conditional	Mean	27.5	13.9	11.2
	S.D.	23.5	7.0	4.8
Biconditional	Mean	39.6	26.1	29.1
	S.D.	26.1	16.3	22.8

TABLE 5. Number of card choices to solution.

An analysis of variance was performed on this data transformed by a square root transformation in order to make the variances more homogeneous. The effect of conceptual rule was significant, $\underline{F}$ (3 , 48) = 3.67, $\underline{p} < .05$, and a significant difference was also found between problems, $\underline{F}$ (2 , 96) = 7.89, $\underline{p} < .001$. No significant interaction between conceptual rules and problems was obtained. A summary of this analysis may be seen in Table 6. A post hoc analysis using Newman-Keuls multiple-range comparisons showed: for rules, the only significant comparison was between the conditional and the biconditional ($\underline{p} < .05$); for problems, Problem 1 required more trials to solution than both Problem 2

($p < .01$) and Problem 3 ($p < .01$). The order of rules therefore with an initial negative instance was seen to be conditional, inclusive disjunction and conjunction, and biconditional.

Source of Variation	S.S.	D.F.	M.S.	F	P
Between subjects	200.67	51			
Rules	37.40	3	12.47	3.67	<.05
Subjects within groups	163.27	48	3.40		
Within subjects	187.39	104			
Problems	25.35	2	12.68	7.89	<.001
Rules X Problems	7.83	6	1.30	<1	
Problems X subjects within groups	154.21	96	1.61		

TABLE 6. Analysis of variance on number of trials to solution transformed by a square root transformation.

The instances selected were again divided into positive and negative instances on each problem for each rule and a summary in terms of the mean percentage of negative instances is given in Table 7. These percentages were compared with the percentage of negative instances occurring in the stimulus population for each type of concept using a sign test. This analysis indicated that a lower percentage of negative instances were selected than were present in the stimulus population for conjunction on all problems ($p < .05$), and on Problem 3 for inclusive disjunction ($p < .05$), while a higher proportion of negative instances were obtained for conditional on all problems ($p < .001$) and on Problem 2 for biconditional ($p < .05$).

Conceptual Rule	Problem		
	1	2	3
Conjunction	75.2	74.1	78.4
Inclusive disjunction	42.3	47.0	41.9
Conditional	48.3	59.5	63.8
Biconditional	43.1	49.7	50.8

TABLE 7. Mean percentage of negative instances selected.

A Spearman rank correlation was computed between the percentage of negative instances chosen and the number of trials to solution for each rule. Significant correlations were found for conjunction, rho (37) = 0.48, $p < .01$, and for conditional, rho (37) = -0.42, $p < .01$. Almost zero correlation was obtained for the other two rules; for inclusive disjunction, rho (37) = -0.04, and for biconditional, rho (37) = 0.00.

Instances were also analyzed into the four stimulus classes - TT, TF, FT, and FF. The mean percentages of each of the four classes of instance selected for each rule over the three problems are given in Table 8. These percentages were compared with the percentage of each type of instance occurring in the stimulus population. A sign test showed: for conjunction, a higher proportion of TT instances ($p < .001$) and a lower proportion of FT ($p < .01$) and FF ($p < .05$) instances were selected; for inclusive disjunction, the distribution of TT, TF, FT, and FF instances was not significantly different from the stimulus population; for conditional, a higher proportion of TF instances ($p < .001$) and a lower proportion of FT ($p < .001$) and FF ($p < .001$) instances were selected; and for the biconditional, the percentage of TT instances was greater ($p < .001$) and of FF instances less ($p < .01$) than the percentages of these instances naturally occurring.

Conceptual Rule	Stimulus Class			
	TT	TF	FT	FF
Conjunction	24.1	19.3	17.6	39.0
Inclusive disjunction	8.8	25.5	22.0	43.7
Conditional	13.1	57.2	4.5	25.2
Biconditional	21.2	26.4	21.4	31.0

TABLE 8. Mean percentages of four stimulus classes selected.

The correlation between the percentage of each class of positive instance and the number of trials to solution was computed for inclusive disjunction and biconditional using Spearman's method. For inclusive disjunction, no significant correlations were found: for TT instances, rho (37) = - 0.11; for (TF + FT) instances, rho (37) = 0.21. For biconditional, the correlations were significant: for TT instances, rho (37) = - 0.34, $p < .05$; for FF instances, rho (37) = 0.40, $p < .05$.

A comparison was made between the results obtained in this experiment and the data collected in Experiment I on the first three problems for each rule. An analysis of variance was performed on the number of trials to solution transformed by a square root transformation. Significant effects were obtained for initial instance, $F\ (1,96) = 16.01$, $p < .001$, conceptual rule, $F\ (3,96) = 4.38$, $p < .01$, and problems, $F\ (2,192) = 12.35$, $p < .001$. A significant interaction was also obtained between type of initial instance and conceptual rule, $F\ (3,96) = 9.32$, $p < .001$. A post hoc analysis using Newman-Keuls multiple range comparisons revealed more trials to solution were required when the initial instance was negative than when it was positive for conjunction ($p < .01$) and biconditional ($p < .01$), while for the conditional more trials were taken with an initial positive instance than with an initial negative instance ($p < .05$). No significant difference was obtained for inclusive disjunction.

Since more trials to solution were taken following an initial negative instance than an initial positive instance for conjunction and biconditional, an analysis was performed in order to determine whether the number of trials to solution following the first positive instance was different for problems having the two types of initial instances on these rules. Where the initial instance was negative, the mean number of trials to solution following the first positive instance encountered was: for conjunction, 11.9 on Problem 1, 5.5 on Problem 2, and 5.3 on Problem 3; for biconditional, 37.5 on Problem 1, 22.5 on Problem 2, and 27.1 on Problem 3. An analysis of variance showed this data was not different from the data obtained when the first positive instance was the initial instance for conjunction, but for biconditional problems, significantly more trials were required following the first positive instance encountered when this instance was not the initial instance as compared with the case where the initial instance was positive, $\underline{F}\ (1,24) = 5.69$, $\underline{p} < .05$. A similar analysis was conducted for conditional to determine whether the number of trials to solution required following the first negative instance encountered differed for an initial positive instance and an initial negative instance. For an initial positive instance, the mean number of trials following the first negative instance selected on conditional concepts was 22.9 on Problem 1, 21.8 on Problem 2, and 14.2 on Problem 3. No difference was found between this data and that for an initial negative instance by an analysis of variance.

Since no difference was found for conjunction and conditional in the analysis immediately above, a comparison was made for these rules of the number of card choices taken before arriving at the first positive and first negative instances, respectively, with that expected if the subject selected cards at random. It can be expected if cards were chosen randomly that a mean of 9 card choices to a positive instance would be required for conjunctive concepts, and a mean of 4.5 card choices would be needed to locate a negative instance for conditional concepts. The obtained mean number of instances for conjunction was 14.5 , which was significantly greater than expected for random selection according to a sign test ($\underline{p} < .01$). For conditional, no difference was found between the obtained mean of 4.2 instances and the mean expected for random selection.

Further analysis of the biconditional was carried out to see whether the number of trials to solution following the first TT instance encountered differed for problems having an initial TT instance (Experiment I) and an initial TF or

FT instance (this experiment). This analysis was suggested by the finding in both experiments that performance on the biconditional was directly related to the proportion of TT instances selected. Where the initial instance was negative (i.e., TF or FT), the mean number of trials following the first TT instance selected was 27.9 on Problem 1, 16.6 on Problem 2, and 21.9 on Problem 3. An analysis of variance comparing this data with the data from Experiment I revealed no difference for the number of trials to solution. A subsequent analysis was performed to determine whether the number of choices to the first TT instance in this experiment for the biconditional was different from the mean number of instances expected with random selection. If cards are selected at random, the expected mean is 9; the obtained mean over the three problems was 9.5 (median = 6). A sign test showed no difference between these means.

An analysis of variance was also used to compare the percentage of negative instances selected in this experiment with Experiment I. A significant effect was obtained for initial instance, $\underline{F}$ (1,96) = 224.68, $\underline{p} < .001$, conceptual rule, $\underline{F}$ (3.96) = 28.57, $\underline{p} < .001$, and the interaction between initial instance and rule was also significant, $\underline{F}$ (3.96)= 18.08, $\underline{p} < .001$. Newman-Keuls comparisons indicated that more negative instances were selected when the initial instance was negative than when it was positive for all conceptual rules ($\underline{p} < .01$).

3. Discussion

The results indicate that the nature of the initial instance has an important effect both on the type of instances selected and on subsequent performance. Firstly, the nature of the initial instance appeared to predispose the subject to select more instances of the same type as the initial instance for all rules, thus extending the original suggestion of Bruner, Goodnow, and Austin (1956). Secondly, for the stimulus population employed in this experiment and in Experiment I, an initial negative instance was found to retard performance for conjunctive and biconditional rules, but to improve performance for the conditional, compared with an initial positive instance. However, no difference was found between the two types of initial instances on attribute-identification difficulty for inclusive disjunction.

It is evident that the predictions about the effects on performance of the initial instance based on the information transmitted by the instance were not confirmed for all rules. Specifically, the predictions from this analysis were supported for conjunctive and conditional rules, but not for disjunctive and biconditional rules. This finding is in accordance with similar findings by Bourne and Guy (1968b), and Bourne, Ekstrand, and Montgomery (1969) in which the ratio of positive to negative instances for each rule was the same as in the present experiment.

For conjunction, better performers were found to choose a higher proportion of positive instances than poorer performers. Moreover, subjects given a negative instance to begin with, did not gain any information about the concept until they had located a positive instance. This is suggested by the finding of no difference between initial positive instance problems and initial negative instance problems for the number of trials to solution following the first positive instance. This conclusion is in agreement with the results of Nahinsky and Slaymaker (1970), who varied the number of negative instances prior to the first positive instance using the reception paradigm. It might be inferred from the present findings and the work of Nahinsky and Slaymaker that subjects are unable to make use of negative information in conjunctive concept identification. However, this inference has previously been shown to be false by Hovland and Weiss (1953) and Bourne and Guy (1968b) using sequences of negative instances only. Apparently, in a mixed sequence of positive and negative instances, subjects are unwilling to store the meagre amount of information which is gained from each negative instance when so much more information can be gained with minimal load on memory from one positive instance.

Parenthetically it may be noted in the present experiment on conjunction the number of choices required before a positive instance was located was more than if instances had been selected at random. This too may be a function of the type of initial instance. In this experiment the initial negative instance presented was always an FF instance. If subjects tend to select instances which show a high degree of similarity with the initial instance as measured by the number of attribute values in common with that initial instance, then instances will be drawn mostly from the FF, FT, and TF stimulus classes; only rarely will a TT instance therefore be chosen. In this regard it is noticeable that the percentage of FF instances selected in this experiment is very much greater than the percentage of FF instances chosen in Experiment I, while conversely the percentage of TT instances selected in Experiment I is substantially more than the percentage of TT instances selected in the present experiment. Perhaps if the initial negative instance were a TF or FT instance, fewer choices would be found prior to the first positive instance.

For conditional concepts, better performers were found to choose a higher proportion of negative instances than poorer performers. Moreover, subjects given a positive instance to begin with, did not gain any information about the concept until they had selected a negative instance. Again, this should not be interpreted to mean that subjects are unable to identify conditional concepts from positive instances only

(Bourne and Guy, 1968b). Rather it suggests that subjects are unwilling to process and store the small amounts of information from several positive instances when, with less information processing and storage, the same result can be achieved with a single negative instance.

For the biconditional, performance was found to be unrelated to the proportion of positive instances selected. A division of positive instances into TT and FF instances indicated, however, that better performers chose a higher percentage of TT instances and a lower percentage of FF instances then poorer performers. Moreover, subjects given an initial negative instance to begin with, apparently did not gain any information about the concept until they had selected a TT instance. No evidence is currently available showing whether or not subjects can identify biconditional concepts from sequences of instances not including TT instances. If they can, this means that they must be unwilling to utilize information from other instances until they have found a TT instance, presumably because information from other instances is somewhat more difficult to use in making inferences about the concept. The relative ease of handling TT instances was referred to in Chapter 2. One problem with this latter account concerns how the subject identifies a positive instance as TT or FF when no such feedback is provided by the experimenter.

Some ideal strategies can be suggested, but evidence is needed to show that these strategies are employed. Obviously further work must be done to provide answers to these questions.

For inclusive disjunction, the type of initial instance had no effect on overall performance. However, some contrasting results were obtained relating performance to the type of instances selected in this experiment and in Experiment I. In this experiment performance was found to be unrelated to the proportion of positive instances selected while a small but significant correlation in Experiment I suggested that better performers selected a greater percentage of positive instances than poorer performers. Moreover, in the present experiment performance was unrelated to the proportion of instances with both relevant attribute values (i.e., TT instances) and to the proportion of instances with one relevant attribute value only (i.e., TF and FT instances). In the previous experiment better performers were found to choose a higher percentage of TT instances than poorer performers. The results in Experiment I seem to confirm what Bourne and Guy (1968b) suggested for disjunctive concepts, "that subjects find it easier and more natural to use a positive as opposed to a negative focussing strategy or that the necessity to convert or transform negative information (what the attributes are not) to positive form contributes a significant additional complication to the task." In the present experiment, however, there appears to be no difference in the facility with which subjects utilize positive and negative instances. It would seem thus that the effect of a negative initial instance for this rule is to reduce the dependence on positive instances during learning which was evident when a positive initial instance was presented.

This effect seems also to be true for the other rules examined. Inspection of the correlation coefficients computed between the percentage of negative instances selected and the number of card choices to solution reveals that smaller positive correlations were obtained for conjunction, inclusive disjunction, and biconditional and a larger negative correlation for conditional for an initial negative instance compared with an initial positive instance. It appears therefore that when a negative instance is presented at the beginning of a problem instead of a positive instance, performance is less dependent on the proportion of positive instances and more dependent on the proportion of negative instances.

Examination of the relative difficulty of the four conceptual rules considered indicates that the order of increasing difficulty of rules found using a selection procedure differs according to the nature of the initial instance presented. In Experiment I with an initial positive instance, the order of increasing difficulty was conjunction, biconditional, inclusive disjunction, and conditional, an order which is similar to that found by Laughlin (1968b). With an initial negative instance, however, the order obtained was conditional, inclusive disjunction and conjunction, and biconditional.

An interesting finding is the extreme relative difficulty of the biconditional with an initial negative instance. This is reminiscent of the result of Bourne (1967) who presented subjects with instances drawn equally from the four stimulus classes TT, TF, FT, and FF. Since instances were randomised, this means, on the average, half of the biconditional problems began with an negative instance. Conventionally, experiments which have employed a selection procedure have presented a positive instance at the start of each problem. The present results suggest that when a negative instance is used initially, the order of difficulty is significantly different from when a positive instance is provided. A combination of the two types of initial instances may yield an overall order which is similar to that found by Bourne.

The results in the present experiment and in Experiment I on the effects of the type of initial instance presented on the attribute identification of concepts using a selection procedure cause some surprise when compared with the work of Giambra (1969). Giambra did report a significant interaction between the type of concept and the type of initial instance, but he attributed this merely to an increase in difficulty for conjunctive concepts with an initial negative instance and an increase in difficulty for disjunctive concepts with an initial positive instance. No completely satisfactory account of Giambra's result is possible without some information on the nature of the selections made by subjects.

A comparison between Bourne, Ekstrand, and Montgomery (1969) and the present experiments can also be made. It will be recalled that Bourne et al. used the selection paradigm but did not provide subjects with a sample instance at the start of a problem. Earlier in this chapter it was argued that, other things being equal, the percentages of each class of instance selected when no initial instance was presented should fall between the percentages of instances selected in the respective stimulus classes with an initial negative instance and an initial positive instance. In a comparison of Bourne et al's experiment with the present experiments, other things are not equal: Bourne et al. employed a random array of instances, whereas in the present experiments stimuli were systematically arranged. If the conclusions made on the effects of presenting or not presenting an initial instance are correct, then any deviations from expectation in the distributions of selections in the four stimulus classes should be due to the effect of the orderliness of the stimulus display.

A summary of the type of instances selected by subjects in Bourne et al's experiment is given in Table 9. Comparing the percentages in this table with percentages shown in Tables 4 and 8 reveals that in most cases the percentage of instances in the various stimulus classes chosen by subjects in Bourne et al. lie in between the percentages selected by subjects in Experiment I and Experiment II. Some notable exceptions, however, occur. More TT instances for conjunction, TF instances for conditional, and TT instances for biconditional were selected in both Experiments I and II than in Bourne et al..

Conceptual Rule	Stimulus Class			
	TT	TF	FT	FF
Conjunction	21.5	23.5	26.9	31.3
Inclusive disjunction	12.3	24.8	24.1	38.9
Conditional	13.5	27.6	18.9	39.9
Biconditional	12.4	20.5	19.9	47.2

TABLE 9. Mean percentages of four stimulus classes selected (adapted from Bourne, Ekstrand, and Montgomery, 1969).

The observation has been made in all three experiments that better performers chose more TT instances on conjunction, more TF instances on conditional, and more TT instances on biconditional than poorer performers. Thus, one prediction in a comparison of Bourne et al's results with the present results is that subjects performed better in the present experiments than in Bourne et al's, irrespective of the type of initial instance. This prediction is confirmed for all but conjunction with a negative initial instance. The failure by subjects in Bourne et al's experiment to select a higher proportion of the most informative instances must be attributed to the effect of the random arrangement of instances.

CHAPTER 4

STRATEGIES AND MEMORY IN CONCEPT IDENTIFICATION

In the previous two chapters an attempt has been made to account for concept identification in terms of a number of directly observable and controllable variables. An alternative approach in studies of concept learning has been to seek to infer the nature of the processes underlying learning and then to relate these processes to performance. At least three assumptions are fundamental to this approach. The first is that such underlying processes exist. Few theorists have questioned the reality of certain covert experiences or events, although in the absence of objective criteria for recognizing them, it may be wondered whether some postulated intervening processes are not mere inventions. Secondly, these underlying processes can be reliably identified. The necessity for objective criteria for recognizing such implicit processes has already been referred to in discussing the first assumption and problems with some attempts at inferring internalized activity will form part of later discussion in this chapter. Thirdly, it is assumed that conceptual behaviour cannot be fully explained without some assertions about intervening processes operating within the subject between input and output, which have the property of controlling and producing behaviour. This assumption has been challenged by some theorists (e.g.

Bourne, 1969) who contend that reference to hypothetical underlying mechanisms adds nothing to what can be ascertained about behaviour from a study of objective factors in the past and present states of the organism and his environment. An attempt will be made to evaluate a process approach in this chapter and the next chapter, and to consider the contributions, if any, it has made to an understanding of concept attainment.

The process of concept attainment

Bruner, Goodnow, and Austin (1956) were the first workers to attempt an analysis of the processes underlying concept attainment. Their aim was to externalize for observation as many of the decisions made prior to solution as possible in the hope that regularities in these decisions might provide the basis for making inferences about the processes involved in learning a concept. These regularities in decision-making they called "strategies". Strategies, it was suggested, may have at least three possible objectives each of which may be incompatible with the other:

(1) to maximize the amount of information gained from each instance chosen;

(2) to minimize the amount of strain on inference and memory capacity;

(3) to control the degree of risk involved.

Bruner et al's approach was firstly, to construct a set of ideal strategies which could be stated in strict logical terms, having the formal properties necessary to meet certain objectives. Much of their work was concerned with strategies for identifying conjunctive concepts and four ideal strategies were proposed using the selection paradigm. Reception strategies were also suggested but in this discussion we shall confine ourselves to the selection situation which allows the subject more freedom in his choice of instances. The four ideal selection strategies were labelled simultaneous scanning, successive scanning, conservative focussing, and focus gambling.

In the simultaneous scanning strategy, after the first positive instance has been given, all possible hypotheses are enumerated. Each succeeding instance is then chosen to eliminate the greatest number of hypotheses still remaining. The main advantage of this strategy is to attain the concept in a minimum number of instances. However, a very considerable amount of strain on inference and memory capacity is produced since many independent hypotheses must be dealt with and remembered.

In the successive scanning strategy a subject tests one hypothesis at a time and limits his choices to those instances that provide a direct test of his hypothesis. This technique for choosing instances cannot assure that the subject will encounter instances containing the maximum information possible since he is quite likely to select logically redundant cards some feature of which has been used to test some previous hypothesis.

The load on memory and inference is less for this strategy than for simultaneous scanning since the subject is required to test and remember only one hypothesis at a time, and no retention of past instances is assumed. Successive scanning has little utility in regulating risk.

The conservative focussing strategy may be described as finding a positive instance to use as a focus, then making a sequence of choices each of which alters one attribute value of the first focus card and testing it to see if the change yields a positive or negative instance. Those attribute values of the focus card which, when changed, still yield positive instances are not relevant to the concept. Those attribute values of the focus card that yield negative instances when changed are relevant to the concept. If no positive instance is initially presented, this strategy may be highly inefficient since it requires of necessity a positive instance to use as a focus. Given a positive instance, however, the strategy is very efficient whilst imposing very little strain on inference and memory. As its name implies, it ensures that every choice yields information about the concept.

In the focus gambling strategy, the subject again uses a positive instance as a focus, but then changes more than one attribute value at a time. If two attributes are changed and the instance is positive, then both attribute values are irrelevant, so that with this strategy two attribute values may be eliminated in one choice instead of only one as with conservative focussing. However, if the instance is negative, then either one or both attribute values are relevant so that little information is gained. Thus this strategy allows for some regulation of risk. Cognitive strain with this strategy is similar to that for conservative focussing.

The next step in Bruner et al's analysis was to identify focussing and scanning strategies. A focusser was defined as a subject whose choices in the main varied only in one attribute value from those attribute values of the focus card that had been found relevant or were still untested. The only other types of choices permitted subjects in this group were redundant choices where the third value of an attribute was checked, or focus-gamble choices shown by later choices to be utilized as such. When the majority of the choices was of these types, the subject was considered a focusser. All others were treated as scanners.

Using this method for identifying strategies, Bruner et al. noted several variations in the types of choices made by subjects in the focussing and scanning groups from those expected according to ideal focussing and scanning strategies. With conservative focussing two modifications were evident. Firstly rather than alter an attribute value of the focus only once, subjects sometimes showed a tendency to test a second alteration.

Bruner et al. refer to this as a "thirst for confirming redundancy". Secondly, some subjects displayed a tendency to change their focus card from the original positive instance to some other positive instance. The chief modification to successive scanning consisted in subjects attempting to remember as many instances encountered as they could. No subject was reported to employ an ideal simultaneous scanning strategy.

The incidence of focussing and scanning was found to vary according to the conditions of the task. Thus, if stimuli were arranged in an ordered array, subjects tended to adopt focussing strategies. Conversely, with random arrays, scanning was more evident. If the array was one suggesting a particular grouping of stimuli as in the case of concrete or thematic material, then the subject tended to utilize a scanning strategy which took into account such expected groupings.

Relating strategies to the number of choices to concept attainment showed that focussing and scanning gave rise to different levels of performance. With an abstract, ordered array of stimuli, focussers were found to require fewer choices to solution than scanners. When the stimuli were perceptually unavailable, the difference in efficiency of the two types of strategies was even more marked. Scanners showed a decrement in performance under such conditions, while focussers showed little or no change. A similar picture was evident when the performance measure was the number of redundant choices to solution.

Criticisms and alternatives

Both theoretical and empirical objections to Bruner et al's analysis have been made by Eifermann (1965a). Her theoretical criticism was to correct an impression given that simultaneous scanning is more efficient in terms of information-getting than conservative focussing. In fact, it is not possible to distinguish the two strategies on this criterion.

Her criticism of the method used by Bruner et al. to infer strategies from the choices made by subjects is more telling. She has pointed out that it is not possible to differentiate operationally between a sequence of choices representing simultaneous scanning and one representing conservative focussing. Moreover, the same sequence of choices may be produced on some occasions by successive scanning as by conservative focussing, and on other occasions successive scanning may be indistinguishable from focus gambling. In other words, Bruner et al's procedure for classifying the types of choices generated by subjects into one of the ideal strategies proposed is invalid.

Other objections have also been raised that the method for identifying strategies is lacking in objectivity, that too much reliance is placed on inference by the experimenter. Hunt (1962) has suggested that two independent judges should be used to determine strategies. However, from the criticism made by

Eifermann above, no number of judges could reliably or validly classify subjects as employing one of Bruner et al's ideal strategies.

Byers (1963) has proposed that an objective classification of strategies is possible if a strategy is redefined as the selection of cards from certain response classes, which differ according to the number of attribute values changed from the focus card. As a technique for making inferences about the process of concept attainment, Byers' procedure suffers from several deficiencies. Firstly, it assumes that the subject uses the given instance as a focus card for the whole of a problem, even though Bruner et al suggested that sometimes subjects changed their focus. Secondly, it provides no indication of the redundancy of card choices also noted by Bruner et al. Thirdly, it assumes that only one strategy may be employed during any one problem, an assumption which has been found to be not always true (Bruner et al., 1956; Eifermann, 1965b). These deficiencies undoubtedly arise because Byers' redefinition of strategy is clearly not intended to be used as a basis for making inferences about some intervening process. Rather his definition makes strategy a directly measurable index of overt behaviour during concept attainment. Perhaps it would be better, therefore, if a different label to strategy was employed in this case. The need to redefine strategy in this way may be an admission of the likelihood of success of a process approach to concept attainment.

Laughlin (1965, 1968b) has developed a quantitative index of strategy usage by conceiving of focussing as a continuous dimension rather than a strict dichotomy with scanning. On this proposal a greater usage of the focussing strategy by a subject reflected a lesser usage by that same subject of the scanning strategy. The amount of focussing was scored according to two rules; (1) each card choice had to obtain information on a new attribute; and (2) if an hypothesis was made, it had to be tenable considering the information available. The number of trials which satisfied both of these two rules was divided by the total number of trials to solution to obtain a focussing score from zero (no focussing) to one (complete focussing). This approach by Laughlin, whilst objective, still does not overcome the criticism of Eifermann on how to differentiate between conservative focussing and simultaneous scanning; both ideal strategies would be expected to yield a focussing score of one. Moreover, Laughlin's rules for scoring focussing seem to be more a measure of the efficiency of performance than of the process involved in arriving at the concept, a possibility admitted by Laughlin himself (Laughlin, 1968b). Thus, it is not surprising that more difficult concepts show less focussing.

Giambra (1971a, 1971b) has suggested that what Laughlin is measuring is really the subject's de facto strategy rather than his intended one. The intended strategy used by a subject may be ascertained, according to Giambra, by requiring the subject in addition to indicate whether he expects the selected card to be a positive or a negative instance. The justification of this claim is given in the form of an example. Consider a subject who, in attempting to identify a conjunctive concept, has selected an instance which varies from the initial positive instance by two attribute values. If this selected instance is positive, then a considerable amount of information is gained about the concept, viz., that the two attribute values changed are irrelevent. If, on the other hand, the instance is negative, very little information about the concept is acquired. Giambra argues that a subject using a focussing strategy must have expected the card selected to be a positive instance. Therefore, he deduces, if the subject intended the card to be negative, then he could not have been using a focussing strategy. By implication he must have been using a scanning strategy. Giambra does not specifically state his conclusion if the subject intended the instance to be a positive instance, but it is apparent that he would classify this as indicative of a focussing strategy. This assumption, unfortunately, overlooks the assertion of Bruner et al. that subjects using successive scanning prefer direct tests of their hypothesis. If Bruner et al. are correct, this means that, like focussing subjects, scanning subjects expect most of their instances selected also to be positive.

Thus, having the subject indicate whether he expects the instance selected to be positive or negative still does not enable an operational distinction to be made between focussing and scanning strategies. The same objections therefore apply also to Giambra's approach as to Laughlin's.

The failure of attempts to differentiate between strategies on the basis of the sequence of card choices forces a retreat to more subjective methods. Eifermann (1965b) has suggested that one approach to the identification of strategies is to analyze not only the choices made, but also the reasons the subjects give for making these choices. The use of subjective reports has been rejected by many people in the past because of the belief that subjects who successfully solve problems are often unable to describe how they arrived at their solution (e.g., Bruner et al., 1956). Contrary to this common observation, Eifermann claimed that the adequacy of the verbal report in tasks which require logical thinking is positively correlated with efficiency in performance. Using this method for inferring strategies, therefore, Eifermann noted in line with Bruner et al's classification that subjects may be broadly grouped into those who employ a concept-centred approach (equivalent to a scanning strategy) and those who employ a component-centred approach (equivalent to a focussing strategy). In the early stages of the task, however, considerable variation in the type of responses and reports given was evident even for a single subject and it was only after being exposed to several problems of the same type that the subject developed any consistent strategy.

In summary, severe criticisms have been made of the reliability and validity of the inference of strategies in conjunctive concept attainment. It seems that no satisfactory description of such processes is possible without reference to subjective reports. Because of the tentativeness of process descriptions obtained in this way, no evidence is available on the relationship between strategies and performance. Indeed, it is likely that Eifermann's method of inferring strategies may itself have an effect on performance. Byers and Davidson (1967) have shown that concept identification is facilitated when subjects are required to state an hypothesis of the concept on each trial. A pilot study by the present author using the Eifermann technique confirmed her classification of strategies not only for conjunction but also for inclusive disjunction as well. However, comments by subjects after the experiment seemed to indicate that requiring a reason for each card choice affected performance either by suggesting a solution which may not otherwise have been seen or, in some cases, by interfering with the process of evaluating and storing information about the concept.

Experiment III: Selection strategies and performance following training on 25 attribute-identification problems for four conceptual rules.

Perhaps an alternative method of assessing the strategies employed by subjects is to ask the subject at the end of a problem to describe how he would tackle another problem of the same type if one were presented. This method suffers from the disadvantage observed by Eifermann (1965b) that the process of concept attainment for early problems tends to be quite variable. Accordingly, it is unlikely that many subjects could state their method of attack until they had some considerable practice at solving problems of the same type. An experiment was conducted in which subjects were given practice at solving a series of 25 successive problems before they were asked to describe their strategy. Clearly any strategies inferred after this amount of practice are going to differ from the sorts of strategies identified for naive subjects. The major source of difference, it is hoped, will be the elimination of any non-systematic or redundant features of responding. There is also the possibility, of course, that some subjects may have developed a completely new approach to a problem from the one utilized at the beginning of the series. Strategies inferred, therefore, cannot be regarded as necessarily indicative of the types of strategies utilized by naive subjects. Nevertheless, classification of strategies by matching with Bruner et al's ideal strategies, for example, should be somewhat more reliable following practice than in the absence of practice effects.

If this technique is successful, the degree to which performance may be predicted from the strategy employed can be evaluated. Providing subjects with training on the task has an additional advantage here in that performance can be expected to reach a stable or asymptotic level which should facilitate comparisons between actual performance and performance predicted by the subject's strategy.

Four conceptual rules were examined. From Bourne (1967), it might be expected that any differences in difficulty between rules which are evident on initial problems should disappear with practice. A consideration of the ideal strategies possible for each rule, on the other hand, leads to a different prediction. According to an optimal (maximum information per choice, no loss of information through forgetting, no risk) focussing strategy, the order of difficulty of rules after training should still be conjunction, biconditional and inclusive disjunction, and conditional, as estimated by the minimum number of trials to solution necessary. This order is calculated for an orderly arranged stimulus display having four attributes with three values per attribute. The initial instance on each problem may be drawn from any of the positive stimulus classes for each rule: TT for conjunction, TT, TF, and FT for inclusive disjunction, TT, FT, and FF for conditional, and TT and FF for biconditional.

If the prediction derived from a consideration of ideal strategies is true, then this negates the simple familiarity hypothesis about differences between rules being due to differing familiarity with the rules and their appropriate strategies (Haygood and Bourne, 1965). Of course, the familiarity hypothesis can be modified to overcome this difficulty. This is to suggest that the number of choices which are redundant to an appropriate strategy can be eliminated by practice. At any rate, any account of differences between rules which is based on the notion of familiarity is scientifically untestable since a failure to eliminate the differences between rules with practice could always be attributed to the possibility that a difference in familiarity still existed between the rules.

Ideal strategies for the four rules are outlined below. In line with the formulation of Bruner et al., strategies are divided into hypothesis-testing or scanning and attribute-testing or focussing. Little needs to be added here to the description of scanning given by these authors. Simultaneous scanning and successive scanning seem equally applicable in the form described to all conceptual problems, irrespective of the rule. The same is not true, however, for focussing in which the ideal plan shows considerable variation according to the rule. Moreover, even for the same rule, a strategy which is appropriate for one type of initial positive instance is sometimes inappropriate when other types of initial positive instances are presented. What is needed, of course, is a strategy which will produce the concept under all conditions of initial instance. The objectives of the task may be assumed to be to find the concept in the minimum number of choices, with minimum load on memory and inference, and without risk.

For conjunction, the ideal strategy given these objectives is conservative focussing. On this strategy, cards are selected which vary from the focus card by one attribute value. If a card chosen in this way is positive, then the changed attribute value is irrelevant; if negative, the changed value is relevant. If after two choices, two negatives or two positives have been found, the concept consists of the two attribute values in the focus found relevant or the two attribute values in the focus not found irrelevant respectively. A further choice is necessary if one positive and one negative card are found in the first two choices. Since half the problems on the average are solved in 2 choices and the other half in 3 choices, the mean number of choices with an ideal strategy is 2.5 choices.

For inclusive disjunction, the ideal focussing strategy depends on whether the initial instance contains both or one relevant attribute value. Thus the first step is determine the nature of the initial instance. This is done by selecting two cards which do not have any attribute values in common with each other or with the focus card. If these two cards are negative, then the initial instance contains both relevant attribute values. If only one is negative, then the initial instance contains only one relevant attribute value. If the former is the case, then the concept may be found by choosing cards which vary all but one attribute value of the focus. If a positive instance results, then the unchanged value is relevant; if negative, then the unchanged value is irrelevant. In this way the concept may be found on the average after 4.5 choices. When the initial instance contains only one relevant value, the focus must be changed to the negative instance already selected. Cards are then selected which vary one attribute value at a time from the new focus. If a positive instance results, the changed value is relevant; if negative, the changed value is irrelevant. All values on each attribute are tested until one relevant value is found. If this value is present in the initial instance, then the remaining values tested need be only those not in the focus nor in the initial instance. Conversely, if the obtained relevant value is not in the initial instance, then only those values in the initial instance need be tested to find the remaining relevant value. Using this strategy, the concept may be found after an average of 5.9 choices.

For conditional, two parts can be distinguished in the ideal strategy: (1) a search for a negative instance, and (2) identification of the relevant attribute values using the negative instance as a new focus. The strategy outlined here works for the three possible initial instances, but different mean performances are predicted for each. To find a negative instance, cards are chosen which vary all attribute values but one in the initial instance. If no relevant values are present in the initial instance, then an average of 2.6 choices are required; if one relevant value is present, 2.0 choices are required; and if both relevant values are present, 4.0 choices will be required on the average. Having located a negative instance, relevant attribute values may be found by using this negative instance as a new focus and varying one attribute value at a time, making a direct test of each value of each attribute. If all values of an attribute yield negative instances, then the attribute is irrelevant. If one value of an attribute yields a positive instance, then this value is the "then" part of the concept. If two values of an attribute produce positive instances, then the value of this attribute in the focus card is the "if" part of the concept. This strategy requires a mean of 5.2 trials for identifying the concept. Thus, if no relevant values are present in the initial instance, the overall mean number of choices required is 7.8; if one value is present, the mean is 7.2; if both are present, the mean is 9.2.

For biconditional, an ideal focussing strategy can be developed which can be used to identify the concept irrespective of the type of initial instance, but once again performance is predicted to differ for the two types of initial instances. Using the initial instance as a focus, instances are selected which vary one attribute value at a time so as to test all values of each attribute. If all tests for a particular attribute are positive, then that attribute is irrelevant. If two of the three tests for an attribute yield positive instances, then the attribute value which gives a negative instance is relevant. If one of the three attribute values yields a positive instance (i.e., the initial instance), then that value is relevant. On this strategy a mean of 4.5 choices are required to solution if both attribute values are present in the initial instance; if, however, neither of the relevant values are present in the focus card, 5.9 choices are required.

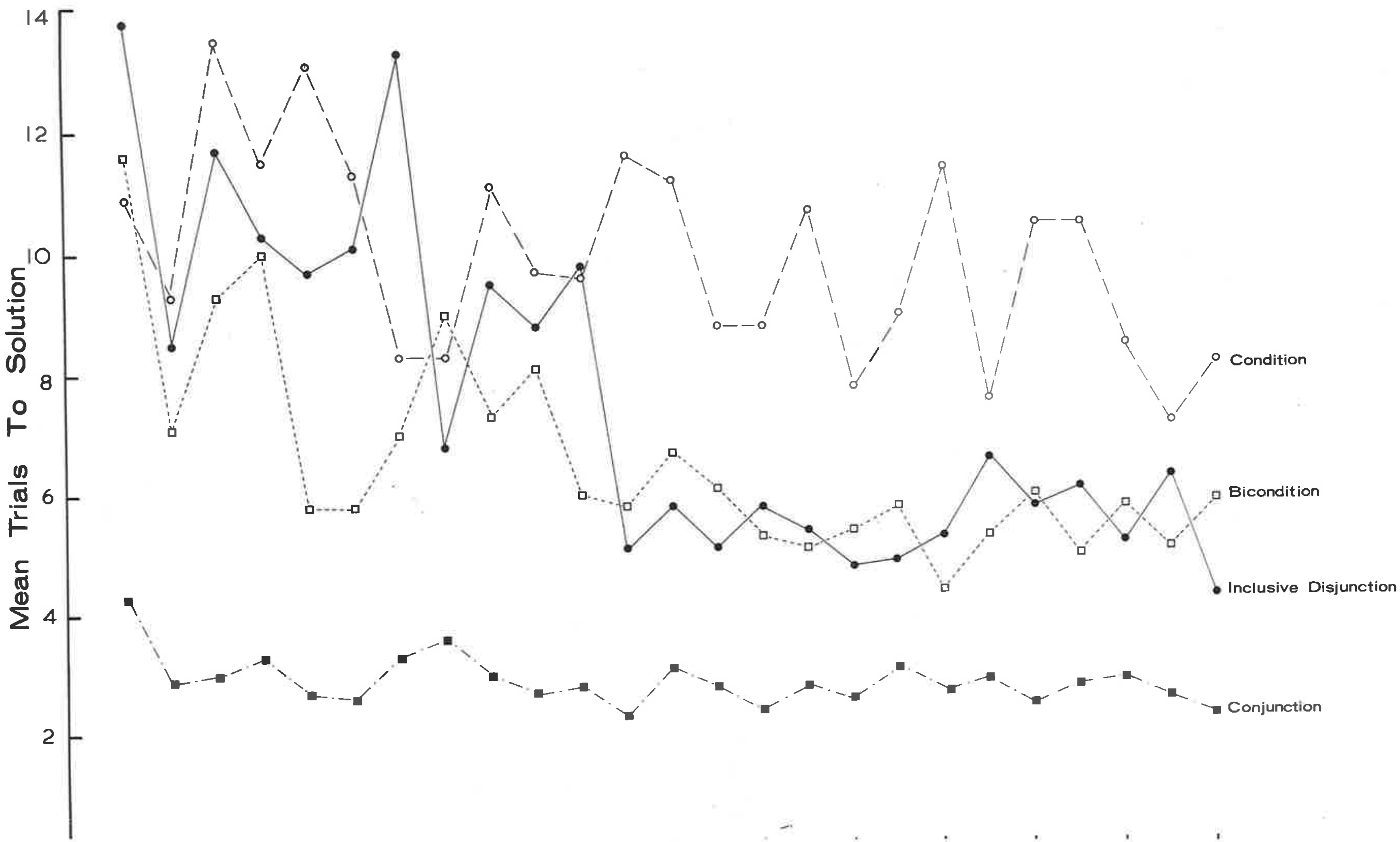
In the experiment which follows, on half of the problems the initial positive instance was drawn from the subset of positive instances containing both relevant attribute values. For the remaining problems the nature of the initial positive instance was dependent on the type of concept: for conjunction, the initial positive instance on the remaining problems also contained both relevant values; for inclusive disjunction, the remaining problems had initial instances with only one relevant attribute value; for the conditional, half the remaining problems had initial positive instances with one relevant value, while half had neither relevant values; for the biconditional, neither relevant values were present in the initial positive instance for the remaining problems. These proportions were adopted arbitrarily. This means that the mean number of choices predicted by an ideal focussing strategy is 2.5 for conjunction, 5.2 for biconditional and inclusive disjunction, and 8.4 for conditional.

1. Method

a. Design and Subjects. - This experiment was conducted as a class practical for third-year psychology students at the University of Adelaide. The kind permission of Professor M. A. Jeeves in running this experiment is acknowledged. The 48 students who took part as subjects were not completely naive to concept learning tasks.

A 4 x 25 repeated-measures factorial design was used with the variables, (a) conceptual rule (conjunction, inclusive disjunction, conditional, and biconditional), and (b) problems (25 for each subject). Twelve subjects were assigned to one of the four rules. The attribute values for the 25 concepts were randomly chosen for each subject from the 54 possible pairs of attribute-value combinations of two attributes. Two experimental sessions of approximately two hours each,

FIGURE 2. Effect of training on 25 problems on the mean number of card choices to attribute identification for four conceptual rules.



14
12
10
8
6
4
2
Mean Trials To Solution
Condition
Bicondition
Inclusive Disjunction
Conjunction

one week apart, were required for each subject.

The initial positive instances on each problem were randomly chosen with the following provisions. Half the problems for each rule began with a TT instance. For the remaining problems the type of initial instance presented was dependent on the rule: for conjunction, these were TT instances also; for inclusive disjunction, the remaining problems had TF or FT initial instances; for the conditional, half the remaining problems began with FT instances, while the other half began with FF instances; for the biconditional, FF instances were presented on the other problems. The order of the different types of initial positive instances was randomized for each subject.

b. Stimuli. - The same stimuli were used as in Experiment I.

c. Procedure and Instructions. - The details of the procedure and instructions were the same as in Experiment I.

In addition, after 25 problems had been completed, the subject was asked to write down an account of the approach he would use to find the concept if he was presented with another problem similar to those he had solved.

2. Results and Discussion

The mean number of card choices to solution on each of the 25 problems for the four rules may be seen in Figure 2. An analysis of variance on this data transformed by a square root transformation so as to make the variances more homogeneous revealed a significant effect of rules, F (3,44) = 52.40, $p < .001$, of problems, F (24,1056) = 5.36, $p < .001$, and also a significant interaction between rules and problems, F (72,1056) = 1.56, $p < .05$. From the graph this interaction seems to be due to inclusive disjunction and biconditional having similar difficulty to conditional on early problems, but having less difficulty on later problems.

The results for the last 5 problems are summarized in Table 10 in terms of the mean and standard deviation of the number of card choices to solution for each conceptual rule and problem. An analysis of variance was carried out on this data transformed by a square root transformation. The effect of conceptual rule was significant, F (3,44) = 40.29, $p < .001$. No significant difference was found between problems, nor was there a significant interaction effect between rule and problems. A summary of this analysis is given in Table 11. A post hoc analysis on rules using Newman-Keuls multiple range comparisons revealed significant differences between conjunction and other conceptual rules ($p < .01$), between biconditional and conditional ($p < .01$), and between inclusive disjunction and conditional ($p < .01$). Thus the ascending order of difficulty of rules after 25 problems is conjunction, biconditional and

	Problem					
	21	22	23	24	25	Average
Conjunction						
Mean	2.5	2.8	2.9	2.6	2.3	2.6
S.D.	0.7	0.6	1.0	0.8	0.6	0.8
Inclusive disjunction						
Mean	5.8	6.1	5.2	6.3	4.3	5.5
S.D.	2.5	2.6	1.5	2.1	2.3	2.2
Conditional						
Mean	10.5	10.5	8.5	7.2	8.2	9.0
S.D.	5.9	6.4	3.7	3.1	3.0	4.6
Biconditional						
Mean	6.0	5.0	5.8	5.1	5.9	5.6
S.D.	3.2	1.9	1.0	1.9	1.6	2.0

TABLE 10. Mean and standard deviation of number of instances selectea to solution on the last five problems.

inclusive disjunction, and conditional, which closely resembles the order obtained in Experiment I and in Laughlin (1968b, 1969b). The finding of no difference between problems suggests that subjects have reached a stable or asymptotic level in their performance. It is evident therefore that a simple familiarity hypothesis cannot account for differences between rules on attribute-identification.

Source of variation	S.S.	D.F.	M.S.	F	P
Between subjects	67.63	47			
Rules	49.58	3	16.53	40.29	<.001
Subjects within groups	18.05	44	0.41		
Within subjects	49.34	192			
Problems	1.04	4	0.26	1.02	N.S.
Rules X Problems	3.45	12	0.29	1.13	N.S.
Problems X Subjects within groups	44.85	176	0.25		

TABLE 11. Summary of analysis of variance on the trials to solution data for the last five problems transformed using a square-root transformation.

It may also be noted that the obtained order of rules is the same order which was predicted by ideal focussing strategies. Moreover, the mean number of choices required over the last five problems closely resembles the predicted mean for the ideal strategies. It should not be assumed from this, however, that subjects are necessarily utilizing an ideal strategy or even a focussing strategy on each rule. Examination of the subjective reports gathered at the end of the 25 problems indicates, nevertheless, that in almost every case subjects did employ a focussing approach and in some instances these resembled the ideal strategies outlined in the introduction to this experiment.

For conjunction, 11 of the 12 subjects reported using a strategy similar to conservative focussing, that is, an ideal focussing strategy . The remaining subject utilized a simultaneous scanning strategy where all possible hypotheses remaining after the given instance were tested. The observed mean of 2.6 trials to solution over the last five problems closely approximates the mean number of 2.5 trials expected theoretically for both conservative focussing and simultaneous scanning.

For biconditional, three strategies were reported. Seven subjects adopted the ideal focussing strategy. One subject employed a similar approach except that he changed his focus to the first instance which indicated one of the relevant attribute values. Four subjects began by assuming that the given instance contained both relevant attribute values. Using this instance as a focus, they employed a conservative focussing strategy similar to that found for conjunction. If three positive instances were obtained, then neither of the relevant values were present in the given card. In this case a direct test was made on each attribute value as in the ideal strategy. On the other hand, if two negative instances were obtained, it was concluded that the two relevant attribute values were present in the given instance. These values were identified to be those values which, when changed, yielded negative instances. It will be realised that the last strategy involves an incorrect assumption which would lead to an occasional incorrect hypothesis, viz., that where two negative instances were obtained, the relevant attribute values were present in the initial positive instance. A comparison between the mean of 5.6 trials to solution over the last 5 problems and the expected mean of 5.2 trials required by the first or ideal strategy utilized optimally with no loss of information through forgetting indicates again a close approximation between observed performance and that theoretically expected from a consideration of strategies.

For inclusive disjunction, four kinds of strategies can be identified from the subjective reports. None utilized the ideal strategy described earlier. Six subjects reported using a strategy which involved changing one attribute value at a time using the previous card tested as a focus card each time. Whenever a change in an attribute value of a positive instance resulted in a negative instance, that attribute value was relevant. Whenever a change in an attribute value of a negative instance resulted in a positive instance, the value of the attribute in the positive instance was relevant. Two subjects reported using the given card as a focus, and changing one attribute value at a time until a negative instance had been found. That attribute value which, when changed, yielded a negative instance, was relevant. The second relevant attribute value was then determined by taking the negative instance as a new focus, and changing values of other attributes, one at a time, until a positive instance was located. It may be seen that the method employed here for seeking a negative instance would be unsuccessful whenever the initial positive instance contained both relevant attribute values. Three subjects adopted the given card as a focus, and changed three attribute values at a time, holding one attribute value constant on each trial. If two positive and two negative instances resulted, then the given instance contained both relevant attribute values, and these were determined to be those values which, when unchanged, yielded positive instances. If this result did not occur, the relevant attribute values were found by using a negative instance as a new focus, and changing one attribute value at a time with a direct test of each attribute value. It may be seen that this strategy too may occasionally lead to an incorrect hypothesis when the initial instance contains only one relevant value. The fact that two positive and two negative instances result from the first part of the strategy does not necessarily mean that both relevant values were present in the initial instance. One subject elected to change all attribute values from those present in the given card. This was done twice, using different attribute values in each case. If both card choices proved negative, then the initial instance contained both relevant values; if only one was negative, then the initial instance contained one relevant attribute value. With this information, individual hypotheses were tested until the concept had been attained. It is difficult to provide a reliable prediction of the mean trials to solution for these strategies, but the third strategy (excluding the possibility of an incorrect hypothesis) requires a mean of 5.7 trials to solution which can be compared with the obtained mean of 5.5 trials over the last 5 problems.

For the conditional, as in the ideal strategy two parts can be distinguished in the strategies reported: (1) a search for a negative instance, and (2) identification of the relevant attribute values. No completely systematic search strategies were used to locate a negative instance. Six subjects began by adopting the given positive instance as a focus card and varied one attribute value at a time until a negative card was selected. If this method was unsuccessful, more than one attribute value was changed and instances were chosen in an apparently unsystematic fashion. The remaining subjects did not utilize any clearly classifiable procedure to find a negative instance. An estimate can be made of the mean number of trials required to locate a negative instance for those problems in which both relevant attribute values ($\bar{x}$ = 2.5 trials) or neither relevant attribute values ($\bar{x}$ = 4.5 trials) were present in the given card. Where one relevant attribute value was present in the given card, changing one attribute value at a time from the given card will never produce a negative instance. In the present task, therefore, a mean of 3.2 trials should be needed to discover a negative instance for two-thirds of the problems; no reliable estimate can be made for the remaining problems. Having found a negative instance, eight subjects used a negative focussing strategy similar to that found in the ideal strategy described earlier. This strategy requires an average of 5.2

trials for identifying the concept. Other strategies employed a method of testing single hypotheses or were unclassifiable. Thus, the best estimate of the mean number of instances selected for a conditional concept is 8.4, which compared with the obtained mean of 9.0 instances over the last 5 problems.

These therefore were the selection strategies inferred from the subjective reports given by subjects at the end of training on 25 problems. From the instructions presented to the subject at the beginning of the experiment, it may be assumed that the subject had at least two objectives on the task: (1) that each selection should enable the maximum possible information about the concept to be gained, and (2) that the concept should be attained without risk of error. A third objective may also be suggested, namely, that the concept should be attained with the minimum load on memory and inference. Some observations can be made about the strategies inferred having these objectives.

Firstly, the nature of the strategies for any concept type shows some variation between individual subjects on the preferred method of attack on a problem. Furthermore, it is apparent that the variety of strategies for identifying a particular type of concept increases as the difficulty of the concept type increases. Not all of these strategies, however, were found to be effective under all conditions. The main factor determining the effectiveness of a strategy seems to be the nature of the initial positive instance. In this regard, a Wilcoxon signed-ranks test indicated that, for biconditional significantly more trials to solution were required over 25 problems when neither of the relevant attribute values were present in the initial instance than when both relevant attribute values were present ($p < .01$); for conditional, fewer trials were needed when both of the relevent attribute values were present in the initial instance than when one or neither were present ($p < .05$); no difference in performance was evident for inclusive disjunction between the two types of initial positive instances. This result extends the evidence obtained in the previous chapter relating performance to the type of initial instance presented. It may also be compared with Haygood and Devine's (1967)finding that attribute identification of biconditional and inclusive disjunctive concepts was improved when the proportion of instances with both attribute values was increased.

Secondly, the subjective reports of strategies indicate that most subjects prefer a focussing technique to attain concepts. They do not, however, suggest that conceptual rules differ in the extent to which focussing is employed, as has been postulated by Laughlin (1968b). This may simply mean that with practice subjects are able to apply focussing techniques equally to all rules. Alternatively it may mean that Laughlin's assertion is a function of his index of the amount of focussing which has already been shown to be an invalid measure of strategy usage. On conjunction and biconditional a positive instance was commonly taken as a focus; on conditional, subjects reported seeking a negative instance which could then be used as a focus; on inclusive disjunction, there was no clear predilection for either positive or negative instances to be used as foci. These findings are compatible with the results obtained in the previous chapter. There it was noted that subjects on conjunction did not gain any information about the concept until they had located a positive instance. Presumably the positive instance once encountered was used as a focus to determine the relevant attribute values. Similarly on the conditional no information about the concept was gained until a negative instance was found which, it would seem was then used as a focus. On the biconditional, the critical instance was a positive instance containing both relevant attribute values. From the strategies reported in this experiment subjects still prefer to assume a TT positive instance on this rule, although several apparently have learned with practice to use an FF positive instance as a focus. The finding of no clear preference for a positive or a negative instance for a focus in this experiment for inclusive disjunction is again consistent with the earlier evidence where it was shown that no difference in performance resulted for this rule when the initial instance was positive or negative.

Bruner et al. (1956) and Byers (1963) have observed with conjunctive concepts that subjects display a marked tendency to select cards which differ in only one attribute value from the focus card. Bruner et al. also noted this tendency with disjunctive concepts. The present study suggests that the same tendency is also to be found for other conceptual rules. The existence of this feature in the behaviour of subjects accounts for the absence of anything approximating a complementary focussing strategy suggested by Laughlin (1968b). The same modifications reported by Bruner et al. to a conservative focussing approach to conjunctive concepts are also observable here, including changing of the focus and direct testing of all attribute values. Whilst these modifications may serve no essential purpose for conjunctive concepts, they are frequently crucial to successful attainment of other types of concepts.

Strategies and memory effects

In the previous section differences in difficulty between rules were accounted for in terms of the number of choices required by certain inferred focussing strategies. It was noticeable that almost all subjects elected to utilize a focussing approach. Several reasons may be offered for this finding. Firstly, Bruner et al. suggested that an ordered array of instances produced more focussing than scanning, while scanning was more evident with a random array. In the previous experiment an ordered array was presented. A second possible reason is that scanning is more likely when subjects are required to state their hypothesis on each trial. A further possibility is that the presentation of an initial instance is more conducive to focussing than scanning. Another reason may be that focussing and scanning have different memory requirements. Byers, Davidson, and Rohwer (1968) showed that subjects given instructions on the use of the conservative focussing strategy for attaining conjunctive concepts showed no improvement in performance when provided with a memory aid, whereas the performance of subjects not given focussing instructions was facilitated when a memory aid was available. Bruner et al. also reported no decrement in performance for focussers required to solve problems "in-the-head", but a decrement was found for scanners under the same conditions.

This last point may be relevant to an experiment by Laughlin (1968b) who found that providing subjects with pencil and paper during an attribute-identification task did not facilitate their performance compared with a control group of subjects not permitted to use such a memory aid. The arrangement of stimuli and the presentation of an initial positive instance suggests that his subjects may have been predominantly using focussing strategies, although they were required to state an hypothesis on each trial. If a focussing strategy was employed, then this may account for the failure to find any improvement with pencil and paper as a memory aid. Alternatively the failure may be simply due to the ineffective use of the memory aid. Since Laughlin does not indicate the ways in which the pencil and paper were utilized, it is difficult to evaluate this possibility.

From the experiments reviewed in Chaper 2, it was concluded that performance is improved by increasing the degree to which prior information is available to the subject during learning, provided sufficient time is allowed for inspection and inference. In these studies, however, there were two major procedural differences from Laughlin's experiment. Firstly, most of them employed the reception paradigm, whereas Laughlin used the selection procedure for presenting instances. Whether this is likely to affect the influence of the availability of previous instances on performance is uncertain. However, one experiment by Bourne, Ekstrand, and Montgomery (1969) also used the selection procedure and their results supported the conclusion made in Chapter 2. In their experiment, however, no initial instance was presented, hypotheses were required after each five selections, and stimuli were randomly arranged, all of which have been suggested to increase the incidence of scanning. Thus, it is likely that scanning was more frequently used in Bourne et al. than in Laughlin where it has been suggested that most subjects employed a focussing strategy. This difference may account for why Bourne et al. found a memory aid facilitated performance whereas Laughlin did not.

Secondly, in the studies reviewed the form of availability of past information was clearly proscribed; in Laughlin's study the subject had to develop his own method of storing information in an accessible form. Moreover, whereas the actual stimuli were retained in previous studies, Laughlin's subjects presumably translated the stimuli into some representational symbol. This suggests that retrieval of information from a memory aid such as pencil and paper is likely to be far more difficult and therefore less beneficial to performance.

An experiment was conducted to explore further these questions raised by Laughlin's experiment. Data from Experiment I was used for the control group. It was desired to attempt to replicate Laughlin's finding of no improvement with pen and paper and to relate this finding, if confirmed, to the ways in which the memory aid is used. Such information may also indirectly suggest the strategies which subjects are utilizing.

It was also desired to investigate the relationship between memory load and conceptual rule difficulty. Studies by Denny (1969a) suggest that recognition-memory error for previously presented stimuli within a concept problem is greater for rules of greater difficulty. When the memory load for these past stimuli was reduced, the difference in difficulty between rules was also attenuated (Denny, 1969b). A similar result was reported by Shepard, Hovland, and Jenkins (1961) who presented subjects with the task of remembering the category to which each of eight stimulus cards were assigned for six different types of concepts. They found that ranking of concept types by the accuracy and speed with which the assignment of stimuli was memorized yielded the same order as was obtained when subjects had to learn classifications for the various types. Both Denny and Shepard et al. employed the reception paradigm. On the other hand, neither Laughlin (1968b) using the selection paradigm, nor Laughlin (1969b) using yoked selection and reception paradigms were able to find any interaction effect between rule and memory aid. In the experiment to be described four conceptual rules were examined and an attempt was made to see whether any interaction effect between rule and memory aid was obtained which might be in agreement with the results of Denny and Shepard et al..

Experiment IV: Usage of pen and paper as a memory aid in concept identification using a selection procedure.

1. Method

a. Design and Subjects. - The subjects were 52 undergraduate students from the first-year psychology subject pool. A 4 X 5 repeated-measures factorial design was used with the variables (a) conceptual rule (conjunction, inclusive disjunction, conditional, and biconditional) and (b) problems (five for each subject). Thirteen subjects were assigned to each of the rules at random with the limitation that there be approximately equal numbers of each sex on each rule. The attribute values for the five problems were randomly drawn from the 54 possible pairs of attribute value combinations of two attributes. The initial positive instance on each problem was randomly selected from the subset of positive instances having both relevant attribute values.

b. Stimuli. - The same stimulus display was used as in Experiment I.

c. Procedure and Instructions. - These were also exactly the same as in Experiment I. In addition subjects were provided with a pen and a foolscap note pad which they were told they could use as they pleased in order to find the concept. No instructions were given on how they might utilize the pen and paper.

2. Results and Discussion

The results are summarised in Table 12 in terms of the mean and standard deviation of the number of card choices to solution for each conceptual rule and problem.

Conceptual Rule		Problem				
		1	2	3	4	5
Conjunction	Mean	5.5	5.2	3.9	5.5	4.0
	S.D.	2.2	2.9	1.5	2.2	2.1
Inclusive disjunction	Mean	24.5	30.8	20.8	14.5	12.7
	S.D.	11.0	23.7	16.4	7.3	7.5
Conditional	Mean	28.2	30.8	22.7	20.8	21.0
	S.D.	25.6	20.2	14.6	13.3	16.4
Biconditional	Mean	24.4	22.5	11.4	15.9	13.0
	S.D.	18.6	21.9	8.6	8.7	10.8

TABLE 12. Number of card choices to solution when pen and paper are provided.

Using the data from Table 1 as a control group (no pen and paper), an analysis of variance was carried out on this data transformed by a square-root transformation so as to make the variances more homogeneous. Significant effects were obtained for conceptual rules, $\underline{F}$ (3,96) = 25.35, $\underline{p} < .001$, and for successive problems, $\underline{F}$ (4,384) = 14.11, $\underline{p} < .001$. No significant difference was obtained between the control group and the group provided with pen and paper, nor were there any significant interaction effects. A summary of this analysis may be found in Table 13.

Source of variation	S.S.	D.F.	M.S.	F	
Between subjects	890.02	103			
Rules	388.38	3	129.46	25.35	.001
Memory aid	4.67	1	4.67	1	N.S.
Rules X Memory aid	6.76	3	2.25	1	N.S.
Subjects within groups	490.21	96	5.11		
Within subjects	583.62	416			
Problems	69.99	4	17.50	14.11	.001
Rules X Problems	17.62	12	1.47	1.18	N.S.
Memory aid X Problems	4.11	4	1.03	1	N.S.
Rules X Memory aid X Problems	15.82	12	1.32	1.06	N.S.
Problems X Subjects within groups	476.08	384	1.24		

TABLE 13. Summary of analysis of variance.

A post hoc analysis using Newman-Keuls multiple range comparisons revealed for rules, significant differences between conjunction and other rules ($p < .01$), between biconditional and conditional ($p < .01$), and between inclusive disjunction and conditional ($p < .05$). Thus, the order of difficulty of rules is conjunction, biconditional, inclusive disjunction, and conditional from easiest to most difficult. Post hoc comparisons for problems showed Problems 1 and 2 required significantly more trials to solution than other problems ($p < .01$).

These results support Laughlin's (1968b) finding that there is no facilitatory effect on concept identification of providing subjects with pen and paper during the task. An examination was therefore made of the various methods in which the pen and paper was used by subjects. A summary of these methods may be found in Table 14. It can be seen for conjunction that the most frequently employed method was to disregard the pen and paper altogether. For these subjects it is not surprising that there is no improvement in performance on the memory aid condition. Other subjects showed clear signs of using a focussing strategy by listing only the attribute values in the given positive instance. Again it may be expected that these subjects should show little improvement with pen and paper available. For inclusive disjunction, conditional, and biconditional, the majority of subjects elected to make a list of the attribute values of each card chosen and whether it was positive or negative. Short-hand symbols were typically used to represent attribute values. Thus, "R" represented red, "□" represented square, and so on. Sometimes, the list of instances selected was divided into two parts - a positive category and a negative category. On other occasions, positive and negative instances were listed together under the one heading.

Methods of use of pen and paper	Conceptual Rule			
	Conjunction	Incl. disjunction	Conditional	Bicon-ditional
All positive and negative instances listed	2	7	8	8
Positive instances only listed	1	2	0	1
Instances recorded using facsimile matrix	1	0	1	1
Attributes under test listed	3	2	3	2
Possible hypotheses listed	0	0	1	1
No use	5	2	0	0
Unclassified	1	0	0	0

TABLE 14. Classification of methods of use of pen and paper and the frequency of use of each method for four conceptual rules.

It is clear that the majority of subjects attempted to store all the instances selected. They were allowed as much time as they required to inspect and make inferences from these available instances. Yet the availability of past instances did not facilitate performance. This suggests a further limitation on the conclusion made in Chapter 2 on the effects of providing subjects access to previous stimuli encountered. The precise nature of this limitation cannot be specified from the present experiment, but it would seem most likely to be concerned with the problem of retrieval of information from the storage display. Two possibilities exist. Firstly, the problem in retrieval may be due to the fact that the subjects must translate symbolically the attribute values of the cards selected. It may be that whenever subjects must represent instances selected symbolically, the effect of availability of these instances thus transformed is less than if the actual cards themselves were retained. Alternatively, it may be that the storage of instances was not so orderly or well organised as in other studies and it was this factor which hindered retrieval. Nevertheless, it has already been mentioned that several subjects listed positive and negative instances into two groups. Moreover, three subjects in the present experiment developed a method of storing instances which was clearly systematic. These subjects drew on their paper a 9 x 9 matrix similar to the stimulus array, and then proceeded to indicate whether an instance selected from the stimulus array was positive or negative by putting a mark in the corresponding cell of the matrix array. If the degree of orderliness of stored information is the principal factor hindering retrieval of information about the concept, then this method using a facsimile array to represent stimuli should lead to some improvement in performance. This should be the more so because only certain spatial configurations are possible for each type of concept for a systematically arranged stimulus display, so that inspection of the instances stored using the same spatial features as the display should enable subjects to perceive more readily the relationship between the attribute values of positive and/or negative instances.

An experiment was conducted to test this second account of why the provision of pen and paper did not facilitate concept identification. In this experiment subjects were provided with pen and paper ruled up in the form of a 9 x 9 matrix. It was explained to them that they were to regard this matrix as representing the stimulus display in front of them and that when an instance was selected they should place a mark in the appropriate cell of the matrix indicating whether it was a positive or a negative instance. Apart from this modification the procedure was identical to Experiment I, which formed the control condition as in the previous experiment. If the performance of the experimental group is superior to the control group, then clearly the orderliness of the storage display is a significant factor influencing the effect of the availability of past instances on concept identification. If not, then the first alternative discussed above would seem to be principally affecting retrieval. It should be noted that the form of storage display recommended in the experiment to be described still requires symbolic translation, in that instances are represented here as cells in a matrix.

Experiment V: The orderliness of the storage display as a factor affecting concept identification

1. Method.

a. Design and Subjects. - The subjects were 52 naive undergraduate students from the first-year psychology subject pool. The design was identical to that used in Experiment IV.

b. Stimuli. - The same stimulus display was used as in Experiment I.

c. Procedure and Instruction. - These were also exactly the same as in Experiment I. In addition subjects were provided with a pen and a quarto note pad in which each page was drawn up into a 9 x 9 matrix array. It was pointed out to the subject that this array could be regarded as a facsimile of the stimulus display in that each cell in the matrix corresponded to a card in the display. The subject was instructed to label his facsimile matrix and to indicate in the appropriate cell "yes" or "no" each time an instance was selected.

2. Results and Discussion

The results are summarized in Table 15 in terms of the mean and standard deviation of the number of card choices to solution for each conceptual rule and problem.

Using the data from Table 1 as a control group, an analysis of variance was carried out on this data transformed by a square-root transformation. Significant effects were obtained for conceptual rules, F (3,96) = 19.91, $p < .001$, and for successive problems, F (4,384) = 19.12, $p < .001$. The only other significant finding was the interaction between rules and problems, F (12,384) = 2.11, $p < .05$. A summary of this analysis may be seen in Table 16.

These results indicate that performance has not been improved by the provision of a storage display which is an analogue of the stimulus display. It would appear therefore that the major factor preventing facilitation of performance as a result of the availability of past instances in this and the previous experiment and in Laughlin (1968b) is that the instances have had to be recoded using a set of symbols to represent attribute values which are different from those in the stimulus display.

Conceptual Rule		Problem				
		1	2	3	4	5
Conjunction	Mean	9.0	4.8	5.1	5.2	5.1
	S.D.	5.4	2.6	2.0	2.2	1.8
Inclusive Disjunction	Mean	23.8	14.9	12.5	13.1	8.8
	S.D.	10.5	9.7	5.9	5.4	2.8
Conditional	Mean	20.0	15.2	14.8	19.3	10.5
	S.D.	7.0	6.6	9.4	12.3	5.3
Biconditional	Mean	17.3	18.1	10.9	18.4	13.9
	S.D.	12.5	12.1	5.5	16.7	13.1

TABLE 15. Number of card choices to solution when pen and matrix paper are provided.

Source of variation	S.S.	D.F.	M.S.	F	P
Between subjects	647.74	103			
Rules	243.41	3	81.14	19.91	.001
Memory aid	2.62	1	2.62	1	N.S.
Rules X Memory Aid	10.57	3	3.52	1	N.S.
Subjects within groups	391.14	96	4.07		
Within subjects	430.52	416			
Problems	66.26	4	16.57	19.12	.001
Problems X Rules	21.92	12	1.83	2.11	.05
Problems X Memory aid	3.07	4	0.77	1	N.S.
Problems X Rules X Memory aid	6.53	12	0.54	1	N.S.
Problems X Subjects within groups	332.74	384	0.87		

TABLE 16. Summary of analysis of variance.

Of course, it has still not been shown that the failure to obtain any facilitating effect is not due to the use of a focussing strategy in these experiments. The task conditions in these experiments have been such as to maximize the use of focussing, and it has been argued that memory aids have little or no effect on performance when subjects are utilizing this type of strategy. A test of this hypothesis could be made by arranging the task conditions so as to increase the likelihood of scanning strategies instead of focussing. Thus, the stimulus display could be randomly arranged, hypotheses required on each trial, and the initial positive instance eliminated or replaced by an initial hypothesis. Under these conditions control and experimental groups could be compared in the same way as in Experiment IV.

The finding of no significant interaction between memory aid and rules in this experiment and in Experiment IV fails to confirm the result reported by Denny (1969a, 1969b), but is in agreement with Laughlin (1968b, 1969b). Several reasons may be offered to account for this discrepancy. Firstly, the failure to find a significant Rule X Memory aid interaction may be due to the factors already referred to above in accounting for the lack of improvement in performance with pen and paper. In addition, it may be noted that Denny employed a reception procedure whereas the present experiments and those of Laughlin have utilized the selection paradigm. The selection procedure was also used by Bourne et al. (1969) who reported a statistically marginal Rule X Availability interaction, $\underline{p} < .10$. These workers considered four levels of availability - none, all positive, all negative, and all instances - so that their interaction effect may well be due to the type of instances available. It should also be noted that Denny's results pertained only to affirmation, inclusive disjunction, and exclusive disjunction for recognition-memory error, while the attenuation in the differences between rules was demonstrated only for the last two of these rules mentioned. The generality of his finding for other rules may therefore also be in question. Further work is necessary to account for Denny's findings.

CHAPTER 5

HYPOTHESIS TESTING

Within concept attainment two major types of processes may be distinguished - component-centred processes and concept-centred processes (Eifermann, 1965b). In the previous chapter most of the discussion was concerned with component-centred or focussing approaches. This was mainly because the conditions of the tasks examined were such as to elicit in most cases this type of strategy. Equally task conditions may be manipulated to give rise to predominantly concept-centred or scanning strategies. For example, Bruner, Goodrow, and Austin (1956) have suggested that scanning is more evident when the stimulus array is randomly arranged. Also, in the previous chapter, it was suggested that focussing is less common when no initial positive instance is presented. These modifications to the procedure were employed in some of the experiments to be described in this chapter in an attempt to learn more about concept-centred behaviour. For such behaviour two aspects may be discerned: (1) the sampling of hypotheses, and (2) the evaluation of each hypothesis sampled.

A. Hypothesis Sampling

Present interest in the selection of hypotheses comes from a model originally proposed by Restle (1962). In this model it is assumed that each problem defines a population of hypotheses for the subject. The subject samples one of these hypotheses at random and then makes his response by predicting on the basis of this hypothesis whether a particular instance is positive or negative. If his response is correct he continues to use the hypothesis on the next trial; if his response is incorrect, then he returns the hypothesis to the pool of hypotheses and randomly resamples.

Levine (1969) has pointed out that Restle's model may be divided into two parts: (1) an assertion about how hypotheses are sampled and that the hypothesis determines the choice response, and (2) statements about the effects of feedback upon retaining and rejecting hypotheses. It is possible to subscribe to the first, or, as Levine calls it, basic part of Restle's theory, without necessarily subscribing to the second part on the effects of feedback. A survey of experimental studies which have examined the various assumptions made by Restle indicates that most of them have concerned themselves with the second part of the theory. This is not to say that some of the assumptions in the first part cannot be challenged or are beyond question. In the discussion to follow an attempt will be made to review those studies which have examined Restle's assumptions.

Examination of assumptions in Restle's model

1. Random sampling

Studies in a number of areas of psychology have considered the question of the "effective stimulus" and these suggest that, in many experimental situations, the subject does not respond to all components of the stimulus situation (Kintsch, 1970). Relating this to the sampling of hypotheses, it would seem likely that the set of hypotheses from which the subject is sampling is actually a subset of the total number possible. Moreover, it is likely that this subset is not merely a random list of hypotheses, but rather a group of hypotheses having certain characteristics in common; for example, they all may involve the same stimulus dimension. Which stimulus dimension is involved may depend on the salience or attention value of the different stimulus dimensions. Trabasso and Bower (1968) have discussed the effects of weighting stimuli according to their attention value on the likelihood of sampling different hypotheses. However, even when the attention values of the various dimensions are the same, a subset of hypotheses is likely to be used with certain features in common.

For the first trial of a problem, nonetheless, to the experimenter the a priori probability that any particular hypothesis is sampled is the same as for any other hypothesis, unless he has information on the size and nature of the subset beforehand. However, when the probability of new hypotheses being sampled on subsequent trials is considered, a better than chance prediction should be able to be made on the hypothesis selected. According to Restle's model, hypotheses are selected from the pool at random, so that the choice of an hypothesis on a particular trial is independent of previous choices. The concept of sampling proposed here, however, suggests that hypothesis sampling is not random and that successive hypotheses may well possess features in common.

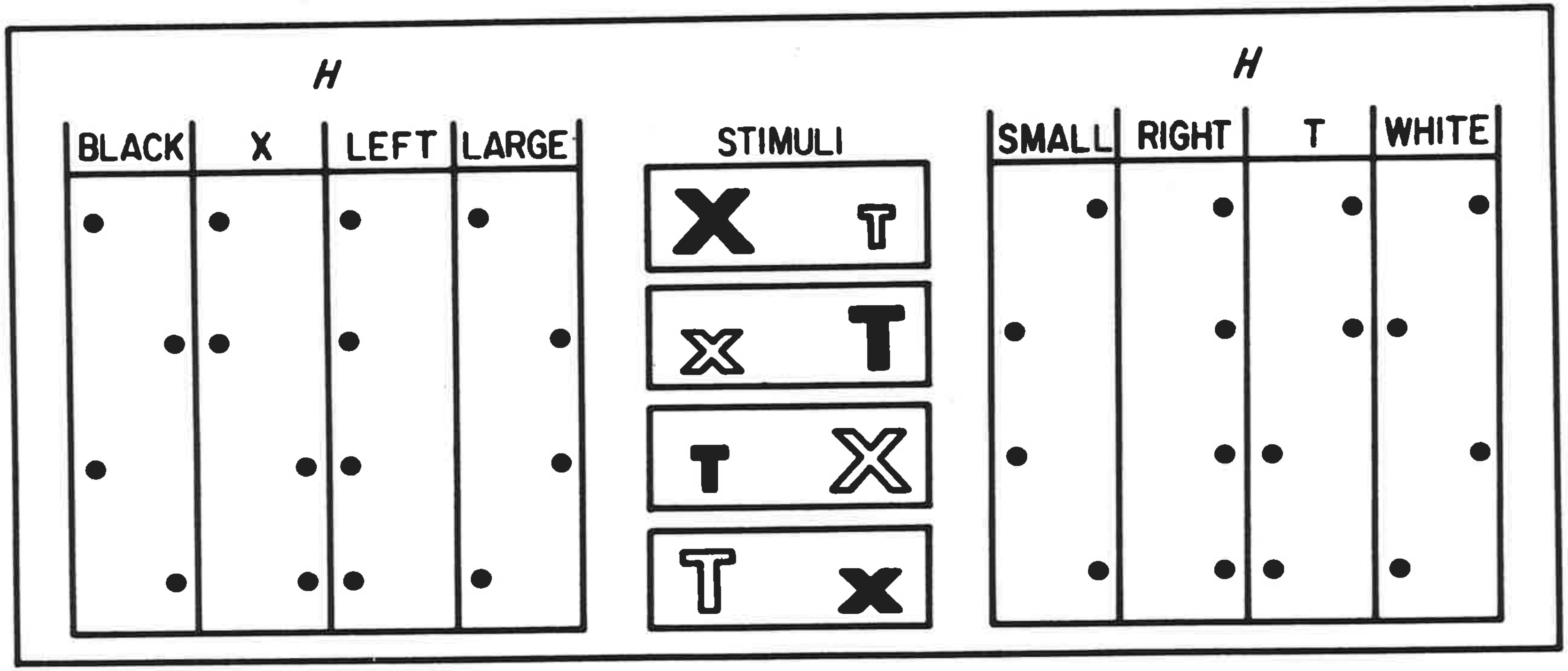
Indeed, any suggestion of random selection must surely be, at best, an approximation in view of considerable evidence that subjects seldom produce a random sequence of events and that consistent biases occur even when subjects are instructed to be random (Chapanis, 1953; Ture, 1964; Weiss, 1964; Baddeley, 1966). Moreover, the extent to which a series of selections is non-random has been found to be a function of such variables as the number of items from which selections are made, the speed at which selections are made, and the age of the person making the selections.

A pilot study by the present writer in which subjects were set the task of identifying conjunctive concepts using a selection procedure where they were required to state their hypothesis after each trial, seems to support the suggestion that sampling is not random. Inspection of subject protocols indicated that, for some subjects at least, when a new hypothesis was proposed, it frequently had some characteristic in common with the previous hypothesis tested. Often the new hypothesis differed from the previous hypothesis by only one attribute value. Thus, for one subject a proportion of 0.6 of all successive hypothesis-pairs showed a difference in only one attribute value. The probability of this event occurring if hypotheses are randomly selected from the total universe of conjunctive hypotheses possible in this experiment is 0.3. Insufficient subjects were used to reliably demonstrate that the selection of an hypothesis in this task was not independent of previous hypotheses, but the results tentatively suggest that Restle's random sampling assumption does not hold for some subjects on conjunctive concept identification. Similar suggestions have been made by Guy, Van Fleet, and Bourne (1966) for single-cue hypothesis sampling and Deffenbacher (1970) for conjunctive hypothesis sampling. On the other hand, evidence supporting a random sampling assumption has been recently provided for conjunctive concept identification by Nahinsky and Slaymaker (1969), and Nahinsky, Penrod, and Slaymaker (1970).

2. <u>The hypothesis and the choice response</u>

Levine (1966, 1969) has examined the assumption that the hypothesis determines the choice response made on each trial, by means of a technique developed for inferring the particular hypothesis held by the subject from his responses to a series of consecutive trials ("blank" trials) on which no feedback is given (Levine, 1963). A description of the method of inference

FIGURE 3. Eight patterns of choices corresponding to each of the eight hypotheses for a set of four stimuli (from Levine, 1966).


H
BLACK
X
LEFT
LARGE
STIMULI
H
SMALL
RIGHT
T
WHITE

used is appropriate.

Consider a subject given the task of identifying the relevant attribute value from a stimulus universe having four attributes with two values per attribute. At the outset of a trial the subject selects an hypothesis from a set of hypotheses which is defined by the stimulus display, that is, a set of eight hypotheses. If no feedback is given to the subject on his choice on this trial, then it is assumed that the hypothesis is retained for the next trial. Evidence to support this assumption has been provided by Levine, Leitenberg, and Richter (1964) who showed that no feedback has the same effect on response probabilities as positive feedback. For a stimulus universe such as considered here, it is possible to present four consecutive stimuli to which the choice response patterns yielded by each of the eight hypotheses is unique. An illustration displaying the eight hypotheses for a set of stimuli like this and used by Levine (1966) is shown in Figure 3. Now, if these four stimuli or another set of four stimuli possessing the same property are presented to the subject without feedback, then the particular hypothesis being utilized by the subject may be inferred from the pattern of choice responses produced. Thus, Levine's technique involves a sequence of trials in which each feedback trial is separated by four blank trials.

Using this technique, Levine (1966) was able to infer the hypothesis sampled on 92.4% of the four trial sets. From the inferred hypothesis, the choice response on the next feedback trial was predicted and Levine found that this prediction was correct on 97.5% of the feedback trials following blank trials where hypotheses could be inferred. It seems clear that on the majority of occasions the subject's behaviour according to this method of inference was systematic, with the response made on a particular trial depending on the hypothesis under test. Nonetheless, it was not possible to determine the hypothesis for 7.6% of the four trial sets, while the predicted response was incorrect 2.5% of the time. This result may be attributed either to the subject changing his hypothesis following a blank trial (which is a violation of the blank-trials assumption) or to the subject "accidentally" making the wrong choice response according to his hypothesis. Levine has suggested that it was the latter and referred to this as an "oops-error". The probability of an oops-error on any trial for Levine's (1966) data was .02. A similar probability has been obtained by Levine (1969) using an eight-dimensional problem, although it was apparent that the probability of an oops-error decreased during a problem. It would appear then that Restle's assumption that the hypothesis determines the choice response is true for all except a very small proportion of trials for which a small correction term can be estimated.

The high degree of consistency in Levine's results is important not only to Restle's theory but also to his own method of inferring hypotheses since it is crucial to this method that the hypothesis held by the subject should determine his response. Where oops-errors do occur during blank trials, two consequences can happen: either a response pattern will be formed from which a different hypothesis is inferred from the one currently held or a response pattern will be formed from which no hypothesis can be inferred. The probabilities of each of these occurrences is once again small and able to be estimated. Further evidence favouring the blank-trials assumption and Levine's technique for inferring hypotheses has been provided by Frankel, Levine, and Kampf (1970). They presented subjects with problems in each of which was a series of 30 consecutive blank trials. It was found that subjects responded in a manner consistent with only one stimulus value on 25 or more blank trials for 80% of the problems.

3. Learning Curve

According to Restle, the subject may sample either incorrect hypotheses or the correct hypothesis. If an incorrect hypothesis is sampled, then the probability that a correct response is made is 0.5 in a two-category task. If the correct hypothesis is sampled, then no further error responses occur. The theory thus implies an all-or-none process which gives rise to a learning curve such that prior to the last error the probability of a correct response is stationary at 0.5. This implication differs from the incremental assumption of earlier conditioning theories of concept identification (Bourne and Restle, 1959), which predicted that the probability of a correct response increases until the final error is made.

For response probabilities, evidence supporting the all-or-none assumption has been found by Bower and Trabasso (1963, 1964), Erickson, Zajkowski, and Ehmann (1966), and Guy et al. (1966). Levine, Miller and Steinmeyer (1967), Levine, Yoder, Kleinberg, and Rosenberg (1968), and Levine (1969) have also confirmed the all-or-none nature of the learning curve, but unlike the other studies their results indicated a probability of a correct response which was significantly below 0.5 prior to solution. Levine and his co-workers have argued that this impairment in presolution performance is compatible with hypothesis-sampling theory, and is attributable to the internally orthogonal stimulus sequences employed (i.e. no stimulus dimension was correlated with any other dimension for two or three trials). According to Levine, earlier studies did not obtain a similar impairment because they used randomized stimulus sequences in which irrelevant dimensions were correlated with the relevant dimension for runs of varying length. Thus, the likelihood that a subject holding an incorrect hypothesis may produce correct responses for several consecutive trials is increased for these sequences compared with the orthogonal stimulus sequences used by Levine. A detailed explanation may be found in Levine et al. (1968).

Latency data casts a different light on the all-or-none versus incremental process controversy to the choice data. An experiment by Erickson et al. (1966) showed that not only do latencies decline during the presolution period, but also during the criterion run following the last error when the subject is supposed to be holding only the correct hypothesis. A similar result has been obtained by Falmagne (1970). It appears therefore that although choice responses show a sudden change in strength, latencies show a gradual decrease as might be expected for a continuous conditioning or incremental process.

Two attempts have been made to account for this conflicting evidence. Trabasso and Bower (1968) propose that on each trial the subject attends to a subset of the possible hypotheses (called the focus sample), instead of only one hypothesis. The response probabilities depend only on the hypotheses in the focus, and only these hypotheses are affected by the outcome of the trial. An erroneous response initiates resampling of a new focus of fixed size $\underline{N}$. After a correct response, the hypotheses which are incompatible with that response are eliminated from the focus, without replacement, i.e. information processing does take place after a correct response. The size of the focus thus decreases on the average during a sequence of correct responses. On the assumption that latency is a function of the number of hypotheses in the focus, Trabasso and Bower concluded that latency would decrease over the first few trials of a criterion run. It is apparent that the modification made to the original theory requires the subject to process considerably more information on any trial. Discussion on the means by which subjects may evaluate hypotheses will be introduced later in this chapter. For the present, however, it can be suggested that, while such information processing requirements may be feasible for simple concepts and a small stimulus universe, doubts must be expressed about its use on more complex problems, e.g., when subjects are required to solve several concept identification problems concurrently. It is not clear therefore how Trabasso and Bower would account for Erickson and Zajkowski's (1967) finding of a decline in latencies in the postcriterion period for concurrent concept identification.

Falmagne (1970) has proposed an alternative account in which an incremental process is posited. In her model every hypothesis is assumed to have associated with it a "strength" or subjective plausibility which is modified by the feedback provided on each trial. A similar objection may be raised to Falmagne as to Trabasso and Bower, that the amount of information processing demanded by her assumption is not feasible except for very simple problems. The dilemma facing these two models is the consequence of their conception of the process of evaluation of hypotheses as a process of accepting one hypothesis by rejecting others. An alternative method of testing an hypothesis will be proposed later which resolves this dilemma.

4. Effect of feedback

a. Positive feedback

The assumption that hypotheses are changed only after error responses has received attention in several studies, and unequivocal evidence has been obtained that subjects sometimes do change their hypotheses after correct responses, although estimates of the probability that a subject will revise his hypothesis after a correct response vary. The range of the estimates appears to be a function of the type of procedure employed.

Dodd and Bourne (1969) have estimated the probability value, $\underline{P}_{\underline{r}}$, to be .03 in an experiment in which subjects' responses to stimuli varying on five binary dimensions were randomly reinforced. These authors also re-analysed data collected by Trabasso and Bower (1966) using a procedure involving multiple shifts in the relevance of two stimulus dimensions within a set of five binary stimulus dimensions during the course of learning. The value of $\underline{P}_{\underline{r}}$ in this experiment was estimated to be approximately .21. The difference obtained between the two estimates led Dodd and Bourne to suggest that $\underline{P}_{\underline{r}}$ may not be constant across consecutively correct responses.

Suppes and Schlag-Rey (1965) found that the proportion of correct trials on which the hypothesis was changed was 0.20 using a string of three letters, each letter being either D or K, as stimuli, but when the stimuli consisted of strings of 1, 2, 3, 4, and 5 letters, the proportion was much higher, ranging between 0.48 and 0.67. Suppes and Schlag-Rey suggest that the high values of $\underline{P}_{\underline{r}}$ obtained in their experiments may be accounted for by the fact that subjects were constantly informed of all past reinforcements.

Levine has also found that subjects change their hypothesis after a correct response, but the probability estimated in his studies is much lower, of the order of .05 (Levine, 1966) for a four-dimensional problem or .06 (Levine, 1969) for an eight-dimensional problem. Dodd and Bourne (1969) have suggested that the low value of $\underline{P}_{\underline{r}}$ may be due to the high degree to which the problem and the responses of the subject were structured by pretraining. If so, this has a bearing not only on Restle's assumption about the effects of positive feedback, but also Levine's blank-trials assumption in which no feedback is supposed to have the same effect as positive feedback, an assumption which forms the basis of his technique for inferring hypotheses. It should be noted that it is sometimes possible on Levine's procedure to infer an hypothesis from a sequence of responses on blank trials, even though the subject is responding randomly.

Obviously further work is needed to show systematically how $\underline{P}_{\underline{r}}$ is affected by variation in the task procedure used. At any rate, it seems that Restle's assumption that an hypothesis is not changed after a correct response has been made is not strictly correct. Where $\underline{P}_{\underline{r}}$ is only very small, changes in the hypothesis may be accounted for in a similar fashion to Levine's notion of oops-errors. For larger values of $\underline{P}_{\underline{r}}$, however, it seems that perhaps the subject may be simultaneously sampling a subset of hypotheses some of which are equivalent and interchangeable following a correct response.

b. <u>Negative feedback</u>

A few studies have also examined the analogous probability that the hypothesis will remain unchanged when an error response has been made. Using the method of inferring the subject's hypothesis through his responses to a series of blank trials, Levine (1966) has obtained an overall probability that the hypothesis is unchanged after an incorrect response of .02 for a four-dimensional problem. The fact that the hypothesis is unchanged after an error need not cause any embarrassment to Restle's theory since he postulates that the old hypothesis is returned to the pool of hypotheses before resampling takes place.

Thus, the probability that the response should be unchanged in this case according to Restle is .125. It can be seen that the mean probability obtained by Levine (1966) is less than that theoretically expected, suggesting that Restle's replacement assumption is not correct. Levine further demonstrated that if negative feedback is administered after both the first and second responses, the third hypothesis tends to be different not only from the second hypothesis, but also the first hypothesis as well.

Levine (1969) replicated this result using an eight-dimensional problem. In this experiment for a replacement assumption the theoretical probability that an hypothesis will be selected again after an error is .0625, but the probability obtained by Levine was .01 on the first trial, and even after five intervening feedback trials and 20 blank trials the probability was below .03. It would appear then that Restle's assumption concerning the effect of an error is incorrect, that replacement does not occur before resampling.

5. Retention of information

In addition to the work of Levine reported above, several other studies have also raised questions about the validity of the sampling-with-replacement assumption. Levine (1962) and Holstein and Premack (1965) showed that a block of random reinforcement trials followed by consistent reinforcement produced a decrement in the solution of problems involving two-dimensional and six-dimensional stimuli. Trabasso and

Bower (1966) found that shifts in the relevance of two dimensions prior to solution retarded learning. These findings suggest that subjects do retain something of their past reinforcement history which is contrary to the idea that there is no memory of past information implied by sampling-with-replacement. Trabasso and Bower (1964b) and Bourne and O'Banion (1969) have also demonstrated some retention of information, allbeit small, during concept identification. Recall of stimuli and the associated feedback was best for the first instance (primacy effect) and the last instance (recency effect) presented. A review of other studies which have indicated an important role of memory in concept identification may be found in Dominowski (1965). Erickson and Zajkowski (1967) compared subjects' performance when three concepts were identified concurrently with performance when these same concepts were learned successively. From their results, Erickson and Zajkowski concluded that the concurrent procedure led to interference in the short-term storage of hypotheses tried but disconfirmed, such that data obtained using this procedure conformed more closely with the predictions of a sampling-with-replacement assumption.

If the sampling-with-replacement assumption is to be revised so as to take into account the retention of information, there are at least two ways to proceed. One approach is to suppose that the subject remembers some previous instances encountered. An alternative approach is to suppose that he remembers some hypotheses that he has tested and disconfirmed.

Using the former approach, it may be assumed that the subject samples only those hypotheses which are consistent with the last instance encountered. This has been referred to as a local-consistency assumption (Gregg and Simon, 1967). Equally, it may be postulated that the hypothesis sampled is consistent with the last two instance presented - referred to by Erickson (1968) as local consistency plus memory for one stimulus back. It can be readily seen that these assumptions concerning memory for past stimuli may in fact be extended until the new hypothesis selected is consistent with all the instances which have been presented to him (global consistency).

Using the latter approach, it may be assumed that the subject samples from all hypotheses except the one held on the last trial. This has been referred to as a local-nonreplacement assumption (Gregg and Simon, 1967). In the same way as was shown with the local-consistency assumption and memory for past stimuli, the local-nonreplacement assumption can be extended to allow for more than one hypothesis. Thus a two-hypothesis memory assumption would postulate that the new hypothesis sampled is drawn from the pool of all hypotheses except the last two hypotheses tested.

A combination of the two approaches is also possible. From these assumptions, quantitative predictions may be derived for the probability of a true hypothesis over a series of trials and these predictions can be compared with empirical evidence to see which of the assumptions provides the best fit. For stimuli having four binary dimensions, Erickson (1968) observed that local-consistency models fit the data fairly well, with the best fit by the local-consistency-with-two-hypothesis-memory assumption.

Levine (1967) has suggested a technique for estimating the size of the hypothesis pool for any given trial. Essentially this estimate is derived from the reciprocal of the proportion of subjects starting the criterion run at that trial. Using this technique on data collected by Levine et al. (1967) for a four-dimensional problem, the size of the hypothesis set was found to decrease for the first three trials to about three hypotheses, thereafter oscillating around three on succeeding trials. This result can be compared with the predictions made by the various memory assumptions. Once again, a local-consistency-with-two-hypothesis-memory assumption seems to fit the data fairly well. On this assumption, the number of hypotheses in the set would be 8 and 4 on the first two trials, and either 4 or 3 on the succeeding trials. Also providing a reasonable fit is a local-consistency-with-memory-for-one-stimulus-back assumption. On this assumption, the number of hypotheses in the set would be 8 and 4 on the first two trials, with a mean of 2.25 on the succeeding trials.

Levine (1969) has estimated the size of the hypothesis set for an eight-dimensional problem. He found that the number of hypotheses decreased for the first three trials to slightly more than five hypotheses. On later trials the number of hypotheses appeared to reach an asymptote at about this level and then to increase slightly. This last observation can also be seen in Levine's (1967) estimate although to a lesser extent. Comparisons with the various memory assumptions this time reveals that the best fit is provided by a local-consistency-with-memory-for-one-stimulus-back assumption. For this assumption the number of hypotheses in the set is 16 and 8 on the first two trials with a mean of 5.1 hypotheses on succeeding trials.

Closer examination of the subject's hypotheses, however, indicates that the proportion of hypotheses consistent with the information from the immediately preceding feedback trial is never equal to 1.0, and indeed decreases over trials, suggesting an intraproblem proactive interference effect. If the proportion of hypotheses consistent with each of the preceding feedback trials is plotted for each trial, it is evident that the subject's hypothesis is never perfectly consistent with any trial, but greatest consistency is apparent for the first trial in the series (i.e., primacy effect) and the trial immediately prior to the hypothesis (i.e., recency effect). This result is reminiscent of the finding of Trabasso and Bower (1964b) and Bourne and O'Banion (1969) of primacy and recency effects in the immediate recall of the stimuli and their associated feedback in concept identification. It therefore appears that a more likely memory assumption is local consistency with memory for the first stimulus presented, although it should be emphasized that even this assumption is at best an approximation only since perfect consistency is not found with any trial.

Extension to multidimensional concepts

Thus far, discussion of the assumptions underlying hypothesis-sampling models has concentrated mainly on the identification of single-attibute concepts. Some attempts have been made recently to extend findings from these studies to concepts having more than one relevant attribute.

An early study by Trabasso and Bower (1964a) provided evidence to support an assumption originally made by Bourne and Restle (1959) that in a bidimensional problem the subject analyses the two relevant dimensions as two independent unidimensional subproblems and that these two subproblems are worked on concurrently. Trabasso and Bower found that, in a four-category concept identification task with two independent relevant dimensions, the performance of subjects could be predicted from the performance of other subjects who were presented with only the unidimensional subproblems.

An unpublished study by Miller (1970) attempted to apply this model of hypothesis-sampling to disjunctive concept identification. In her experiment she presented post-shift problems involving both (intradimensional shift), one (half-shift), or neither (interdimensional shift) of two dimensions relevant in a pre-shift problem. From Trabasso and Bower's model it was predicted that the order of increasing difficulty of post-shift problems would be intradimensional shift, half-shift, and interdimensional shift. The results indicated that the intradimensional shift problems were less difficult than both the half shift and the interdimensional shift problems, but no difference was obtained between the last two types of shifts. This suggested that, instead of treating the bidimensional problem as two separate single-dimension subproblems, subjects must be using some method of sampling pairs of dimensions concurrently. Hence, when one member of a pair is found to be irrelevant, both members of the pair are discarded. Whether pairs of dimensions are sampled randomly is a question which has been raised earlier.

Nahinsky (1968) investigated the identification of conjunctive concepts having two relevant attributes by asking subjects to guess the concept after a fixed number of trials.

He found that a significantly higher proportion of subjects attained the concept without errors than would be expected if no information processing was assumed following a correct response. Furthermore, the data were not compatible with either sampling-with-replacement or local-consistency-with-memory-for-one-stimulus-back assumptions. These findings were corroborated in a later study by Nahinsky and Slaymaker (1969) in which blank trials were used to infer the subject's hypotheses. In this experiment each subject was presented with an initial positive instance followed by a random series of stimuli. Each positive instance was followed by a series of blank trials. The results suggested that subjects sample randomly and independently some subset of possible hypotheses compatible with information provided by the first stimulus and then reduce the number of hypotheses in this subset on the basis of information provided by succeeding stimuli. A model based on these results has been reported by Nahinsky (1970).

Deffenbacher (1970) assessed the fit of models incorporating four of the memory assumptions discussed earlier - zero memory, local nonreplacement, local consistency, and global consistency - to conjunctive concept identification data obtained for four levels of stimulus complexity. In examining each of these assumptions, it was further assumed that subjects were able to recall the initial instance. Analysis of the results, however, indicated that none of the assumptions satisfactorily fitted the data for all four levels of stimulus complexity, although the best approximation was provided by the local-consistency (strictly, local-consistency with memory for the first stimulus presented) assumption. Two reasons were advanced by Deffenbacher as possible explanations why a good fit was not obtained: (1) the subjects may not have retained information from the initial exemplar over trials, and (2) hypotheses may not have been selected with equal probability.

Comment

It is apparent from this review of hypothesis sampling that Restle's original assumptions have been tested and in a number of instances found to require some modification. Where possible in the discussion particular modifications have been considered. However, an integrated account is yet to emerge. One of the difficulties seems to be the paucity of information relating procedural constraints and task conditions to the sampling process. Until more is known of their effect, exceptions to any sampling model are likely to be found.

B. Hypothesis Evaluation

It has been typically assumed in the foregoing discussion that hypotheses are evaluated according to an elimination process in which the concept (i.e., the true hypothesis) is established by the rejection of all alternative hypotheses inconsistent with the stimuli presented. This may be termed an information-processing approach. There is an alternative approach to hypothesis evaluation which should also be considered. This proposes that evaluation of the truth of an hypothesis proceeds by a statistical decision-making process. Regrettably few studies have compared these approaches or examined the assumptions underlying them. In the succeeding pages an argument will be developed to suggest that, except in a few special cases, subjects tend to utilize the latter rather than the former approach. If this is so, then certain implications for some of the phenomena reported during the previous discussion on sampling follow. For example, the observation that latencies decrease during the criterion run of correct responses may be attributed to increasing confidence in the decision on the truth of the hypothesis. There is consequently no need to assume here that subjects are simultaneously monitoring several hypotheses as proposed by Trabasso and Bower (1968).

An information-processing approach

Essentially this approach proposes that the subject arrives at the true hypothesis by eliminating by some means all other possible hypotheses. This assumes therefore that a concept-centred subject possesses information about (1) the population of hypotheses from which the concept is drawn, and (2) the truth value of each hypothesis in the population on any trial. It is plausible in most concept identification experiments to assume that the number of hypotheses is defined by the instructions and stimuli presented, but a number of reasons can be given to suggest that the second assumption is untenable.

In the first place, Bruner et al. (1956) reported that no subject employed a simultaneous scanning strategy, i.e., examined all hypotheses simultaneously, but rather dealt successively with single hypotheses for conjunctive concept attainment from a four-attribute, three-value stimulus display. Secondly, theoretical models of hypothesis sampling (see earlier) have postulated that subjects are able to recall up to only two previous hypotheses during concept identification for unidimensional concepts. It would be interesting to know the maximum number of hypotheses a subject can recall, assuming hypotheses are sampled independently of each other (Nahinsky et al., 1970). This estimate was made in an experiment in which immediate recall of a list of ten random bidimensional hypotheses based on one of four conceptual rules was tested for a stimulus population with four three-value attributes. The estimate will be, of course, an upper limit since there is no testing of the hypotheses required, nor do subjects have to recall the truth value of each hypothesis.

Experiment VI: Immediate recall of random lists of bidimensional hypotheses.

1. Method. - The subjects were 36 undergraduates randomly drawn from the first-year psychology subject pool at the University of Adelaide. These subjects were assigned to one of four groups so that there were 9 subjects in each group. Each group was presented with hypotheses involving a different conceptual rule. The four rules were conjunction, inclusive disjunction, conditional, and biconditional.

In front of the subject was a stimulus display consisting of geometrical designs printed on 7 x 10 cm. white cards. Designs varied according to four attributes, with each attribute having three values, making 81 different patterns. The attributes and their values were colour (red, green, and black), shape (cross, circle, and square), number of figures (1, 2, and 3), and number of borders (0, 1, and 2). The 81 cards were systematically arranged in a 9 x 9 array.

The instructions explained to the subject that he would be presented with a number of hypotheses which would consist of combinations of pairs of attribute values from the stimulus population before him. Each pair of attribute values involved the same form of combination. For conjunction, hypotheses were of the form, "green and square", for inclusive disjunction, "either green or square or both", for conditional, "if green then square," and for biconditional, "green if and only if square". The meaning of the relevant conceptual rule was demonstrated for the example provided using the stimulus display. No indication was provided on how many hypotheses were to be presented. The subjects were told that they should try to remember as many of the hypotheses as they could, and that opportunity would be given them to write down those they could remember when all hypotheses had been read out. Ten hypotheses were read aloud at intervals of 5 seconds. Each of the ten hypotheses was randomly selected from the 54 pairs of attribute-value combinations of two attributes. Subjects were not required to recall the order in which the hypotheses were presented, only the hypotheses themselves.

2. Results and Discussion. - The results are summarized in Table 17 in terms of the mean and standard deviation of the number of hypotheses correctly recalled.

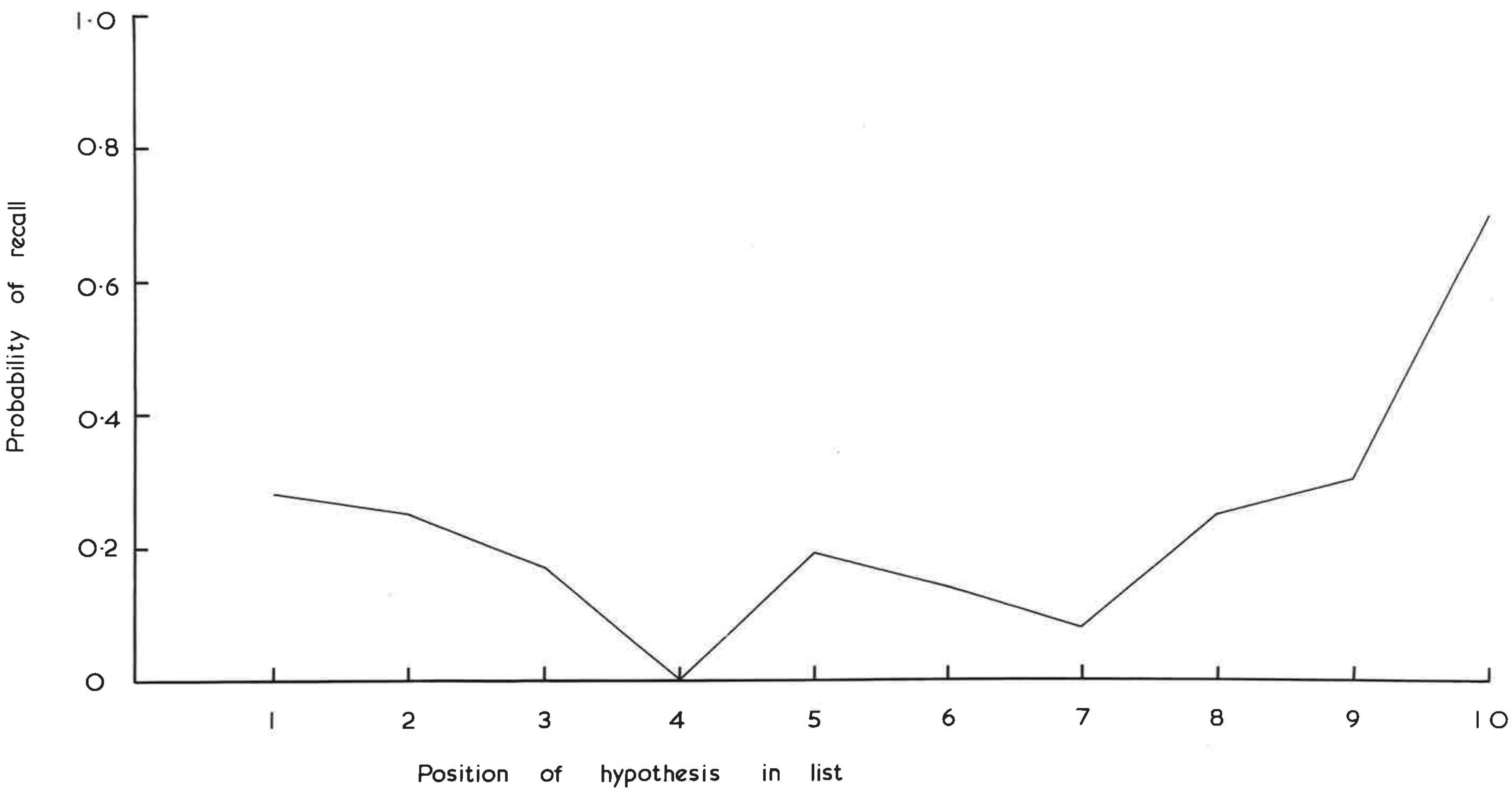
Conceptual Rule	Mean	S.D.
Conjunction	2.2	0.6
Inclusive disjunction	2.1	1.5
Conditional	2.2	0.9
Biconditional	3.0	0.9

TABLE 17. Number of hypotheses recalled for each rule.

An analysis of variance was performed on this data, but no significant difference was obtained between rules. The mean number of hypotheses recalled over all rules was 2.4. A curve showing the probability that an hypothesis was correctly recalled plotted against the position of the hypothesis in the list may be seen in Figure 4. Examination of this curve suggests a marked recency effect in recall, but no primacy effect was obtained.

Errors made by subjects suggest that for bidimensional hypotheses, a high degree of confusability exists between hypotheses. This can be seen in that each attribute value is associated with another attribute value on 9 occasions within a complete list of 54 hypotheses. Moreover, there are 9 hypotheses involving combinations of values taken from the same pair of attributes. The effect of this was shown by the number of times in which one of the two components was incorrectly recalled. Often the error consisted of simply choosing the wrong value within the right dimension. Some subjects reported trying to connect each hypothesis with a card from the stimulus display which contained both relevant attribute values. This method apparently did not improve recall, however. The reason for no improvement is probably due to the fact that recall of cards is not very much better than that for hypotheses, and anyway, should the card be

FIGURE 4. Probability that an hypothesis is correctly recalled plotted against the position of the hypothesis in the presentation list.



Probability of recall
1·0
0·8
0·6
0·4
0·2
0
1
2
3
4
5
6
7
8
9
10
Position of hypothesis in list

recalled correctly, it still has to be decided which of six possible hypotheses was the basis for choosing the card.

It is evident that for this stimulus population, recall of a random list of ten hypotheses was only slightly more than two, a finding which must be regarded as the upper limit of recall for the reasons stated earlier. This result, when considered along with the other evidence cited, strongly suggests that a logical elimination method is not used by concept-centred subjects in evaluating a hypothesis. If subjects are not able simultaneously to test all hypotheses, and if they are not able to recall all previous hypotheses tested, then it is difficult to see in what way they can logically eliminate all hypotheses except the true hypothesis and consequently establish that this remaining hypothesis is true. If subjects were using an elimination strategy, then one would expect them to be able to describe, with more precision than has been commonly found, the approach they used.

Moreover, the finding that only two hypotheses were correctly recalled in this task also casts doubt on the subset-sampling assumption which states that the subject samples and evaluates several hypotheses at the same time (Levine, 1970). Clearly in this task the size of the subset must be only one or two, with one being more likely since subjects here were required only to recall hypotheses, not to evaluate them. It should be admitted, however, that the present task is somewhat more difficult than has generally been considered, involving as it does a larger population of hypotheses with a high degree of confusability between them. Perhaps for simple problems subset sampling is still a possibility. However, other evidence supposedly favouring subset sampling, for example, the decrease in latencies during the criterion run of responding can also be accounted for without this assumption. It would seem therefore that most of the data can be adequately explained by assuming that the subject handles only one hypothesis on each trial. The strongest evidence against the single-hypothesis assumption would seem to be the finding of a high probability that the hypothesis is changed following a correct response. More work is needed to determine the conditions under which this observation occurs.

Considerations such as the above do not apply to the subject who elects to utilize an attribute-testing strategy, for example, conservative focussing. For this strategy the subject effectively eliminates all other hypotheses by examining the relevance of individual attribute values in a focus card. The capacity of the subject to retain information about the concept is not exceeded with this strategy. Indeed, Byers, Davidson, and Rohwer (1968) have shown that the memory load for subjects given instructions on conservative focussing is less than for other subjects not given these instructions. Attribute-testing or focussing strategies may be used on all conceptual rules (see previous chapter), but their use seems confined mostly to special task conditions, e.g. a systematically arranged array of instances from which one instance is given for a focus.

A statistical decision approach

Since it is evident that many subjects do use an hypothesis-testing or concept-centred strategy, and since it is also evident that this strategy cannot be one in which all alternative hypotheses are eliminated, then clearly some other approach is necessary for them to solve the problem. It is suggested that this other approach may be termed a statistical decision-making process.

Essentially the subject given an hypothesis to test is viewed as having to make a decision between two alternatives; (1) the hypothesis is true, or (2) the hypothesis is false. Any instance which confirms the hypothesis may be regarded as influencing this decision in favour of the former alternative. Any instance which disconfirms the hypothesis may be regarded as influencing the decision in the direction of the latter alternative. In fact the criterial number of disconfirming instances for choosing the alternative, the hypothesis is false, is one. The criterial number of confirming instances for choosing the alternative, the hypothesis is true, will vary with the consequences of making this decision.

A model of this sort is not subject to the demands on memory required by a process of logical elimination. Apart from assumptions made concerning the role of memory by hypothesis-sampling models, the only requirement is that the subject remember how many confirming instances of the hypothesis under test have been found. Furthermore, it does not suppose that the subject knows the number and type of hypotheses from which the concept has been chosen. This assumption raises the possibility that some hypotheses may not be eliminated during the process of evaluation. Evidence illustrating this possible eventuality has been provided by Wason (1960, 1968a). The degree of risk involved in deciding that a particular hypothesis is true depends on the number of confirming instances obtained for the hypothesis. The finding that subjects may, on occasions, change their hypothesis after confirming feedback (see earlier) does present some difficulty to this model of hypothesis-evaluation, unless it may be supposed that these occurrences are due to oops-errors.

The success with which an hypothesis is evaluated clearly depends not only on the number of instances selected, but also on the type of instances which are chosen. Bruner et al. (1956) have suggested that subjects prefer a direct test of an hypothesis. The example they give to illustrate the difference between a direct and an indirect test of an hypothesis is to consider a subject faced with deciding whether a white door or a black door is the correct entrance to a reward chamber. If the subject adopts the hypothesis that the white door is correct, then a direct test of this hypothesis is to try the white door; an indirect test is to try the black door. Bruner et al. suggest that subjects are either unwilling or unable to employ an indirect method of testing the hypothesis. In effect, this means that subjects prefer to select instances which would be positive if their hypothesis was true.

In Bruner et al's investigation the apparent preference for direct testing of hypotheses may be related to the fact that for conjunctive concepts a direct test or positive instance transmits more information about the concept than an indirect test or negative instance. Thus it may be that subjects do employ a partial information-processing strategy by choosing instances which convey the maximum amount of information about the truth of their hypothesis; if the hypothesis should be false, on the other hand, little information is likely to be gained about the true hypothesis through a direct test. Evidence from matching problems (Donaldson, 1959) and compound code-item tasks (Campbell, 1965) in which the optimal solution involved the use of negative or indirect information, suggests, however, that subjects prefer the direct test method even when this conveys only a small amount of information towards the solution.

An experiment was performed to investigate some of the characteristics of hypothesis evaluation discussed above. Hypotheses involving eight rules were presented to subjects to test. According to a statistical-decision model the number of instances chosen to evaluate an hypothesis is independent of the rule, being a function only of the criterion adopted for accepting the hypothesis. Thus no difference was expected between rules on this response measure. This is contrary to the prediction of an information-processing model. A simultaneous scanning strategy for a stimulus population with four three-value attributes requires 2 choices for conjunction and disjunctive absence, 3 for inclusive disjunction and conjunctive absence, and conditional and exclusion, and 4 for biconditional and exclusive disjunction in order to determine the truth or falsity of a given hypothesis, assuming that the true hypothesis involves the same rule as the given hypothesis. Secondly, it was also desired to investigate the type of instances chosen to evaluate hypotheses for each rule. From Bruner et al. it may be expected direct test instances are preferred for all rules. If, however, the informativeness of instances is also taken into account by subjects, then on some rules, e.g., conditional and disjunctive absence, indirect test instances may be preferred.

Experiment VII: An investigation of the number and type of instances selected to test an hypothesis.

1. Method

a. Design and Subjects. - The subjects were 24 undergraduates randomly selected from the first year psychology subject pool at the University of Adelaide, with the provision that there be equal numbers of each sex.

A randomized single factorial design was employed in which the 24 subjects were presented with eight hypotheses, one at a time, each hypothesis involving a different rule for combining two attribute values (conjunction, inclusive disjunction, conditional, biconditional, disjunctive absence, conjunctive absence, exclusion, and exclusive disjunction). The order of the rules was balanced over all subjects. The pair of attribute values present in each hypothesis were different, being randomly drawn from the 54 possible pairs of attribute-value combinations of two attributes.

b. Stimuli. - The stimuli consisted of geometrical designs printed on 7 x 10 cm. white cards. Designs varied according to four attributes with each attribute having three values, making 81 different patterns. The attributes and their values were colour (red, green, and black), shape (cross, circle, and square), number of figures (1, 2, and 3), and number of borders (0, 1, and 2). The 81 cards were arranged on a table before the subject in a systematic 9 x 9 array.

c. Procedure. - The details of the task were explained in the instructions which were presented on a typed sheet. The nature of the stimulus display was described and a definition given of an "hypothesis" as a rule specifying the presence or absence of two of the attributes of the cards in the display. Three examples of hypotheses were provided, viz., red and square, one border or not cross, not two figures or not black. It was pointed out that hypotheses such as these described different sets of cards.

Eight problems were presented, and at the start of each problem the subject was given an hypothesis. He was told that on each problem there was one and one only hypothesis which could describe and include the set of cards which had been adopted by the experimenter as the correct set. The subject's task was to choose those cards which would enable him to decide whether the hypothesis he was given was the one which defined the correct set. There was no limit to the time the subject could spend on each problem, nor was there a limit to the number of cards he could choose, but he was instructed to endeavour to point only to those he considered necessary to test the hypothesis he was given. The subject was told that the purpose of the experiment was not merely to see whether he obtained the right answer, but more importantly to see how he went about finding this answer. Thus, when he considered he had chosen enough cards to test the hypothesis, he should indicate accordingly, and that would end the problem. No decision was required about the truth or falsity of the hypothesis, and therefore no feedback was provided for the cards selected.

Hypotheses were presented to the subject typed on 7.5 x 13 cm. file cards. The hypotheses were: for conjunction, "black and cross"; for inclusive disjunction, "three figures or circle or both"; for conditional, "green or not no border"; for biconditional, "one figure and two borders, or, not one figure and not two borders"; for disjunctive absence, "not red or not one border"; for conjunctive absence, "not square and not two figures"; for exclusion, "green and not square"; and for exclusive disjunction, "either no border or circle but not no border and circle."

2. Results

The results are summarized in Table 18 in terms of the mean and standard deviation of the number of card choices required in order to test the hypothesis provided for each rule.

Basic Rule	Mean	S.D.	Complementary Rule	Mean	S.D.
Conjunction	5.6	3.8	Disjunctive absence	9.8	9.1
Incl.disjunction	8.6	6.2	Conjunctive absence	6.9	4.5
Conditional	7.2	5.0	Exclusion	7.4	4.9
Biconditional	8.4	5.9	Exclusive disjunction	7.6	6.7

TABLE 18. Number of instances required to test hypothesis.

An analysis of variance was used to analyse this date, but no significant effect was obtained for rules - see Table 19.

Source	S.S.		D.F.		M.S.	F
Between subjects		3536.00		23		
Within subjects		3527.25		168		
Rules	261.42		7		37.35	1.84
Residual	3265.83		161		20.28	
Total		7063.25		191		

TABLE 19. Analysis of variance on number of card choices

Thus, the mean number of instances selected for all rules was 7.7 with a standard error of the mean of 0.4.

The nature of the instances selected was examined and a summary may be found in Table 20 in terms of the mean percentage of direct test instances selected for each rule. These percentages were compared with the percentage of direct test instances present in the stimulus population. A sign test showed that more direct test instances were selected than were present in the stimulus population for conjunction ($p < .01$), inclusive disjunction ($p < .05$), biconditional ($p < .05$), exclusion ($p < .01$), and exclusive disjunction ($p < .05$).

No differences were obtained for disjunctive absence, conjunctive absence, and conditional.

Basic Rule	Direct test selected	Direct test present	Complementary Rule	Direct test selected	Direct test present
Conjunction	48.8	11.1	Disjunctive absence	75.2	88.9
Inclusive disjunction	72.1	55.6	Conjunctive absence	46.6	44.4
Conditional	67.6	77.8	Exclusion	50.9	22.2
Biconditional	70.9	55.6	Exclusive disjunction	62.5	44.4

TABLE 20. Percentage of direct test instances selected and present in the stimulus population.

Instances selected were further classified according to the percentage of instances selected containing both (TT), one (TF and FT), or neither of the attribute values specified in the hypothesis for each rule. A summary of this data is given in Table 21. These percentages were compared with the percentages of each class of instance occurring in the stimulus population. For all rules there were 11.1% TT instances, 22.2% TF instances, 22.2% FT instances and 44.4% FF instances. A sign test showed: a higher percentage of TT ($\underline{p} < .01$) and a lower proportion of FF ($\underline{p} < .001$) instances were selected for conjunction; a higher percentage of TF ($p < .05$) and a lower percentage of FF ($p < .05$) instances for inclusive disjunction; for biconditional, the proportion of TT ($p < .05$) instances selected was higher while a lower percentage was found for TF ($p < .05$), FT ($p < .001$), and FF ($p < .05$) instances; for conjunctive absence, a lower percentage of TF ($p < .05$) and FT ($p < .01$) instances were selected; more TF ($p < .01$), and less FT ($p < .01$) and FF ($p < .001$) instances on exclusion; and more TT ($p < .05$) and FT ($p < .05$) and less FF ($p < .001$) instances on exclusive disjunction. No significant differences were found between the percentage of each type of instance selected and present in the stimulus population for disjunctive absence and for conditional.

Basic Rule	TT	TF	FT	FF	Complementary Rule	TT	TF	FT	FF
Conjunction	48.8	25.7	17.0	8.5	Disjunctive absence	24.7	17.3	20.3	37.6
Incl. Disjunction	15.3	30.6	26.2	27.8	Conjunctive absence	27.9	13.2	12.2	46.6
Conditional	13.6	17.7	32.4	36.3	Exclusion	21.3	50.9	10.7	17.1
Biconditional	36.1	16.3	12.8	34.8	Exclusive disjunction	25.5	33.1	29.4	11.9

TABLE 21. Percentage of TT, TF, FT, and FF instances selected.

Instances may also be classified as being logically informative or redundant according to whether or not they eliminate any alternative hypotheses. Alternative hypotheses in this analysis refers to those having two attribute values combined by the same rule which is present in the given hypothesis. It should be noted that subjects were not told to assume that the rule was correct. However, subject's comments following the task suggested that the majority did not consider testing the rule, only the relevance of the attributes. A summary of this analysis may be found in Table 22 in terms of the mean number of informative and redundant instances chosen for each rule.

Basic Rule	Informative	Redundant	Complementary Rule	Informative	Redundant
Conjunction	2.7	2.9	Disjunctive absence	5.8	4.0
Incl. disjunction	5.1	3.5	Conjunctive absence	4.2	2.7
Conditional	4.8	2.4	Exclusion	4.2	3.2
Biconditional	4.8	3.7	Exclusive disjunction	4.3	3.3

TABLE 22. Mean number of informative and redundant choices.

An analysis of variance was performed on the number of informative choices and on the number of redundant choices made. No difference was found between rules for redundant choices, but a significant effect was obtained for rules for the number of informative instances selected, F (7,161) = 5.41, p .001. Newman-Keuls post hoc comparisons showed that this effect was due to fewer informative choices being made for conjunction than for other rules.

The likelihood that a subject might be correct in his evaluation of the hypothesis provided can be estimated by assuming that all instances chosen confirmed the hypothesis and then determining the number of alternative hypotheses which could also account for the sequence of instances chosen. According to the statistical decision process model, under these circumstances the subject would conclude that the given hypothesis is true. Thus the greater the number of alternative hypotheses which could also have produced the sequence, the more likely is it that the subject would have been wrong in his conclusion. Table 23 shows the frequency distribution for the number of possible hypotheses remaining at the end of the evaluation process. Clearly more than half the subjects did not completely eliminate all alternative hypotheses. A chi-square test indicated however that there was no difference in the frequency distributions for the different rules.

Basic Rule	H=1	2≤H≤5	H>5	Complementary Rule	H=1	2≤H≤5	H>5
Conjunction	14	5	5	Disjunctive absence	7	7	10
Inclusive disjunction	11	5	8	Conjunctive absence	10	7	7
Conditional	7	5	12	Exclusion	12	4	8
Biconditional	11	6	7	Exclusive disjunction	8	11	5

TABLE 23. Number of subjects and the number of hypotheses not eliminated.

3. Discussion

The results indicate that there is no difference in the number of instances chosen to evaluate an hypothesis for different rules. This finding confirms the prediction made by a statistical-decision process model in which it is assumed that the number of instances required is independent of the type of hypothesis. It is not consistent, however, with an information-processing model for which some difference was expected between rules. In this regard a difference may be noted between the present result and the results in earlier experiments where focussing strategies were inferred to be utilized.

An analysis of the type of instances selected confirms Bruner et al's suggestion that subjects tend to make direct tests of hypotheses and extends this suggestion to nonconjunctive rules. There was no evidence that subjects preferred an indirect method even on rules like disjunctive absence and conditional where such instances were likely to yield substantially more information on the truth of the hypothesis.

Further analysis shows that a preference for direct tests and an avoidance of indirect tests is not general for all classes of direct and indirect instances. Indeed there is evident for some rules a tendency not to select some classes of direct test instances and also a tendency to select some classes of indirect test instances. For example, for the biconditional the proportion of instances with neither of the attribute values in the hypothesis was less than that occurring in the stimulus population; for exclusive disjunction, a higher proportion of instances with both attribute values in the hypothesis were selected than were present in the stimulus population. For the latter rule, it might perhaps be argued that instances possessing both of the two attribute values specified in the hypothesis are in fact direct test instances since the wording of the rule does say, "... but not both". A similar explanation cannot be given, however, for the biconditional.

For those rules on which a direct test is a logically more informative method for evaluating the hypothesis, it may be expected that subjects were more successful in eliminating all other hypotheses than for those rules on which an indirect approach is the logically more informative. The results showed, however, that there was no difference between rules for the number of subjects who eliminated all but one hypothesis with their choices. This suggests therefore, that either a greater number of informative choices were made on those rules in which a direct test was less informative than an indirect test or, if a direct test yielded more information, then a greater number of redundant cards were selected. Analysis of informative and redundant choices indicated only that the former alternative was true for conjunction for which fewer informative choices were made than for other rules, but no other differences were obtained between rules. This analysis, however, did not distinguish between the extent to which an instance was logically informative; an instance which eliminates one hypothesis is classified equally with an instance which eliminates 84.

More importantly, the assertion that subjects are using a statistical-decision as opposed to an information-processing approach reflects on the meaningfulness of logically classifying each card choice as informative or redundant in the sense that the choice eliminates one or more alternative hypotheses. The analysis of the number of informative and redundant choices made on each rule described in the results is based on the false assumption that the subject is utilizing all the information which is contained in every instance, and may therefore be disregarded as having any relevance to a description of any subject's behaviour during the task. To say that a logical analysis of choices made reveals that in the course of their selections 41.7% of subjects eliminated all but one hypothesis is meaningful, however. It is meaningful not due to any description of the process of evaluation, but because, as a consequence of effectively removing all alternative possibilities which could also have produced a similar outcome on each selection, any decision made by these subjects concerning the truth of the hypothesis should be veridical.

At the end of the task all but three subjects indicated when questioned that the task would have been easier if feedback had been provided on each instance selected. An experiment was carried out as a follow-up to this study in which such feedback was supplied. Obviously, if feedback is to be given, the experimenter must know the solution to each problem, and hence, for some problems the initial hypothesis was true, but in others it was false. If the hypothesis is true, then according to a statistical-decision model, the subject will tend to select _n_ confirming instances of the hypothesis after which he will judge the hypothesis to be true. The value of _n_ will depend on the criterion for being right held by the subject. In the experiments described here the criterion was set by the instructions; subjects were to use a high or safe criterion for successful solving of the problem. No difference is expected between rules for this model when the hypothesis is true. If the hypothesis is false, then a new hypothesis must be sampled and tested in similar fashion to the old until an hypothesis is found which meets the criterion for accepting it as true. If no memory of instances is assumed, then no difference should be expected between rules for the number of instances selected before arriving at the solution. However, if a consistency assumption is operative, then some differences may appear between rules when the given hypothesis is found to be false. If, however, subjects tend to evaluate the hypothesis by logically eliminating other possible hypotheses using a focussing strategy, for example, then it can be predicted that the number of instances required is dependent on the logical rule of which the hypothesis consists.

Two other minor modifications were made in the experiment to be described from Experiment VII. The first was to present subjects with three problems instead of one involving the same rule. The subjects were then instructed that the solution to each problem involved the same rule combined by two attribute values. All they had to determine was whether the two attribute values in the hypothesis were the same two attribute values in the solution, and if not, what the two attribute values were for the solution. Thus, the type of solution involved in this study was more defined than in the previous experiment. According to a statistical-decision model, this should not make any difference to the number or type of instances chosen to evaluate the hypothesis. Secondly, it was desired to increase the likelihood of hypothesis-testing behaviour and so the stimuli were randomly placed in a 9 x 9 array instead of being systematically arranged. Following Bruner et al. (1956), this should increase the incidence of scanning.

Experiment VIII: Hypothesis-testing with true and false hypotheses

1. Method

a. Subjects and Design. - The subjects were 96 undergraduate students with equal numbers of both sexes randomly selected from the first-year psychology subject pool at the University of Adelaide. An 8 x 3 repeated-measures factorial design was used with the variables (a) conceptual rule (conjunction, inclusive disjunction, conditional, biconditional, disjunctive absence, conjunctive absence, exclusion, and exclusive disjunction) and (b) problems (three for each subject). For all subjects, on the first two problems initial hypotheses were true, while on the third problem the initial hypothesis was false. No subject was aware, of course, at the beginning of a problem whether his hypothesis was true or false. Twelve subjects were randomly assigned to one of the eight rules with the single limitation that there be equal numbers of each sex on each rule. The attribute values of the initial hypotheses were randomly chosen from the 54 possible pairs of attribute value combinations of two attributes, and a different initial hypothesis was presented to each subject on each problem for each rule. Where the initial hypothesis was false, the concept was randomly selected and was different for each subject.

b. Stimuli. - The stimuli consisted of geometrical designs printed on 7 x 10 cm. white cards. Designs varied according to four attributes with each attribute having three values, making 81 different patterns. The attributes and their values were colour (red, green, and black), shape (cross, circle, and square), number of figures (1, 2, and 3), and number of borders (0, 1, and 2). The 81 cards were arranged on a table before the subject in a random 9 x 9 array.

c. Procedure. - The details of the procedure were explained to the subject in the instructions which were presented to him on a typed sheet. At the beginning, the instructions noted that the experiment in which the subject was about to take part was designed to investigate ways in which people form concepts; it was not an intelligence test, nor would the results be used to assess individuals for examination purposes.

The nature of the stimulus display was then described and a definition given of a "concept" as "a certain set of these cards determined by the presence of two of the above attributes". The relevant type of concept was then explained with the aid of an example. The examples were: for conjunction, "all cards which are green and square"; for inclusive disjunction, "all cards which are either green or square or both"; for conditional, "all cards which are green or not square"; for biconditional, "all cards which are green and square, or, not green and not square"; for disjunctive absence, "all cards which are not green or not square"; for conjunctive absence, "all cards which are not green and not square"; for exclusion, "all cards which are green and not square"; for exclusive disjunction, "all cards which are either green or square but not both". It was ascertained that the subject understood the relevant conceptual rule concerned by pointing to several test instances, randomly selected from each truth-table category.

It was then explained to the subject that his task was to find concepts just like the one described. Three problems were to be presented, and on each problem a new concept was involved. The subject was to attempt to find the concept by selecting cards, one at a time, to which the experimenter would indicate whether or not it was a member of the set of cards determined by the concept. The subject was told that he could at any time volunteer his concept to the experimenter, but since he was allowed only one attempt at the concept, he was advised not to make such an attempt until he had established beyond reasonable doubt what the concept was; time was not important.

Each problem was started by presenting the subject with an hypothesis of the concept typed on a 7.5 x 13 cm. file card. It was explained to the subject that this hypothesis may be either true, i.e., it was the actual concept he was trying to find, or false. If the subject discovered that the hypothesis was false, then he was recommended to form a new hypothesis of his own making and test that to see whether it was true or false, until he considered the concept to be identified.

When the subject had finished reading the typed instructions, the experimenter asked him if there were any questions. An assurance was sought that the subject understood the difference between an hypothesis and a concept. The instructions were then summarized once more verbally. When the subject indicated that he was ready, the first problem was begun by presenting him with the first hypothesis. The instruction sheet was available to the subject throughout the task.

2. Results

The results are summarized in Table 24 in terms of the mean and standard deviation of the number of card choices to the criterion of proposing the concept for each rule and problem.

Basic Rule		Problem one	Problem two	Problem three	Complementary Rule		Problem one	Problem two	Problem three
Conjunction	M	5.9	6.6	11.2	Disjunctive absence	M	7.1	7.3	16.7
	S.D.	3.3	6.2	6.3		S.D.	3.6	4.6	11.2
Inclusive disjunction	M	7.9	6.4	13.0	Conjunctive absence	M	7.9	5.8	12.7
	S.D.	6.3	3.5	12.8		S.D.	5.3	3.0	13.2
Conditional	M	7.2	7.2	22.9	Exclusion	M	7.8	6.8	12.4
	S.D.	4.6	4.4	21.6		S.D.	4.2	4.1	4.1
Biconditional	M	9.1	6.5	25.7	Exclusive disjunction	M	6.3	5.3	17.0
	S.D.	6.1	3.2	33.3		S.D.	2.9	2.2	14.1

TABLE 24. Number of trials to criterion for each rule and problem.

An analysis of variance was performed on this data transformed by a square-root transformation in order to make the variances more homogeneous. A significant effect was obtained for the number of trials to criterion over problems, F (2,176) = 48.42, $p < .001$, but there was no significant effect over rules, nor any interaction between problems and rules. A summary of the analysis of variance may be found in Table 25. A post hoc analysis using Newman-Keuls multiple range comparisons (Winer, 1962) revealed significant differences between Problems 1 and 3 ($p < .01$), and 2 and 3 ($p < .01$), but no difference between Problems 1 and 2.

Source	S.S.	D.F.	M.S.	F	P
Between subjects	168.46	95			
Rules	11.51	7	1.64	0.92	N.S.
Subjects within groups	156.95	88	1.78		
Within subjects	278.61	192			
Problems	93.94	2	46.97	48.42	< .001
Problems x rules	13.99	14	1.00	1.03	N.S.
Problems x subjects within groups	170.68	176	0.97		

TABLE 25. Summary of analysis of variance on number of card choices to criterion transformed by square-root transformation.

An examination was made of the number of subjects whose proposal of the concept was correct - see Table 26. A Cochran Q test (Siegel 1956) was used to determine whether the number of subjects arriving at the correct solution differed on each problem. This analysis showed that significantly different numbers of correct solutions were provided on each problem for inclusive disjunction, conditional, biconditional, conjunctive absence, and exclusive disjunction. For these rules, inspection of Table 3 indicates that there were fewer correct solutions on Problem 3 than on Problems 1 and 2. For the first two problems, in which the initial hypothesis was the concept which subjects were seeking, the probability of an incorrect solution taken over all rules was 0.03. On Problem 3, however, the probability of an incorrect solution computed for all rules was 0.39.

Basic Rule	Problem			Complementary Rule	Problem		
	one	two	three		one	two	three
Conjunction	12	11	9	Disjunctive absence	12	11	9
Incl.disjunction	10	12	5	Conjunctive absence	12	11	6
Conditional	12	12	7	Exclusion	12	12	10
Biconditional	12	11	5	Exclusive disjunction	12	12	8

TABLE 26. Number of subjects attaining the concept on each problem.

A logical analysis was performed on the instances selected by each subject on each problem to determine the number of hypotheses remaining at the time the subject proposed his solution to the problem. A summary of this analysis is in Table 27 showing the number of subjects on each rule who

Basic Rule	Problem			Complementary Rule	Problem		
	one	two	three		one	two	three
Conjunction	10	9	10	Disjunctive absence	4	9	7
Incl. disjunction	6	5	5	Conjunctive absence	4	2	6
Conditional	3	3	10	Exclusion	8	7	11
Biconditional	7	6	8	Exclusive disjunction	9	5	9

TABLE 27. Number of subjects logically eliminating all but one hypothesis.

eliminated all but the correct solution with their card choices according to a logical analysis. A Cochran Q test (Siegel, 1956) was conducted on each rule but no difference was obtained between problems except for conditional ($\underline{p} < .01$) and disjunctive absence ($\underline{p} < .01$). Inspection of Table 27, however, does not indicate that the frequency with which only one hypothesis logically remained was less for Problem 3 than for Problems 1 and 2 as might be expected from the previous analysis of the number of subjects attaining the concept. Indeed, for the conditional, the frequency was higher on Problem 3 than on 1 and 2. It is evident that a logical analysis is not necessarily a good predictor that the concept will be identified. Clearly a number of subjects identified the concept when the initial hypothesis was true even though they had not eliminated all alternative hypotheses. When the initial hypothesis was false, however, subjects who logically eliminated all but one possible solution did display a significantly greater tendency to attain the concept than other subjects. In this case, when one hypothesis logically remained, the probability of a correct solution was 0.73; with more than one hypothesis remaining, the probability was 0.37. A 2 x 2 chi square using Yates correction on this data yielded $\underline{X}^2 = 9.85$, $\underline{p} < .01$.

The nature of the instances selected was analysed and a summary may be found in Table 28 in terms of the median percentage of positive instances chosen on each problem for each rule. These percentages were compared with the percentage of positive instances present in the stimulus population for each rule: conjunction 11.1%, inclusive disjunction 55.6%, conditional 77.8%, biconditional 55.6%, disjunctive absence 88.9%, conjunctive absence 44.4%, exclusion 22.2% and exclusive disjunction 44.4%. A sign test was used for this purpose with a correction on the significance level for each comparison

Basic Rule	Problem			Complementary Rule	Problem		
	one	two	three		one	two	three
Conjunction	100.0	100.0	35.4	Disjunctive absence	76.4	62.8	86.4
Incl. disjunction	87.4	79.2	62.3	Conjunctive absence	46.5	31.0	52.8
Conditional	68.4	74.6	62.9	Exclusion	95.5	85.9	45.3
Biconditional	63.4	63.1	58.4	Exclusive disjunction	66.7	69.1	43.7

TABLE 28. Median percentage of positive instances selected.

to ensure that spurious differences did not occur by chance (Duncan, 1955). The corrected significance level for each comparison was $\alpha = .003$ (two-tail) obtained using the formula $\alpha = 1 - (.95)^{1/k}$ where <u>k</u> represents the number of comparisons made. On this analysis it was found that a higher percentage of positive instances were selected than were present in the stimulus population on all problems for conjunction and exclusion, and on the first two problems for inclusive disjunction and exclusive disjunction. A lower percentage of positive instances were chosen on Problems 1 and 2 for disjunctive absence. No significant differences were obtained for conditional, biconditional, and conjunctive absence.

The type of instances selected on the first two problems were further analysed into those containing both (TT), one (TF and FT), or neither (FF) of the attribute values given in the initial hypothesis. It will be recalled that on the first two problems for each rule the initial hypothesis was true. The median percentage of each of these four classes of instance selected may be found in Table 29.

Rule	Problem							
	one				two			
	TT	TF	FT	FF	TT	TF	FT	FF
Conjunction	100.0	0	0	0	100.0	0	0	0
Incl. disjunction	17.0	25.0	38.2	12.7	19.1	29.2	30.4	20.9
Conditional	13.4	21.6	29.3	25.0	17.5	26.0	23.6	31.0
Biconditional	24.3	8.8	14.3	33.3	31.0	25.0	11.8	33.7
Disjunctive absence	25.0	19.5	27.8	4.6	37.2	25.0	19.4	16.1
Conjunctive absence	14.3	22.0	20.0	46.5	15.5	22.5	22.5	31.0
Exclusion	0	95.5	0	0	7.2	85.9	0	0
Exclusive disjunction	22.5	38.8	31.0	0	28.6	33.3	33.3	0

TABLE 29. Median percentage of instances selected with both (TT), one (TF and FT), and neither (FF) of the attribute values in the initial hypothesis.

These percentages were compared with the percentage of each class of instance occurring in the stimulus population - TT 11.1%, TF 22.2%, FT 22.2%, and FF 44.4% - using a sign test with a corrected significance level of α = .003. For conjunction, the significantly higher percentage of positive (TT) instances was due to a lower percentage of TF, FT, and FF instances being chosen. For exclusion, the higher percentage of positive (TF) instances selected was related to a lower percentage of FT and FF instances only. The higher proportion of positive instances selected for inclusive disjunction was brought about through a higher percentage of FT instances being chosen. For exclusive disjunction, the percentage of both TF and FT instances selected was higher than that which occurred in the stimulus population, but this was related to a lower percentage of only FF negative instances. For disjunctive absence, a lower percentage of FF positive instances was found. A lower percentage of FF instances was also obtained on conditional and biconditional, even though the overall result of the analysis of positive instances was not significant for these rules.

As will already be evident, the eight rules considered in this experiment fall into four pairs having the property that any instance which is positive under one member of the pair is negative under its complement. This means that a positive instance under, say, conjunction has the same information value as a negative instance under disjunctive absence. Previously it has been shown in this experiment that the number of instances selected to criterion has not differed between rule pairs. A comparison between the members of each pair was also made for the type of instances selected on each. For this comparison a Mann Whitney U test was performed on the percentage of positive instances selected on each "basic" rule and the percentage of negative instances selected on its complementary rule for each rule pair and problem. A correction was made to the significance level adopted for each comparison because of the number tested. On the first two problems significantly more positive instances were selected for conjunction, conditional, and biconditional than negative instances for disjunctive absence, exclusion, and exclusive disjunction respectively. This implies also that the percentage of negative instances selected for conjunction, conditional, and biconditional was lower than the percentage of positive instances for disjunctive absence, exclusion, and exclusive disjunction respectively. No differences however, were found between inclusive disjunction and conjunctive absence at α = .004. On Problem 3, where the initial hypothesis was false, significantly more positive instances were chosen for conjunction and inclusive disjunction than negative instances for disjunctive absence and conjunctive absence respectively. The reciprocal of this is therefore also true. No differences, however, were obtained on problem 3 between conditional and exclusion, or between biconditional and exclusive disjunction at $\alpha = .004$.

3. Discussion

The results indicate that there are no differences in the number of instances chosen for the various rules examined when subjects are asked to test an hypothesis which, unknown to them at the beginning of the problem, is true. This finding is reminiscent of Experiment VII where again no difference was found between rules in the number of instances chosen to evaluate an hypothesis, although in the experiment no feedback was provided on the truth value of the selections. Moreover, in the present task almost all subjects evaluated the given hypothesis successfully for all rules.

These findings may be taken as supporting the model proposed earlier where hypothesis-evaluation was considered as a decision-making process, as distinct from a process of logical elimination of alternative hypotheses. According to a logical-elimination process model, the minimum number of instances necessary to test an hypothesis varies with the rule. Two choices are necessary for conjunction and disjunctive absence, 3 for inclusive disjunction and conjunctive absence, and conditional and exclusion, and 4 for biconditional and exclusive disjunction. A statistical-decision process model, however, it will be recalled is independent of the rule and is a function only of the criterion for successful evaluation of the hypothesis held by the subject. In this task subjects on the different rules were all given instructions to adopt a safe or conservative criterion towards success. Assuming that these instructions were equally effective for the different groups of subjects, and the results suggest they were, it can be expected that no significant variation should take place between rules.

Other findings are also consistent with this model. Thus, it has been shown that more instances were required when subjects were provided with an initial hypothesis which was false than when given one which was true. This effect has been obtained in spite of the design of the experiment in which Problems 1 and 2 always began with a true hypothesis and Problem 3 with a false hypothesis. This result is of interest when considering subjects' performance in Experiment VII in the light of this new data. The means and standard errors of the mean of the number of choices made in Experiment VII and in this experiment for an initial true hypothesis were 7.7 and 0.4, and 6.9 and 0.3 respectively. A t test on these two sets of results shows that there is no difference between them. It seems likely, therefore, that the subjects in each case were behaving similarly, even though those in the present experiment were provided with feedback on each instance chosen whereas in the earlier task no feedback was available. The almost unanimous report by subjects in Experiment VII that some feedback on their instances would have helped them in their task is not reflected in the number of instances selected.

The finding that the means of the number of instances selected for the various rules lie between 5.3 and 9.1 with an overall mean of 6.9 is reminiscent of Miller's (1956) observation of a limit to the number of independant items which can be stored in immediate memory of between 5 and 9 with a mean of 7. This may lead one to speculate that the number of instances selected under a high criterion for successful evaluation of the hypothesis may be related to short-term memory capacity for storing information about the hypothesis. Using Miller's informational model of memory, this means that, if the feedback given on an instance agrees with the prediction made by the hypothesis under test, then one item of information concerning the hypothesis is stored in the first pigeonhole in short-term memory. If the second instance selected confirms the hypothesis, another item of information is stored in a second pigeonhole. This process is repeated until either all available pigeonholes in short-term memory have been filled, or a discrepancy occurs between the prediction of the hypothesis and the feedback given by the experimenter. If the former takes place, then a decision will be made to accept the hypothesis. If the latter, then all evaluative information in the pigeonholes is erased and a new hypothesis sampled in the manner discussed earlier in the section on hypothesis-sampling. This model assumes, of course, that the individual is operating under a high criterion for correct evaluation of the hypothesis. At other times when a risky evaluation is demanded, a decision will be made before all pigeonholes have been filled. This proposal is formulated in terms of Miller's informational model of immediate memory, but it will be realised that this is not the only model of short-term memory which could be applied.

It may well be asked what is the item of information which is stored in each pigeonhole. At least two possibilities exist. The first comes from the decision-making model and suggests that only an indicator is placed in the pigeonhole to show that one confirming test of the hypothesis has been made. In other words, the only information stored is the number of instances selected confirming the hypothesis. We may wonder in this case why 7 pigeonholes are needed for this purpose, why the numbers 1 to 7 could not equally well be stored in a single pigeonhole to show the number of confirming tests completed. If the latter system was employed, then the remaining pigeonholes could be utilized for retention of instances selected for use in sampling. This leads to the second major possibility, that the subject seeks to store in the pigeonhole the instance selected together with the associated feedback. In view of the high degree of confusability of the stimuli, it seems unlikely that the subject can remember 7 random instances. However, if some simple strategy is employed for choosing cards, then it may easily be possible to remember more than 7. This would account for the occasional subject who selected almost all positive instances before making an affirmative decision on the hypothesis. Clearly, this discussion is rather speculative. More will be said on the role of memory in hypothesis-testing following the next experiment.

Analysis of the nature of the instances selected for those problems on which a true hypothesis was given indicates that there is a tendency for subjects to select instances which are positive according to the hypothesis under test. This is evidenced by the number of rules for which a greater proportion of positive instances were selected than were present in the stimulus population. The result for disjunctive absence is probably related to the fact that negative instances provide greater information and are therefore a more effective test of the hypothesis involving this rule. That the logical effectiveness of a test of an hypothesis is not a dominant consideration, however, may be seen for inclusive disjunction, conditional, and biconditional. Comparisons between complementary rules also indicates a strong tendency to choose positive instances which constitute a direct test of the hypothesis as suggested by Bruner et al. (1956). The finding of no difference between the proportion of positive instances selected for inclusive disjunction and the proportion of negative instances selected for conjunctive absence may be a reflection of the comments of a number of subjects who found the wording of the conjunctive absence rule confusing and tended to rephrase it as a negated inclusive disjunction. Comparison between the present data and the data obtained in Experiment VII for the proportion of direct test instances reveals a close similarity in the percentages. The only apparent effect of feedback seems to be a slight increase in the percentage of direct test instances for conjunction, exclusion and perhaps inclusive disjunction.

Further analysis of the instances selected into four categories according to the presence or absence of the attribute-values specified in the hypothesis reveals additional similarities between the two sets of results for feedback and no feedback. Thus in both experiments there was a tendency to select instances having one attribute value for inclusive disjunction and exclusive disjunction, and a tendency not to select FT or FF instances for exclusion. The observation made following Experiment VII that a preference for direct testing of an hypothesis does not appear to apply equally to all direct test instances nor against all indirect test instances may be reiterated here. In this experiment, however, this finding is somewhat less general. For example, only FF instances were found to be infrequently chosen in the earlier experiment for conjunction, while in this experiment there was a significant tendency not to select any of the 3 types of indirect test instance for this rule.

On the third problem, when an initial false hypothesis was presented, no difference was again found between the number of instances selected to criterion for different conceptual rules. This result is somewhat unexpected for both models of hypothesis evaluation. It may be predicted from an hypothesis-sampling model which assumes recall of two instances selected that the probability of sampling the correct hypothesis will vary according to the rule since the number of hypotheses consistent with two instances depends on the rule concerned. Thus, if two positive instances are selected, a maximum of 6 conjunctive hypotheses will be available for sampling, but the minimum number of conditional hypotheses from which a choice can be made is 60. A similar expectation may be seen if recall of only one instance is assumed. Even if no memory is assumed and sampling is random with replacement, a difference between rules would be predicted because the hypothesis pool size for conditional and exclusion is twice that for the other rules.

The finding of no difference between rules with an initial false hypothesis cannot be taken as conclusive since, although most subjects were correct in the identification of the concept when the initial hypothesis was true, this was not the case on inclusive disjunction, conditional, biconditional, conjunctive absence, or exclusive disjunction when the initial hypothesis was false. It may be that if subjects whose first attempt at the concept was incorrect, were allowed to continue until they had attained the concept, then some differences may have been found. Further work is necessary to answer this question. The reason for allowing only one attempt at the concept in the present experiment was to minimise guessing, and hence to determine what is acceptable evidence for the truth of an hypothesis using a high or safe criterion.

The finding of a difference in successful concept attainment between initial true and false hypotheses may indicate either that the criterion utilized by subjects on the problems with the false hypotheses was different from that utilized with true hypotheses or that subjects attempted to recall previous instances prior to sampling of the true hypothesis and that errors in the recall of these instances led to erroneous attempts at the concept. Comments spontaneously made by a number of subjects on encountering an invalidating instance of an hypothesis after a series of confirming instances suggest that the former may be more likely. These subjects reported that they had forgotten which cards they had selected and so they would have to start all over again.

A number of questions on behaviour during hypothesis testing have already been raised in the preceding discussion, and there are other observations which also need further examination. For example, it was noted that subjects seemed to perseverate their responding to the initial hypothesis even after an invalidating instance. Some subjects spontaneously reported when they believed the hypothesis was false and in several cases more than one disconfirming instance had been chosen. Adequate data to answer these questions is not possible with the methodology employed in the present experiment. A further experiment is proposed here to provide more information about the processes of hypothesis sampling and evaluation.

In this experiment it is suggested that subjects be presented with either true or false hypotheses as before. True and false hypotheses should be counterbalanced over successive problems. The subject's task would be to determine the concept by testing hypotheses. On each trial the subject is required to select an instance for which the experimenter provides positive or negative feedback whichever is appropriate. Then, before the subject may select another instance, he is required to verbalize the hypothesis he is now testing together with some rating of his confidence in it. Trials should continue until the subject is extremely confident that the hypothesis he is holding is the concept. The introduction of confidence ratings should provide a further test of a statistical-decision model of hypothesis evaluation. According to this model, the confidence that a subject has in his hypothesis should increase with the number of instances chosen, mostlikely in the shape of a learning curve. Moreover, not only should the number of instances chosen to complete confidence be the same for each rule, but also the curve showing confidence plotted against number of instances showed be similar for the various rules. This method will enable us to follow the subject's confidence in his hypothesis at the point of re-sampling, to see whether more than one disconfirming instance is required and to see whether confidence for a new hypothesis returns to zero. Some explanation may then be found why more errors were made in concept identification with an initial false hypothesis. Also, some additional data can be collected on differences between rules when the initial hypothesis is false.

Another matter for further inquiry concerns the role of memory in hypothesis-testing. Both hypothesis sampling and hypothesis evaluation models make some assumptions about the retention of information. Of the hypothesis sampling models discussed earlier, the one which seemed to fit the data best postulated memory for two previous instances, viz., the first and the last in the sequence presented. A statistical-decision model for evaluation of the hypothesis does not require that any instance be recalled. Nevertheless, the possibility has been raised in the preceding discussion that some attempt to recall previous instances may have been made. Accordingly, an experiment was conducted to investigate immediate recall of instances and their associated feedback when testing true and false hypotheses. A similar procedure was employed as in Experiment VIII, except that the number of instances the subject could select was determined by the experimenter. The number permitted was arbitrarily set at 10 for both true and false initial hypotheses. The subject was not warned before the problem that he would be asked to try to recall as many of the first 10 instances selected as he was able.

Experiment IX: Immediate recall of instances during hypothesis-testing.

1. Method

a. Subjects and Design. - The subjects were 192 undergraduate students randomly selected from the first-year psychology subject pool at the University of Adelaide.

An 8 x 2 factorial design was used with the variables, (a) conceptual rule (conjunction, inclusive disjunction, conditional, biconditional, disjunctive absence, conjunctive absence, exclusion, exclusive disjunction), and (b) initial hypothesis (true, false). Twelve subjects were randomly assigned to each of the 16 experimental conditions, with the limitation that there be equal numbers of each sex on each condition. The attribute values of the initial hypotheses were randomly chosen from the 54 possible pairs of attribute value combinations of two attributes, and a different initial hypothesis was presented to each subject within each condition. Where the initial hypothesis was false, the concept was randomly selected and was different for each subject. All subjects were given one practice problem with no recall of instances before the test problem was presented. Again, the attribute values of the initial hypothesis for the practice problem were randomly selected and different for each subject within each condition.

b. Stimuli. - The stimulus display was the same as that used in Experiment VIII.

c. Procedure and Instructions. - The details of the procedure and instructions were identical to Experiment VIII with the following alterations. Firstly, subjects were not told how many problems would be presented. This was designed to prevent guessing of the truth or falsity of the initial hypothesis. Secondly, subjects were instructed that they would be allowed to select only a limited number of cards, but they were not told exactly how many. After this, they were told they would be asked to say what they thought was the concept. It was explained that this would complete the problem. Since the exact number of cards permitted to be selected would not be given at the beginning of a problem, subjects were encouraged to try to work as efficiently as possible. If, by chance, they happened to discover the concept before the experimenter required them to stop, then they should check their solution by choosing further cards until they were interrupted.

Two problems were presented to each subject who was allowed to select 10 instances on each problem. On the first or practice problem, the initial hypothesis was true, and the subject was simply required to give his attempt at the concept as instructed. On the second or test problem, before he was allowed to propose the concept, the subject was asked to point to those cards he was able to remember having just chosen in the second problem, and to say whether they were members of the concept or not. No special order of recall was required. After he had recalled as many instances as he could clearly remember, the experimenter asked him to make his attempt at the concept.

2. Results

The mean and standard deviation of the number of instances correctly recalled out of 10 was calculated for each rule and each type of initial hypothesis. A summary of this data may be found in Table 30.

Basic Rule		Initial Hypothesis		Complementary Rule		Initial Hypothesis	
		True	False			True	False
Conjunction	M	6.6	5.1	Disjunctive absence	M	4.8	5.5
	S.D.	2.9	1.7		S.D.	2.3	2.1
Incl. disjunction	M	5.0	4.8	Conjunctive absence	M	5.0	3.7
	S.D.	2.0	2.2		S.D.	2.4	1.7
Conditional	M	4.0	5.5	Exclusion	M	5.8	6.4
	S.D.	2.1	1.8		S.D.	2.8	2.0
Biconditional	M	3.8	4.6	Exclusive Disjunction	M	5.6	3.9
	S.D.	1.9	2.2		S.D.	2.8	1.8

TABLE 30. Number of instances correctly recalled.

An analysis of variance was performed on this data. A significant effect was obtained between rules, $F\ (7,176) = 2.13$, $p < .05$, but there was no significant effect of initial true and false hypotheses, nor a significant interation between initial hypothesis and rule. A Newman-Keuls post hoc analysis on individual pairs of rules was unable to detect significant differences between any pair of rules at the 0.05 level. A summary of the analysis of variance may be found in Table 31.

Source	S.S.	D.F.	M.S.	F	P
Initial hypothesis (H)	.88	1	.88	0.17	N.S.
Conceptual rule (R)	76.20	7	10.89	2.13	<.05
H X R	63.33	7	9.05	1.77	N.S.
Error	898.58	176	5.11		
Total	1038.99	191			

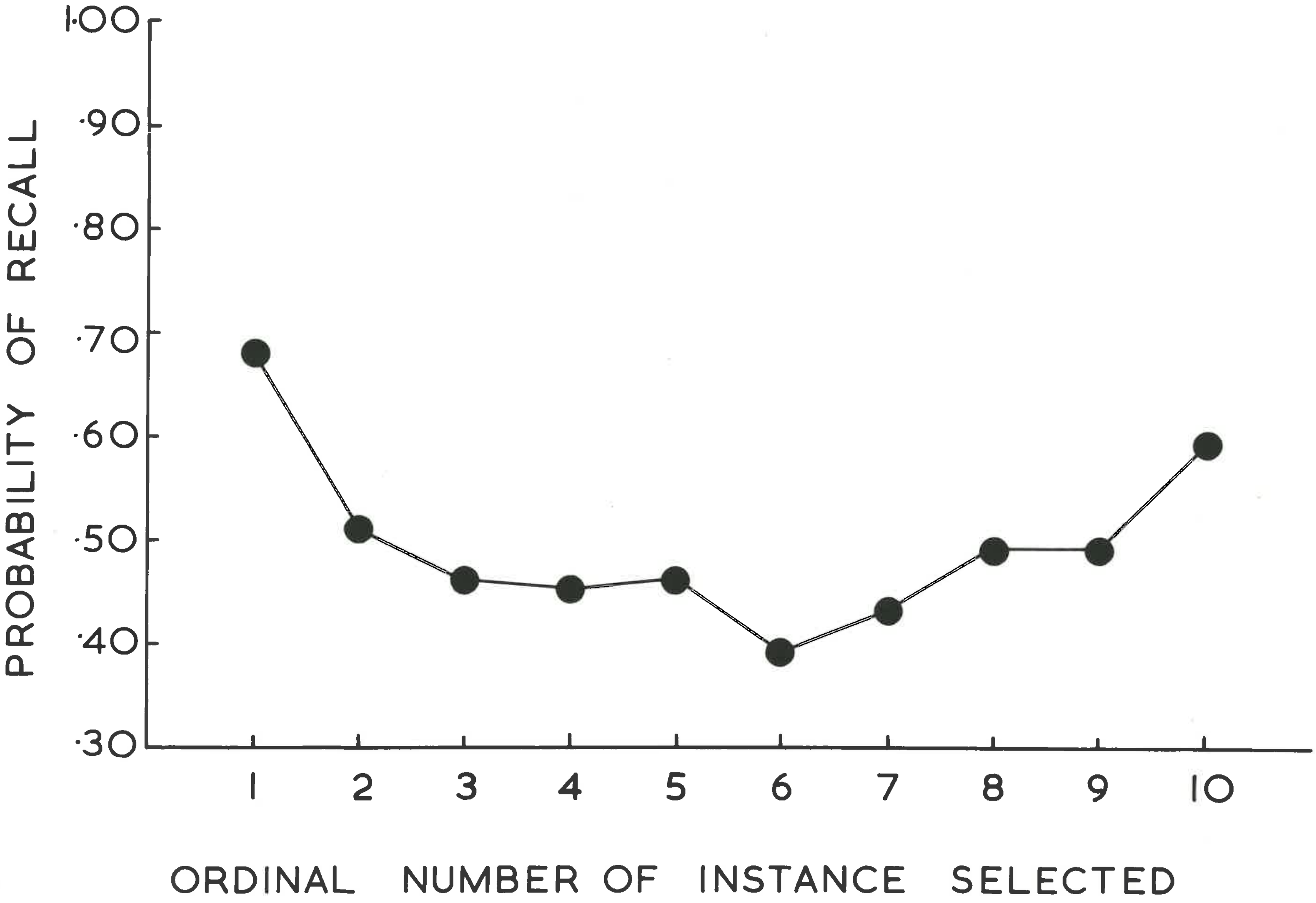
TABLE 31. Analysis of variance on number of instances correctly recalled.

The effect of the position of the instances in the sequence selected on recall was investigated by means of a chi-square analysis. This showed a significant difference between positions for recall; $\chi^2_9 = 67.08$, $p < .001$, but no interaction between position in the sequence and the initial hypothesis nor between position in the sequence and rule. Figure 5 shows the probability that an instance was correctly recalled plotted against the ordinal position of the instances selected. It can be seen that this is a conventional serial position curve with recall best for the first and last instances.

Examination was made not only of the number of instances recalled, but also of the errors committed on recall. It was noticeable that most subjects (81.2%) made at least one error of commission during recall in addition to other errors of omission. Instances falsely recalled were classified into three groups: (1) instances whose truth value was incorrectly reported, (2) instances recalled differing in only one attribute value from an instance selected, and (3) instances with more than one attribute value difference from those selected. Only a relatively small proportion of subjects (21.4%) falsely recalled the truth value of an instance. The frequency of category 1 errors was greater for an initial hypothesis which was false than for one which was true, $\chi_1^2 = 5.49$, $p < .05$. Clearly the majority of errors (71.9% of all subjects) differed by one attribute value from an instance selected. The third type of error was committed by only 28.1% of subjects.

The instances selected were divided into positive and negative instances and an analysis was performed to see whether there was any difference in the recall of these two classes of instances. Since different numbers of positive and negative instances were selected by different subjects, scores were standardized by computing the proportion of positive instances recalled and the proportion of negative instances recalled. Subjects who chose instances all of the same type were discarded from the analysis. The remaining subjects were sorted into

FIGURE 5. Probability that a selected instance is correctly recalled plotted against the ordinal position of the instance selected.



1·00
·90
·80
·70
·60
·50
·40
·30
PROBABILITY OF RECALL
1
2
3
4
5
6
7
8
9
10
ORDINAL NUMBER OF INSTANCE SELECTED

three groups according to whether they recalled: (1) a higher proportion of positive than negative instances, (2) a higher proportion of negative than positive instances, or (3) the same proportion of positive and negative instances. The number of subjects in the first two groups is shown in Table 32 for initial true and false hypotheses and for all conceptual rules examined in the experiment. A chi square analysis on this data revealed that the frequency with which a higher

Rule	More Positives Recalled			More Negatives Recalled		
	Initial Hypothesis			Initial Hypothesis		
	True	False	Total	True	False	Total
Conjunction	4	4	8	1	1	2
Incl. disjunction	5	6	11	2	5	7
Conditional	6	3	9	3	6	9
Biconditional	7	9	16	1	1	2
Disj. absence	3	2	5	5	6	11
Conj. absence	2	5	7	5	6	11
Exclusion	3	5	8	2	4	6
Excl.disjunction	4	9	13	4	2	6
Total	34	43	77	23	31	54

TABLE 32. Frequency with which a higher proportion of positive instances than negative instances were recalled and frequency with which a higher proportion of negative instances than positive instances were recalled.

proportion of positive instances than negative instances were recalled was significantly greater than the frequency with which a higher proportion of negative instances than positive instances were recalled, $\chi^2_1 = 4.04$, $p < .05$. From this it may be concluded that recall of positive instances was better than of negative instances. Recall of positive and negative instances was also found to interact with the rule involved, $\chi^2_7 = 17.88$, $p < .05$, but not with the type of initial hypothesis presented. Further analysis reveals a clear tendency to recall positive instances better than negative instances on biconditional, $\chi^2_1 = 10.89$, $p < .001$, a similar but marginal tendency on conjunction, $\chi^2_1 = 3.60$, $p < .10$, but for disjunctive absence and conjunctive absence inspection of Table 32 shows the tendency is in the reverse direction.

After the subject had recalled as many instances as he was able, he was asked to attempt to name the concept. A summary of the number of times the concept was correctly identified is given in Table 33. A chi-square analysis showed a significant effect of the type of initial hypothesis presented on concept identification, $\chi^2_1 = 56.35$, $p < .001$. There was no difference between rules, nor any significant interaction between the initial hypotheses and rules.

3. Discussion

Several findings of interest have been obtained in this

Rule	Initial Hypothesis		Total
	True	False	
Conjunction	11	2	13
Inclusive disjunction	10	2	12
Conditional	10	1	11
Biconditional	10	1	11
Disjunctive absence	11	1	12
Conjunctive absence	10	1	11
Exclusion	11	1	12
Exclusive disjunction	9	1	10
Total	82	10	92

TABLE 33. Frequency of concept identification.

experiment. Firstly, a small difference has been found between rules for the number of instances successfully recalled. This result was not predicted by the models of hypothesis sampling and evaluation considered. In the previous experiment it will be remembered that no difference between rules was found for the number of choices to criterion and this was taken to agree with the predictions of a statistical-decision process of evaluation. This new finding may be related to the amount of information processed over the 10 trials for the various rules since, according to Posner (1965), the amount of information retained is inversely dependent on the amount of information processed. Empirical support for this notion comes from Posner and Rossman (1965) who studied the effect on material in short-term memory of information transformations of varying levels of difficulty and found that more complex transformations led to increased forgetting. Thus, it seems that the amount of information processed may vary according to the rule such that the fewer is the number of instances recalled, the more is the amount of information processed. Since the same number of instances were presented on each rule, this suggests that the amount of information processed on each individual trial must be a function of the rule. A similar conclusion may be derived from the results of Denny (1969a) who showed that recognition-memory error over the first 10 trials of an attribute-identification task was greater for conceptual rules requiring more trials to criterion.

The nature of the information processing in this task which might give rise to differences between rules is not easy to specify. Two possibilities, however, may be considered. Firstly, subjects may be attempting to process information so as to logically eliminate hypotheses. We may reject this possibility as unlikely in view of the evidence already presented against it and because the obtained differences between rules do not conform to the expected order of rules for recall of instances. The second possibility concerns the process of checking on each trial whether the feedback provided by the experimenter on the truth value of the instance selected coincides with the truth value predicted by the hypothesis the subject is testing. The suggestion is that this process may vary slightly in difficulty for different rules. Two stages can be distinguished in the process: (1) determination of the truth value predicted for the instance by the hypothesis, and (2) comparison of this prediction with experimenter's feedback. It may be reasonably expected that the second stage is independent of the rule. Hence, this second possibility proposes that the differences in recall between rules are due to differences in the difficulty of classifying instances as positive and negative. An experiment seeking to explore this possibility is described later in this chapter.

Another finding of interest is that indicating better recall for positive instances than negative instances. This result represents an addition to the evidence already accumulated showing a preference to select positive instances rather than negative instances. If we follow again the argument that the amount of information processed is inversely related to the amount of information recalled, then it would seem that subjects are required to process more information on negative than positive instances. That is, an indirect test of an hypothesis necessitates greater information processing than a direct test. This conclusion, however, requires some qualification because of an interaction between conceptual rule and positive and negative instance recall. Perhaps this can be conceptualized by referring again to the two-stage process suggested above for confirming or disconfirming an hypothesis. For the first stage some differences may be expected in the ease of categorising positive and negative instances for different rules. For example, negative instances should be easier to identify than positive instances for conjunction, while the reverse ought to be true for conditional. In general, the type of instance which is the more probable should be classified more quickly for each rule. Strategies for classifying instances for various rules will be discussed in more detail later. For the second stage, there are only four possible contingencies from which the subject can make his decision whether the hypothesis is confirmed or disconfirmed. A summary of these contingencies and the appropriate decision to be made in each case is given in Table 34. It is not immediately obvious what differences, if any are likely to be found between these four possibilities for the ease of deciding to confirm or disconfirm the hypothesis. Some previous work has sought to compare the time for making judgements of sameness and

Hypothesis Prediction	Instance Feedback	Decision on Hypothesis
Positive	Positive	Confirm
Positive	Negative	Disconfirm
Negative	Positive	Disconfirm
Negative	Negative	Confirm

TABLE 34. Four possible contingencies for making a decision on hypothesis under test.

difference of stimuli but the results are equivocal and clearly subject to a number of procedural variables (Posner and Mitchell, 1967; Nickerson, 1965; Bindra, Williams, and Wise, 1965; Entus and Bindra, 1970). Further experimentation is necessary to determine the effects of the various contingencies of prediction of the hypothesis and feedback on the instance on the time taken to reach a decision about the hypothesis.

The comment made by several subjects in Experiment VIII upon disconfirming an hypothesis that they could not recall which instances they had selected and would have to start again is not supported by the results of this experiment. No difference is evident between initial true and false hypotheses either for the total number of instances recalled or for frequency of recall of each serial instance selected. If the subjects' comment had any validity, one would expect it to appear on the first one or two instances. However, the probability of recall of the first instance was 0.68 for both initial true and false hypotheses and this was the most frequently recalled instance for both types of hypotheses. Examination of Figure 5 shows a clear serial position curve with both primacy and recency effect in recall, which confirms the estimations of Levine (1969), and also Trabasso and Bower (1964b) and Bourne and O'Banion (1969).

The number of instances successfully recalled is rather more than was expected by hypothesis sampling and evaluation models. The hypothesis sampling model which gives the best fit to sampling data proposes retention of the first and last instances presented. A statistical-decision model of evaluation does not necessitate storage of any instances. A combination of these two models, therefore, suggests that only two instances need be retained in hypothesis-testing. Taken over all rules and both initial hypotheses, the mean number of instances recalled in this experiment is 5.0 with a standard error of the mean of 0.2. It seems apparent that subjects are able to recall more instances than these two models predict. Some account of this result may be drawn from the fact that the procedure followed permitted subjects to indicate as many instances as they could remember selecting. In doing so, almost all subjects displayed some intrusions in their recall. Thus, it seems likely that subjects are reporting instances, some of which they may be completely certain that they have previously selected, but also others about which they are not so certain. The importance of the response criterion for recall has been emphasized in the application of signal detection theory to memory (Banks, 1970). In the course of a concept identification task subjects may be somewhat more conservative in recalling past information which they wish to utilize in sampling or evaluating an hypothesis. In other words, in hypothesis testing subjects may tend to utilize only information which is derived from instances for which they are most confident about having selected on a previous trial, even though there may be other instances that they can recall with less certainty. If instructions to subjects specified that they were to recall only those instances which they were certain they had selected and that they should be careful to avoid any errors in recall, the number of instances reported may be nearer to that expected. One implication from this discussion is that the subject may be able to control the degree of risk in hypothesis testing not only by the number of confirming instances selected during evaluation but also the number of instances recalled during sampling.

Experiment X: Classifying instances of a concept

We shall return now to considering more fully stage one of the two-stage process outlined earlier for comparing the feedback provided by the experimenter on the truth value of an instance and the predicted truth value of the instance according to the hypothesis being tested. Stage one involves the classification of the instance as positive or negative by the hypothesis under test. Two inferences were made in the preceding discussion about stage one: (1) that the ease with which instances are classified depends on the rule, and (2) that positive instances are more easily classified on some rules than negative instances.

An experiment was done to examine these inferences. A choice reaction-time task was used in which subjects were required to tap one of two morse keys depending on whether a stimulus presented was a positive or negative exemplar of a concept. The concepts presented to the subjects at the beginning of the task involved one of four basic rules - conjunction, inclusive disjunction, conditional, and biconditional. The stimuli were randomly drawn from the stimulus population employed in the previous experiment. The ease of classification was measured by the time taken to respond.

The subject's task may be viewed in at least two ways. The first is to regard the difficulty of classification as a function of the likelihood of occurrence of each alternative. The more probable is the occurrence of an alternative, the faster is the choice of that alternative likely to be made. Information theory has proved useful in the past for expressing mathematically the relationship between the response time and the probability of an alternative in a choice reaction-time task. The version of Hick's Law (Welford, 1968) which is applicable to the present situation where there are unequal probabilities of each alternative is

$$\text{Mean choice reaction time} = \underline{K} \times \sum_{\underline{i}=1}^{n} \underline{p}_{\underline{i}} \log (1/\underline{p}_{\underline{i}} + 1)$$

This approach does not enable us to account for any differences between rules which may occur but it does make clear predictions about the relative difficulty of categorising positive and negative instances on each rule. Thus, positive instances should have faster reaction times on conditional, inclusive disjunction, and biconditional, while negative instances should be classified more rapidly on conjunction.

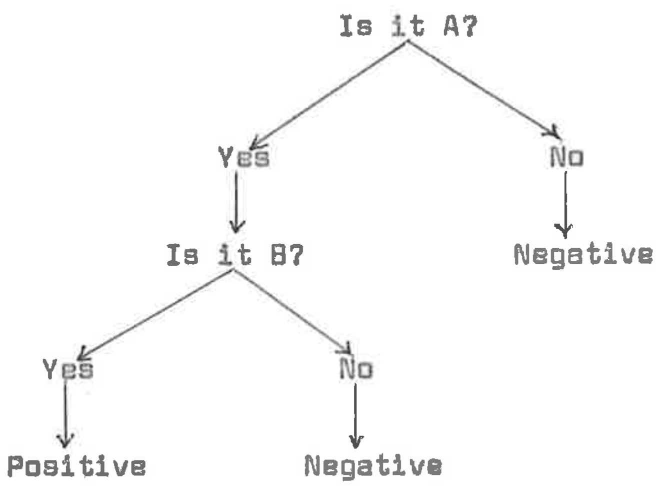
Another way of looking at the subject's task is to consider the strategies which may be used to identify positive and negative instances on each rule. These may be seen best by constructing decision trees in which the presence of each of the attribute values specified in the concept given to the subject is tested. Simple decision trees for the four rules may be found in Figure 6. Using these strategies, negative instances on conjunction and positive instances on inclusive disjunction and conditional should be least difficult, with positive instances on conjunction and biconditional and negative instances on inclusive disjunction, conditional, and biconditional having greater difficulty. Thus positive and negative instances are expected to differ for all rules except biconditional, while the biconditional should be more difficult than other rules.

1. Method

a. Design and Subjects. - The subjects were 80 undergraduate psychology students who took part in this experiment as a course requirement.

A 4 x 2 repeated measures factorial design was used with the variables (a) conceptual rule (conjunction, inclusive disjunction, conditional and biconditional), and (b) instance (positive and negative). Twenty subjects were assigned to one of the four rules at random. Each subject was presented with a different concept in which the relevant attribute values had been randomly selected. The sequence of instances presented to the subject was the same for all subjects on all rules, and consisted of 80 stimuli in random order, i.e., the order of positive and negative instances was random. The ratio of the number of positive to negative instances varied for different

(a) Conjunction A∩B



(b) Inclusive disjunction A∪B

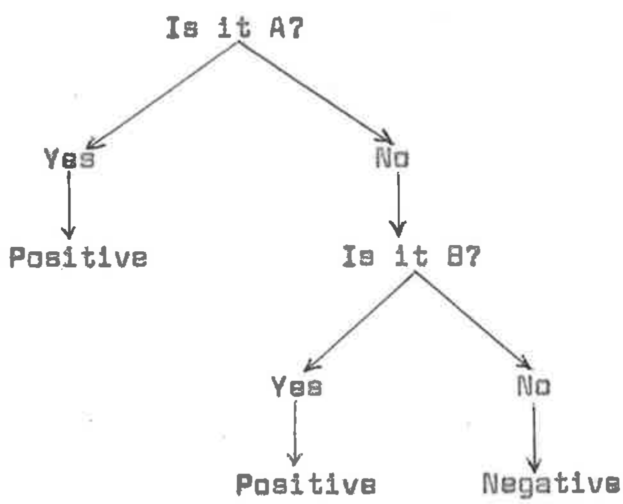
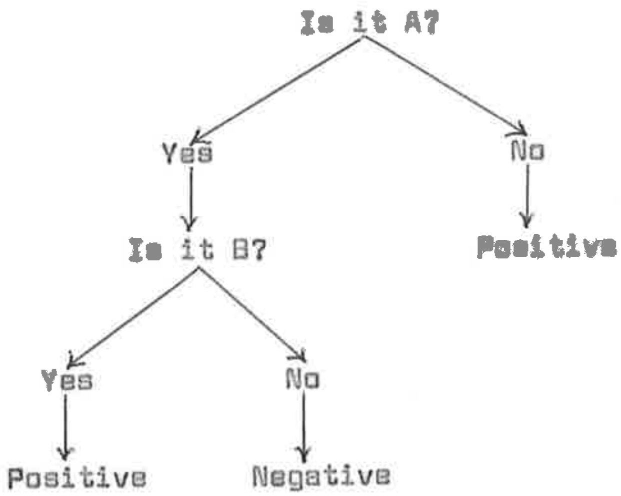
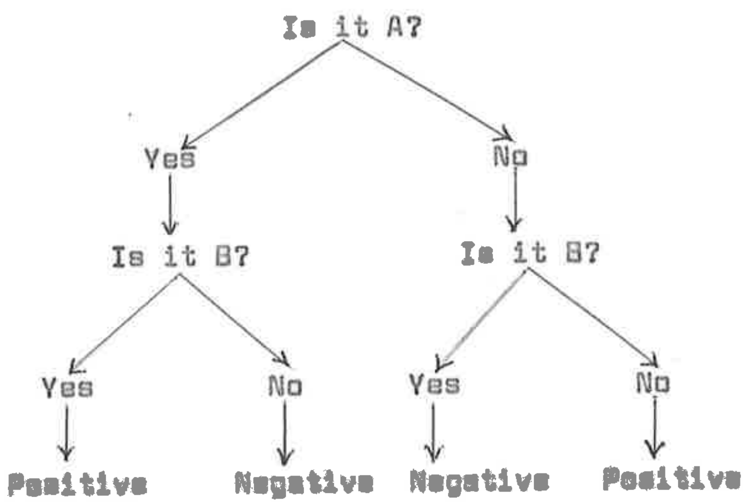


FIGURE 6. Decision trees for classifying instances

(c) <u>Conditional A→B</u>



(d) <u>Biconditional A↔B</u>



FIGURES 6 (cont.) Decision trees for classifying instances conceptual rules - conjunction 1:8, inclusive disjunction 5:4, conditional 7:2, biconditional 5:4.

b. Stimuli. - The stimuli consisted of geometrical designs varying according to four attributes, with each attribute having three values. The attributes and their values were colour (red, green, and black), shape (cross, circle, and square), number of figures (1, 2, and 3), and number of borders (0, 1, and 2). Each stimulus was contained on a 35 mm slide. Out of the 81 different patterns, one was randomly eliminated because of the limited capacity of 80 slides for the projector used.

c. Apparatus. - Slides were projected onto a screen placed approximately 12 feet from the subject with the aid of a Kodak Carousel 35 mm projector. A control timer was connected to the projector so that a new slide was presented every 10 seconds. Whenever a new slide appeared on the screen, an electric timer was activated, and also a buzzer sounded. Both the timer and the buzzer were stopped when the subject depressed one of two morse keys in front of him. Indicator lights enabled the experimenter to determine which of the two keys had been depressed each time. The task was carried out in a darkened room.

d. Procedure and Instructions. - The instructions began by explaining to the subject that his task in this experiment was to use a concept to categorize slides into those which were examples of the concept, and those which were not. For the purposes of this experiment, a concept was a rule which defined a certain set of patterns. An illustration was provided using a display of 81 cards arranged on a table before the subject in a systematic 9 x 9 array. It was pointed out to him that the attributes and their values printed on the cards were the same as those to be found on the slides. The relevant type of concept was then explained with the aid of an example. The examples were: for conjunction, "all cards which are green and square"; for inclusive disjunction, "all cards which are either green or square or both"; for conditional, "all cards which are either green or not square"; and for biconditional, "all cards which are green and square or, not green and not square". It was ascertained that the subject knew the general nature of the relevant concept type by having him count all the positive and negative instances for the sample concept provided.

At the beginning of the task, a concept involving the same rule but different attribute values was presented to the subject. Ten seconds later, a slide was projected onto the screen, portraying a pattern similar to the patterns on the cards. The subject's task was to decide whether the pattern was an exemplar or not of this concept. If it was an exemplar, then he depressed the right-hand key; if it was not, then he depressed the left-hand key. The subject was instructed that he should try to respond as quickly as he could without making errors. Ten seconds after the first slide, a second slide was presented, and so on, until all 80 slides had been shown.

2. Results

Both response time and correctness of response were recorded. Few errors in categorising instances were made on any rule, insufficient for a reliable measure of error reaction time to be calculated. The data used in this analysis, therefore, was response time for correct responses only and a summary of this data may be found in Table 35.

Rule	Type of instance			
	Positive		Negative	
	Mean	S.D.	Mean	S.D.
Conjunction	600	215	459	86
Inclusive disjunction	553	135	584	130
Conditional	506	163	780	284
Biconditional	1057	329	1029	343

TABLE 35. Response times (msec.) for classifying instances as positive or negative.

An analysis of variance was performed on this data transformed by a square root transformation in order to make the variances more homogeneous. A significant difference was obtained between positive and negative instances, $\underline{F}$ (1,76) = 4.00, $\underline{p} < .05$, and between rules, $\underline{F}$ (3,76)= 26.76, $\underline{p} < .001$. The interaction between rules and the type of instance was also significant, $\underline{F}$ (3,76) = 25.26, $\underline{p} < .001$. A summary of the analysis of variance is given in Table 36.

A post hoc analysis using the Newman-Keuls test on the effect of the class of instance showed that the mean response time for positive instances was less the mean response time for negative instances ($\underline{p} < .05$) over all rules. For individual rules, positive instances were classified faster than negative instances on conjunction ($\underline{p} < .01$), and slower on conditional ($\underline{p} < .01$). No differences were found on inclusive disjunction or biconditional. Newman-Keuls post hoc comparisons between rules for both classes of instances indicated that the biconditional took significantly longer to classify than the other rules ($\underline{p} < .01$); no other differences between rules were obtained. A similar result was found when rules were compared for positive instance classification only. For negative instances, biconditional classification was slower than conditional ($\underline{p} < .01$), which was slower than inclusive disjunction ($\underline{p} < .05$), which in turn was slower than conjunction ($\underline{p} < .05$).

Source of variation	S.S.	D.F.	M.S.	F	P
Between subjects	404.45	79			
Rules	207.76	3	69.25	26.97	<.001
Subjects within groups	196.69	76	2.59		
Within groups	71.81	80			
Instance	1.84	1	1.84	4.00	<.05
Instance x rules	34.93	3	11.64	25.26	<.001
Instance x subjects within groups	35.04	76	0.46		
Within cell	231.73	152	1.52		

TABLE 36. Summary of analysis of variance on transformed latencies.

3. Discussion

A comparison may be drawn between these results and the predictions of the two approaches suggested earlier to describe the subject's behaviour on the task. Firstly, the information theory approach can be rejected as providing a most unsatisfactory set of predictions. Not only is the informational model unable to account for differences in time to classify the different rules, but also it is not clear why responses to positive instances were not quicker than to negative instances for inclusive disjunction and biconditional. It is obvious that Hick's Law does not fit the data. The only results which agreed with the expected outcome according to this model were the relative response times for positive and negative instances on conjunction and on conditional.

The decision model given in Figure 6 on the other hand, provides a somewhat better fit to the data. Several predictions have been supported, viz., (1) the relative difficulty as measured by response time for classifying instances for the four rules, and (2) the relative difficulty of classifying positive and negative instances, both over all rules and for each rule except inclusive disjunction. The major discrepancy between predictions and results is the greater relative difficulty of conditional negative and biconditional positive and negative classification.

Another look at the decision trees shown in Figure 6 suggests that these discrepant results may be accounted for by postulating that classification time may vary slightly for different subclasses of instances. Evidence indicating that categorization time is affected by the number of relevant attribute values in an instance has already been demonstrated for conjunction and inclusive disjunction (Nickerson and Feehrer, 1964; Nickerson, 1967). These workers found that response time for negative instances on conjunction and positive instances on inclusive disjunction varied directly with the number of attribute values present in the stimulus. They also found that response times for positive instances on conjunction (i.e. all relevant attribute values present) and for negative instances on inclusive disjunction (i.e. all relevant attribute values absent) were less than expected from extrapolation from times for negative conjunctive and positive disjunctive instances respectively. Indeed, response times and errors made on positive conjunctive instances seem more comparable with the response times and errors made on those negative conjunctive instances in which all relevant attribute values were absent. Similarly, response times and errors on negative disjunctive instances seem to closely resemble the response times and errors for positive disjunctive instances in which all relevant attribute values were present. In other words, the longest latencies appear to have occurred for those instances in which, in the case of two relevant attribute values, one attribute value is present and one absent.

Applying this observation to the decision tree model does help to explain some of the results which are discrepant when it is assumed that all decisions are of the same difficulty. For example, it provides an explanation why negative conditional and negative biconditional instances require long classification times. If a weighting factor is introduced to make allowance for the variable frequencies of each subclass of instance, there is a suggestion why a significant difference was found between positive and negative instances on conjunction, but not on inclusive disjunction. A further postulate, however, is necessary in addition to the observation taken from Nickerson's studies in order to account for the long latencies obtained for classifying positive instances on the biconditional. This is that the classification as positive instances in which both relevant attribute values are absent may be significantly more difficult than the classification of instances containing both attribute values. More experimental work is necessary to test these hypotheses.

Whatever the precise nature of the decision model applicable to this situation, the results of this experiment have largely confirmed the propositions put forward about stage one of the process comparing the feedback on an instance with the prediction for that instance by the hypothesis under test. Firstly, it has been found that the ease with which instances are classified does depend on the rule. In particular biconditional instances have been found to be the most difficult to categorise. This confirms the suggestion made in Experiment IX that the low number of instances recalled on the biconditional was an indication that more information was being processed on each trial for this rule. Secondly, it has been found that positive instances are easier to classify than negative instances. Again, this supports the hypothesis made in Experiment IX that the better recall of positive instances than negative instances was indicative that less information is processed on each positive instance than negative instance. Thirdly, there was an interaction between the rule and the type of instance classified similar to the interaction found in Experiment IX between the rule and the type of instance recalled. In this there is one result which does deviate from expectation. On the biconditional, no difference was found between positive and negative instances on classification, but positive instances were significantly better recalled than negative instances. This result may perhaps reflect the tendency noted in Experiment VIII to avoid selecting FF positive instances on biconditional, since it has already been hypothesized above that this subclass of instances are most difficult to classify.

Conclusion on process descriptions of conceptual behaviour

In the preceding and the present chapters attempts have been made to look at conceptual behaviour in terms of the processes underlying this behaviour. In these attempts two different techniques have been employed to infer such processes. In the first part of Chapter 4, the technique adopted was to try to construe the nature of the process directly from the sequence of choices and/or the verbal report made by the subject. Several examples using this technique were dismissed as unreliable or invalid, but one approach was tried which produced an account of behaviour which appeared to show some consistency with other evidence. In this approach subjects were given extensive training on the task which it was supposed would facilitate the making of inferences about strategies. Despite this, the experimenter was still left with a certain dissatisfaction that he may be imposing his own preconceptions about the process onto his classification. Possibly independent judges could be used to overcome this problem to some extent. In situations where no pretraining is provided, however, it is difficult to see that this approach is likely to yield much helpful information about the underlying process. A similar technique was described in the early part of the present chapter. This was Levine's procedure for inferring the particular hypothesis held by the subject from his responses to a set of blank trials. Once again Levine's subjects were provided with detailed instructions and pretraining. In the absence of such preliminary experience it may be wondered how successfully Levine's method could be used to identify hypotheses. Another factor detracting from the blank-trials technique is that it is possible sometimes to infer an hypothesis even when the subject is responding randomly.

Methods such as these may be useful on exploratory or preliminary investigations. A preferable technique for use on more extensive studies is modelling. In this approach several possible strategies for solving problems are devised and then an experimental situation is set up for which these different strategies make different predictions about behaviour. The particular model which predicts most accurately the subject's performance may be regarded as reflecting most closely the properties of the process underlying behaviour. This is the approach which has been utilized for the most part in the remainder of the two chapters. The usefulness of such an approach is that it depends on the formulation of testable predictions about behaviour, the end product of which may sometimes be the extension of existing knowledge of the conditions influencing behaviour.

The number of suggestions for future experiments in the present chapter bears witness to this. Whether an equal or greater contribution to the establishment of empirical laws between input and output is likely to be made without reference to process models is a matter of opinion.

CHAPTER 6

RULE LEARNING

At the beginning of this dissertation it was stated that a concept consists of two components: the relevant attribute values and the rule relating the relevant values. In the past three chapters we have been concerned with problems of attribute identification in which the subject is instructed on the particular rule involved and is required to discover the unknown attribute values which are relevant. In studies to be considered in this chapter it is the rule which is unknown and the relevant attribute values that are given. Tasks of this type have been termed rule learning (Haygood and Bourne, 1965).

Differences between rules

Experiments on rule learning have concentrated for the most part on the four basic rules, conjunction, inclusive disjunction, conditional, and biconditional. These rules when used to relate two stimulus dimensions have been found to differ significantly in difficulty for naive subjects and the order of difficulty has been generally agreed to be the same as in the previous sentence, conjunction to biconditional.

It would seem that the relative difficulty of rules is independent of whether the stimuli are geometrical designs (e.g., Haygood and Bourne, 1965) or verbal material (Di Vesta and Walls, 1969), whether the task is paced (e.g., Bourne and

Guy, 1968a) or unpaced (e.g., Bourne and Guy, 1968b) of the number of irrelevant stimulus dimensions (Haygood and Stevenson, 1967), and of the age of the subject (Bourne and O'Banion, 1970). All studies, but one by Giambra (1970), have utilized the reception procedure in which a randomized sequence of instances has been presented typically with the constraint that there be either equal numbers of positive and negative instances (e.g., Haygood and Bourne, 1965) or equal numbers of TT, TF, FT, and FF instances (e.g., Bourne, 1967). The same order of rule difficulty has been found in each case. In one experiment Bourne and Guy (1968b) presented subjects with three types of stimulus sequences - all positive, all negative, and a mixture of positive and negative. The results clearly showed that performance on all rules (conjunction, inclusive disjunction, and conditional) was best with the mixed sequence, contrary to expectation from instance contiguity considerations. Moreover, a comparison between the trials to solution data for the all positive and all negative sequences did not reveal the typical result obtained on attribute identification. Indeed, for conjunction a negative sequence seemed to lead to solution in fewer trials than a positive sequence, while for conditional a positive sequence produced slightly better performance than a negative sequence. For inclusive disjunction, there was a clear superiority in performance produced by a positive sequence over that produced by a negative sequence. This finding would seem to suggest that the type of instance, positive or negative, which includes the majority of the four stimulus classes, TT, TF, FT, and FF, leads to more rapid solution.

The plausibility of this explanation is clear; rules can only be fully differentiated when the truth value of each of the four classes is known. It is impossible to discriminate between conjunctive, inclusive disjunctive, conditional, and biconditional rules when the truth value of only TT instances is known. The result ofan experiment by Haygood and Devine (1967), however, would seem to be in conflict with this assertion. They showed that as the proportion of positive instances containing both relevant attribute values was increased for inclusive disjunctive and biconditional rules, the number of errors to criterion became less. The range of p (TT) examined by Haygood and Devine was between 0.2 and 0.8. Clearly, if p(TT) equalled 1.0, there is no way that subjects could learn a disjunctive or a biconditional rule from the stimulus series. Yet an extrapolation of Haygood and Devine's result would lead us to predict that fewer errors to criterion would be made under this condition. Moreover, for the procedure employedwhereby problem solution was concluded from a run of 16 correct categorizations, this is what might be reasonably expected. The confusion seems to be derived from the confounding of the testing sequence and the training sequence. Possibly, Haygood and Devine's finding does hold within the range tested. However, to properly substantiate this, a replication study is necessary in which test trials are not subject to the same constraints as learning trials.

Recently an experiment by Giambra (1970) has claimed an order of rule difficulty which is different from the trend of other studies. In this experiment conditional and biconditional rule learning problems were presented to one group of subjects according to the selection paradigm and another group of subjects yoked to the first group according to the reception paradigm. Giambra reported that conditional rule learning was more difficult than biconditional rule learning. It is assumed that this finding is true for both selection and reception subjects although no direct comparisons were made between the two procedures. Giambra's account of his finding was that the relative difficulty of rules varies according to the percentage of positive and negative instances encountered. No empirical support for this explanation was cited by Giambra and from the evidence discussed above it would seem unlikely. At the same time there appears to be no other alternative explanation to account for the discrepancy between Giambra's result and the results of other studies except that Giambra employed primarily the selection paradigm whereas the reception procedure was used in all other studies. In view of the inexplicable nature of these results, a replication experiment was carried out using a selection procedure to determine the order of difficulty of rules on rule learning and to relate this difficulty to the type of instances selected.

Experiment XI: Rule learning of concepts using a selection procedure.

1. Method

a. Subjects and Design. - The subjects were 12 undergraduate students randomly selected from the first year pool at the University of Adelaide. None of the subjects had taken part in a concept learning experiment before.

A randomized blocks single factorial design (Hays, 1963) was employed in which the 12 subjects were presented with four problems, each problem involving a different rule (conjunction, inclusive disjunction, conditional, and biconditional). The order of individual rules and pairs of rules were balanced over all subjects. The relevant attribute values on each problem were chosen at random from the 54 possible pairs of attribute value combinations of two attributes.

b. Stimuli. - The stimuli consisted of geometrical designs printed on 7 x 10 cm. white cards. Designs varied according to four attributes, with each attribute having three values, making 81 different patterns. The attributes and their values were colour (red, green, and black), shape (cross, circle, and square), number of figures (1, 2, and 3), and number of borders (0, 1, and 2). The 81 cards were arranged on a table before the subject in a systematic 9 x 9 array.

c. Procedure and Instructions. - The details of the procedure were explained in instructions which were given for the subject to read on a typed sheet. The nature of the stimulus display was described and a definition given of a "concept" as representing "a certain set of these cards." It was explained to the subject that his task was to discover the set of cards which made up the concept on each of four problems. For each problem, one concept, consisting of a particular combination of two characteristics of these cards, had been chosen. At the beginning of a problem, the experimenter presented the subject with the two relevant attribute values typed on a 8 x 13 cm. file card. Thus the subject had to find the rule which described the particular combination or relationship between these two given attribute values. The subject was told that the rule was not numerical, but a logical one involving the presence or absence of one or both of the given characteristics. He was also told that each of the four problems involved a different rule.

When the subject had read aloud the relevant attribute values, he was presented with an initial positive instance which was randomly selected for each problem from the subset of positive instances having both relevant attribute values. The subject then selected cards, one at a time, to each of which the experimenter said "yes" or "no", according to whether his choice was a positive or a negative instance. The subject could at any time volunteer an hypothesis of the concept, but only one hypothesis was permitted between successive card choices. When a correct hypothesis was verbalized, the problem was considered finished. The criterion of a correct verbalization was that it should accurately include all cards which were members of the concept set, and exclude cards which were not members of this set. The instructions emphasized that the subject should try to solve each problem with as few card choices as necessary, but warned him not to guess at the concept; time was not important.

2. Results

The results are summarized in Table 37 in terms of the mean and standard deviation of the number of card choices to solution for each rule. An analysis of variance was performed on this data transformed by a square-root transformation to make the variances more homogeneous. A significant effect was obtained on this analysis for conceptual rules, $F\ (3,33) = 9.47$, $p < .001$. A summary of this analysis is given in Table 38. A post hoc analysis using Newman-Keuls multiple range comparisons revealed significant differences between conjunction and inclusive

	Conceptual Rule			
	Conjunction	Inclusive disjunction	Conditional	Biconditional
Mean trials to solution	9.6	18.3	23.6	33.7
S.D. trials to solution	5.4	11.9	15.8	16.4
% negative instances selected	49.6	20.5	24.5	42.9

TABLE 37. Number and type of instances selected on rule-learning.

Source of variation	S.S.	D.F.	M.S.	F	P
Between subjects	33.63	11			
Within subjects	94.13	36			
Rules	43.56	3	14.52	9.47	$<.001$
Residual	50.57	33	1.53		
Total	127.76	47			

TABLE 38. Analysis of variance on trials to solution data transformed by a square-root transformation.

disjunction ($p < .05$), conjunction and conditional ($p < .01$), conjunction and biconditional ($p < .01$), inclusive disjunction and biconditional ($\underline{p} < .05$), and conditional and biconditional ($\underline{p} < .05$). Thus the ascending order of difficulty of rules was conjunction, inclusive disjunction, conditional, and biconditional.

The type of instances chosen were also examined and the mean percentage of negative instances selected for each rule can be seen in Table 37. These percentages may be compared with the percentage of negative instances in the stimulus population for each type of concept - conjunction 88.9%, inclusive disjunction 44.4%, conditional 22.2% and biconditional 44.4%. A sign test indicated that a significantly smaller percentage of negative instances were selected than were present in the stimulus population for conjunction ($\underline{p} < .001$), and for inclusive disjunction ($\underline{p} < .001$). Spearman rank correlations were computed for each conceptual rule between the percentage of negative instances selected and the number of card choices to solution, but no significant correlations were obtained.

Instances were also analyzed into those containing both, either one, or neither of the given attribute values. The mean percentage of each of these four classes of instances selected is given for each rule in Table 39. These percentages can be compared with the percentages of each class in the stimulus population - TT 11.1%, TF 22.2%, FT 22.2%, and FF 44.4%. A sign test showed that subjects selected a higher percentage of TT instances ($\underline{p}$ <.01) and a lower percentage of FF instances ($\underline{p}$ <.01) for each rule; however, there were no differences found on any rule between the percentage of TF and FT instances selected and present in the stimulus population. Thewe were no significant correlations between performance and the percentage of instances selected from any class.

	Truth table class			
	TT	TF	FT	FF
Conjunction	50.5	19.3	19.5	10.7
Inclusive disjunction	30.2	27.7	21.6	20.5
Conditional	24.6	22.3	24.5	28.6
Biconditional	24.2	22.4	20.5	32.9

TABLE 39. Mean percentage of instances selected from each stimulus class for each rule.

3. Discussion

The results indicate that the order of difficulty of rules on a rule-learning task using a selection procedure is conjunction, inclusive disjunction, conditional, and biconditional. This is the same order as has been found in other studies using the reception procedure. On the other hand, the order obtained by Giambra (1970) has not been replicated. Furthermore, the account advanced by him of a relationship between relative difficulty and the percentage of positive and negative instances encountered is not supported, although, on this latter point, it must be acknowledged the data is only slight. On this evidence no explanation can be found for Giambra's results. The present findings suggest that the order of difficulty of rules is unaffected by the type of procedure employed.

What kinds of explanations can therefore be proposed to account for rule differences? Some possibilities (Bourne, 1970) can be eliminated immediately. The first possibility suggests that rules are acquired by the rote learning of stimulus pattern-response category assignments. Since there are the same number of stimuli (81) and the same number of responses (2) for each rule, from this explanation we would expect no difference between rules, which is plainly false. A second suggestion is that rules may be acquired by focussing on positive instances only. In this case the fewer the number of positive instances on a rule, the sooner is that rule likely to be learned. Thus the order of rules should be conjunction (9 positive instances), inclusive disjunction (45) and biconditional (45), and conditional (63), which again does not correspond to the obtained order. A third possibility is that rules are learned by focussing on the smaller and less variable of the two categories, positive or negative. On this argument, conjunction should be the easiest rule, conditional next, and inclusive disjunction and biconditional most difficult. Once again this is not the order found. A fourth possibility comes from a suggestion advanced by Neisser and Weene (1962) that concepts are hierarchically organized into three levels, in which the operations involved for higher level concepts are based on rules present in lower level concepts. Level 1 is composed of affirmation and negation, Level 2 of conjunction, inclusive disjunction, conditional, disjunctive absence, conjunctive absence, and exclusion, and Level 3 of biconditional and exclusive disjunction. This account is unsatisfactory for the present findings insofar as it is unable to explain differences between conjunctive, inclusive disjunctive, and conditional rules. A fifth possible explanation which can also be rejected is that subjects acquire rules by a truth-table strategy (Bourne, 1970) in which instances from each stimulus class in the truth table are sampled. For this optimal strategy no rule differences are expected, which is not true for the naive subjects in this experiment.

Two interpretations seem more plausible on the evidence presented. The first suggests that a naive subject brings to the task a predisposition for solving for conjunctive concepts.

This conjunctive tendency involves three assumptions: (1) TT stimuli are positive, (2) TF and FT stimuli are both negative, and (3) FF stimuli are also negative. Because of these implicit assumptions, conjunctive rules should be the easiest to acquire. Apart from confirming this prediction, two other aspects of the present experiment lend support to this proposal. The first comes from an examination of the hypotheses made by subjects during the task. On almost all problems the first bidimensional hypothesis made involved conjunction. Secondly, in most cases subjects selected only from TT, TF, and FT stimulus classes prior to their initial hypothesis. Moreover, for all rules the proportion of TT instances selected was higher and of FF instances lower than the proportion of these types of instances in the stimulus population. This is what would be expected if subjects were using a conservative focussing strategy for which the initial positive instance (TT) was taken as a focus. If such a strategy was used, then it might be predicted that fewer conjunctive solutions should be hypothesized for inclusive disjunction where TF and FT are both positive. On this rule 7 of the 12 subjects were correct at their first attempt. On the conditional, subjects who tested FT instances prior to their first hypothesis sometimes tended to ignore one of the relevant attribute values presented at the beginning of the problem and to postulate the solution as a single attribute value affirmation.

Since TF and FT instances are chosen in the course of conservative focussing, it is to be expected that the truth values of these instances should be learned sooner than for FF instances. Accordingly, those rules for which the third assumption that FF stimuli are negative is invalid will tend to prove more difficult than those rules for which this assumption is true. Thus, conditional and biconditional rules should be more difficult than conjunction and inclusive disjunction. Inclusive disjunction is more difficult than conjunction because it contravenes the second assumption that TF and FT stimuli are negative. That is, the conjunctive rule may be proposed even without testing TF and FT instances, but the disjunctive rule is never suggested in absence of information about the truth values of these stimuli. The conditional seems to be less difficult than the biconditional for two reasons. Firstly, because his findings on the truth values of TF and FT instances eliminate conjunctive and disjunctive rules as possible solutions, the conditional subject is led to test FF instances sooner than biconditional subjects who try FF instances only after exhaustive examination of the other three classes. In the present experiment the mean number of trials selected to the first FF instance was 11.6 for the conditional and 14.4 for the biconditional. Secondly, even when an FF instance is encountered subjects on the biconditional require more trials to solution (19.3) than conditional subjects (12.0). Clearly subjects find some difficulty in accepting the possibility the TT and FF instances belong to the same category when other instances, TF and FT, represent a different category.

More work is necessary to examine this interpretation further. It may well be asked how it is that subjects display a tendency towards conjunctive solutions instead of disjunctive or conditional or biconditional solutions. One answer to this question is to say that conjunctive concepts are more familiar and more common in everyday thinking than other types of multidimensional concepts. Evidence in support of this contention is rather difficult to find. Perhaps one relevant source of information is the frequency of the connectives, "and", "or", and "if", in the Thorndike-Lorge (1944) word-count. The Lorge magazine count indicated 138,672 occurrences of "and" in 4½ million words, 14,851 occurrences of "or", and 14,506 occurrences of "if", which clearly supports the contention.

Other research may seek to confirm the use of a conservative focussing strategy inferred in this experiment. For example, if conservative focussing is less apparent when a random array of stimuli is displayed, the effect of randomizing the array may be to increase the likelihood of sampling from the FF class. The result may be a reduction in the relative difficulty of the conditional and biconditional. A second suggestion is to consider the effect of the type of initial instance on differences between rules. If subjects tend to use conservative focussing from the initial instance, different predictions may be made depending on the class of the initial instance. In the present experiment the given instancewas from the TT stimulus class. If the initial instance was drawn from the FF class, however, we would expect more FF instances to be chosen which should lead to the conditional and biconditional becoming relatively less difficult. Similarly, if an FT instance was chosen as the initial instance, inclusive disjunction and conditional should be relatively easier. If the initial instance was from the TF class, inclusive disjunction should be the rule to be most facilitated.

A second interpretation of rule differences involves a modification to Neisser and Weene's (1962) proposal that conceptual rules may be hierarchically organized into three levels. This proposal is based on the observation that complex logical rules can be broken in several component operations. At the first level is the simple affirmation or negation of a single attribute value; Level 2 contains rules formed by a single conjunction or disjunction of two attribute values; Level 3 consists of those rules which require a combination of conjunction and disjunction. Neisser and Weene assume that rules on each level of the hierarchy do not differ in difficulty, but this assumption has already been stated to be unsatisfactory. Not making any such assumption has implications which seem to fit the data much better. If we suppose that inclusive disjunction is more difficult than conjunction, and negation is more difficult than affirmation, and if we further suppose that these four rules represent the fundamental operations in which all other rules are expressed, then the difficulty of the conditional and the biconditional may be predicted by the number of fundamental operations required to produce the same classification of instances. Thus the conditional, $p \rightarrow q$, may be expressed as the disjunctive combination of two components where one component is negated, $\overline{p} \cup q$. Similarly the biconditional, $p \leftrightarrow q$, may be analyzed into two negations, two conjunctions, and one disjunction, $(p \wedge q) \cup (\overline{p} \wedge \overline{q})$. The generality of this interpretation needs to be tested for other rules.

Examination of the verbalizations of concepts by subjects in the present experiment confirms the choice of affirmation, negation, conjunction, and inclusive disjunction as fundamental operations. None of the twelve subjects when verbalizing their concepts for the conditional and biconditional rules gave their responses in the conventional "if....then" and "if and only if" forms. For the conditional a variety of verbalizations were given. The most common (3 subjects) involved a disjunctive rule in which one attribute value was negated; two subjects expressed solutions in terms of one conjunction plus two single attribute affirmations; one subject used one conjunction plus one negation; another subject gave the solution as a negated exclusion rule; several other verbalizations were also noted for individual subjects. For the biconditional the most common form (9 subjects) in which the concept was expressed was, for example, "black and two figures, or, not black and not two figures".

This interpretation of rule differences in terms of a modified structural hierarchy of rules still leaves unanswered the questions: (1) why are affirmation, negation, conjunction, and disjunction fundamental operations, and (2) why is disjunction more difficult than conjunction, and negation more difficult that affirmation. (The evidence for disjunction being more difficult than conjunction has already been presented; a suggestion that negation is more difficult than affirmation is found in Jones (1966)). The answer to both questions seems to be familiarity. That is, these rules are fundamental rules because subjects encounter them most often in their everyday experience. Further, any differences between these fundamental operations is again attributable to one of them occurring more commonly than another. Support for these assertions may be found in the Thorndike-Lorge (1944) word-count tables. A query may be raised in this respect about conditionals and biconditionals; why are they not basic operations since "if" is quite common in language? Wason (1966) has made a suggestion about the logical meaning of the connective "if ... then" which seems to indicate why not. Wason's proposal is that when a subject forms a sentence like "if p then q", the truth-table function of this sentence is not the same as that given in propositional logic in that there are three truth values - true, false, and irrelevant - instead of only two. Thus the truth function is pq true, $p\bar{q}$ false, $\bar{p}q$ irrelevant, and $\bar{p}\bar{q}$ irrelevant. Wason is not the first to consider this problem: most logic textbooks (e.g., Jeffrey, 1967) make mention of it. Nor is his proposal of a three-valued logic new: the feasibility and implications of such a system are currently being debated by a number of logicians (e.g., Woodruff, 1970). If the logical meaning of the sentence, if p then q, is in doubt, then so must it be for the biconditional sentence, p if and only if q, which is logically two conditionals, if p then q and if q then p. Extending Wason's proposal to the biconditional, we would expect the truth function to read pq true, $p\bar{q}$ false, $\bar{p}q$ false, and $\bar{p}\,\bar{q}$ irrelevant. The observation made in earlier experiments on attribute identification (see Experiment I) wherein clarification of the type example of a conditional rule was required when it was given in the instructions in the form, "if a card is green, then it must be square to be a member of the concept" and of a biconditional rule given in the form, "green cards are members if and only if they are squares" reinforces the doubt on the logical meaning of these connectives.

In an attempt to learn more on this question of the truth functions of the conditional and biconditional rules an experiment was devised in which subjects were presented with three concepts involving the same two attribute values, but three different rules. The rules were conjunction, conditional, and biconditional, and were expressed in their conventional form. The subject's task was to try to discriminate between these concepts by selecting instances from a stimulus display systematically arranged in front of them. If subjects possess the truth functions for these three rules which is given in propositional calculus, then several strategies may be used. One may be to test instances, one from each of the four stimulus classes TT, TF, FT, and FF. The appropriate rule may then be determined from the truth values assigned to each stimulus class. A more efficient strategy would be to choose an instance from each of the two stimulus classes for which the three rules predict different truth values, that is, FT and FF. On the other hand, if subjects possess the truth functions claimed by Wason for the conditional and extended for the biconditional here, then no discrimination between the three conceptual rules is possible.

Experiment XII: Discrimination between conceptual rules.

1. Method

a. Subjects. - The subjects were 12 undergraduates randomly selected from the subject pool of first-year psychology students at the University of Adelaide. None of the subjects, according to their student record cards, was currently or had been previously enrolled in a course in which instruction in logic was given. Furthermore, none of the subjects had previously taken part in a concept learning experiment.

b. Stimuli. - The stimuli were the same as those used in Experiment XI.

c. Procedure and Instructions. - The instructions explained to the subject that he was to be presented with the "necessary and sufficient" specifications of three different classes within the set of cards displayed before him. One of these classes had been chosen by the experimenter and the task of the subject was to find which one. To do this he was to choose cards, one at a time, to each of which the experimenter would indicate whether it was in the particular class chosen or not. The subject was told that the time taken was unimportant, only the number of cards required to discover the class. Only one attempt at the class would be allowed.

Following these instructions a 12.5 x 20 cm file card was presented to the subject on which were typed three sentences specifying the three classes. The sentences were:

(1) Cards which are green and square are members of the class;

(2) If a card is green, then it must be square to be a member of the class;

(3) Green cards are members of the class if and only if they are squares.

It was decided in advance by the experimenter that the relevant class would be the third or biconditional. It was felt that if subjects could successfully identify this class as the one chosen, then this constituted evidence of their possession of a truth function similar to that in propositional calculus for the three rules examined.

2. Results and Discussion

None of the 12 subjects was successful in discriminating between the three sentences as they referred to the stimulus display before them. This result is significant on a binomial test ($p < .001$). Indeed, only two subjects even began to select cards as instructed, and even they gave up because they could not see any difference between the sentences with regard to the class specified, given that these were the necessary and sufficient specifications of the class. These subjects wondered whether perhaps more information was needed on the conditional sentence to know what class cards which were not green belonged. When asked to describe the class indicated in each sentence, all subjects described the set formed by the intersection of green and square for all three sentences. That is, the only true category in the truth function for each of these sentences is pq.

It is clear then that the truth functions for "if, then" and "if and only if" connectives are not the same as those expected from propositional calculus. Rather they would seem to resemble more the truth functions suggested in Wason (1966) and in this chapter. Further examination of the nature of the truth-table function for conditional sentences is made in the next chapter.

On the present findings the conditional and the biconditional are obviously precluded as fundamental operations in the same sense as affirmation, negation, conjunction, and inclusive disjunction. For naive adult subjects these latter rules appear to have "psychological truth functions" which are equivalent to their truth functions in propositional logic. Neimark and Slotnick (1970) and Neimark (1970) have investigated the development in the understanding of these connectives. Their findings are of interest to the present discussion. They found that affirmation and negation were understood by the majority of subjects, even by American third grade children.

Conjunction likewise was understood by the majority of subjects at all ages, except third grade. The proportion of each age group correctly answering disjunctive items increased gradually with age but this connective was not correctly understood by the majority in any group except the college-age level. Other findings by Neimark, however, cast doubt on the second interpretation of rule differences considered here. For example, no difference was found between comprehension of affirmation and negation; moreover, there was no difference between conjunction, conjunctive absence, and exclusion rules. This would seem to militate against affirmation and negation being considered separately, where negation is more difficult than affirmation. At the same time disjunction appeared to be more difficult than conjunction. Furthermore, inclusive disjunction and disjunctive absence were found easier than the conditional expressed as a disjunction with one component negated. These results seem difficult to reconcile with a structural hierarchy interpretation of rule differences.

Further evidence against the structural hierarchy interpretation comes from studies of inter- and intra- rule transfer effect. It is to be expected from this interpretation that any rule differences for naive subjects are chronic, since the number of fundamental operations on each rule is unchanged. Transfer of training studies indicate, however, that these differences are only transient, disappearing when subjects are presented with a series of problems based on the same rule (Bourne, 1967, 1970). Moreover, when transfer takes place between rules, the transfer is usually positive (Bourne and Guy, 1968a; Bourne, 1970); the largest positive transfer to a biconditional rule is produced by training on the conditional. Parenthetically it should be pointed out that transfer effects and rule difficulty are confounded in Experiment XI. In view of this evidence, however, no change in the order of difficulty of rules would be expected for an experimental design eliminating these effects. Indeed the outcome of such a design should be to increase the size of differences already apparent.

It seems unlikely that changes in the relative difficulty of rules can be accounted for without reference to changes in the strategies for acquiring these rules. The conjunctive-tendency interpretation has suggested the type of strategy which gives rise to rule differences in naive subjects. Bourne (1970) has attributed the inter-rule and intra-rule transfer effects noted above to the acquisition by subjects of a simple, stimulus-coding or truth-table strategy for which no rule differences are predicted. Given two relevant attribute values, the subject can codify an indefinitely large stimulus population into four classes corresponding to the four contingencies of the truth table, i.e., TT, TF, FT, and FF.

Thus, to solve any new rule-learning problem, the subject needs only to learn the assignment of these four classes to the two response categories, positive and negative. The major difficulty in acquiring this optimal strategy seems to be to attend only to those dimensions exemplified by the relevant attribute values which are given, and to collapse over irrelevant variations in the stimuli (Bower and King, 1967). Several ways to overcome this difficulty have been suggested. For example, Bower and King found that requiring subjects to verbalize a reason for each choice before making it facilitated solution. Bourne (1967) suggested a different method, providing subjects with a sample pattern illustrating the four stimulus classes and their response assignments.

Implications for complete learning

In this chapter we have been concerned with conceptual problems in which subjects are given the relevant attribute values of the concept and required to learn the rule for combining them. In earlier chapters subjects were given the rule and required to identify the relevant attribute values. It may reasonably be asked what the findings from studies of rule learning and attribute identification contribute to an understanding of behaviour on a task in which both the rule and the relevant attribute values are unknown, that is, complete concept learning.

One fundamental question in this regard is concerned with whether the difficulty of complete learning is equal to the sum of the difficulties of rule learning and attribute identification. Very little attention seems to have been devoted to this basic question and there is a need for further research. However, according to Haygood and Bourne (1965), the difficulty of complete learning measured by the numbers of trials and errors to solution is greater than the simple sum of rule learning and attribute identification difficulties. The increased number of trials and errors may merely be an indication of a greater amount of forgetting of information on the more difficult complete-learning task. On the other hand, it may be suggestive of a qualitatively different strategy employed on complete-learning tasks compared with rule-learning and attribute-identification tasks.

What kind of approach to complete learning is appropriate? Clearly a truth-table strategy cannot be used since the relevant attribute values are unknown. Likewise a conservative focussing strategy is inappropriate unless it is known that the rule is conjunction. It would appear that the subject has open to him at least two alternatives; (1) assume a particular rule and then attempt to identify the relevant attribute values, or (2) assume a particular set of relevant attribute values and attempt to formulate the particular rule combining them. The findings from rule learning suggest that the former alternative is the more probable. Most likely subjects begin a complete learning problem by assuming a conjunctive rule and then attempt to identify the relevant attribute values for this assumed rule. If the task conditions are appropriate, a conservative focussing strategy will be employed in this endeavour. In the course of these selections, he may find (when a nonconjunctive solution is applicable) exceptions to his assumed classification rule. From these findings he may modify his strategy to determine the relevant attribute values. These modifications may occur several times before solution is attained. This descriptive outline for a complete-learning strategy makes the assumption that subjects know how many relevant dimensions there are in a concept. If this information is also not presented to the subject at the beginning at a problem, further possibilities arise concerning the size of the set of relevant attribute values assumed by the subject to be in the concept.

Hunt, Marin, and Stone (1966) describe several versions of a model for complete learning. This model differs from the preliminary model outlined above in that the rule and the relevant attribute values are acquired in parallel instead of sequentially. A basic version (CLS - 1) of Hunt et al.'s model is composed of the following steps:

(1) A search is made for some characteristic or set of characteristics which appear in the description of all positive instances and never appear in the description of any negative instance. The characteristic or set of characteristics so found constitute the concept.

(2) If the procedure in step (1) cannot be carried out, it is reversed. A search is made for a set of characteristics which appear in all negative and no positive instances. If such a set is found, then the concept is the negative aspect of this set.

(3) If steps (1) and (2) fail, note which characteristic appears most frequently in the positive instances encountered. Then split the available stimuli into two subsamples, one containing this characteristic, the other not containing the characteristic. Then, treating these two subsamples as two subproblems, proceed as before in step (1), etc..

No claim is made by Hunt et al. that their model is intended as a representation of the sequence of decisions which human subjects make during complete learning. Clearly it makes assumptions about memory capacity which are quite unrealistic to attribute to human performance. Later versions, however, consider the effect of a system with limited memory capacity. A comparison made by Hunt et al. of the relative difficulty of conjunction, inclusive disjunction, conditional, exclusive disjunction, and biconditional predicted by his model and obtained empirically for human subjects indicated the same order of rules in each case.

Other possibilities have also been considered by Neisser and Weene (1962). One approach which they were able to eliminate was rote learning. Clearly a great deal more work must be done on complete learning before an adequate understanding of behaviour on this task is possible. Research is needed on the influence on performance of a number of task variables. Some work has been done on rule differences (Neisser and Weene, 1962; Haygood and Bourne, 1965; Hunt, Marin, and Stone, 1966) and evidence is also available on the effects of other parameters like positive and negative instances (Hovland, 1952) and instance contiguity (Haygood, Sandlin, Yoder, and Dodd, 1969), but information is lacking on interaction effects between variables. Furthermore, a study should be undertaken to explore transfer effects on complete learning, particularly inter-rule transfer. To answer questions on the relationship between complete learning on the one hand, and rule learning and attribute identification on the other, it would also be useful to investigate transfer from training on tasks involving one or both of the latter to a task involving the former.

<u>Implications for other problem-solving tasks</u>

One of the outstanding features of research which has come to be included under the rubric of "problem-solving" is the wide variety of tasks and techniques that have been devised. A recent review of some of these procedures is given in Bourne and Battig (1966). According to Duncan (1959), "this diversity is a major reason why the area of problem solving seems so chaotic, and is a serious obstacle to systematic progress." It is suggested here that some order may be achieved out of this "chaos" by regarding concept learning as continuous with problem solving, such that many of the laws underlying behaviour derived in experiments involving the former may also be true in the latter. Thus, many problem-solving studies have been concerned with the effect of set (e.g., Luchins, 1946; Saugstad and Raaheim, 1960) which can be compared with the work on the salience of stimuli in concept learning referred to in Chapter 2. Other studies have examined variables such as amount of stimulus variability (e.g., Brush, 1956), concreteness or abstractness (e.g., Lorge, Tuckman, Aikman, Spiegel, and Moss, 1955a, 1955b), spatial contiguity of problem elements (e.g., Kay, 1954), etc., all of which have been examined in concept learning.

It is further suggested that a number of problem-solving studies may be understood in terms of the rules which are necessary for combining the relevant elements of a problem.

In some cases these rules may be numerical, but in others logical relationships similar to those investigated in this and preceding chapters may be identified. One obvious example of this is an experiment described by Moore and Anderson (1954). In their task the subject was given certain premises from which he had to reach a specified conclusion through the use of given transformational rules. Another example comes from the work of Dienes and Jeeves (1965, 1970) on the solving of various embodiments of mathematical groups. The application of logical rules to problem solving will be illustrated by a consideration of the latter example cited.

A mathematical group "is a set of elements over which a binary operation is defined, which is closed, in the sense that applying the operation to any two elements, another element is obtained" (Jeeves, 1968). The mathematical group with two elements, say *a* and *b*, has the rule structure given in Figure 7. According to Dienes and Jeeves, "multiplying" by *a* leaves the

		State	
		a	b
Play	a	a	b
	b	b	a

FIGURE 7. A two element group.

"multiplicand" unchanged, and "multiplying" by *b* changes the "multiplicand" from *a* to *b*, or from *b* to *a*. In mathematical group terminology the element which leaves things unchanged is called the neutral element; the element which produces the change is called the alternator. There is an alternative way of viewing thus rule structure, however. This is to consider the two-group as equivalent to a biconditional rule in which the outcome a is produced whenever the same element is present on the two variables, state and play, and the outcome b whenever the elements for these two variables differ. Thus we have two stimulus dimensions, state and play, on each of which there are two values, a and b, and there are two responses, a and b, which are fully determined by the particular combination of stimuli chosen. If S represents the state and P the play, then the rule producing the outcome a is given by $S_a \leftrightarrow P_a$. On this reasoning we might expect some similarities between the behaviour exhibited by Dienes and Jeeves's subjects and subjects on a biconditional concept learning problem.

Dienes and Jeeves were not interested simply in the behaviour of subjects in solving the two-group per se, but in the effects of transfer to more complex mathematical groups formed from the two-group. We need consider here only those groups with four elements. Two such groups can be conceived: the cyclic or modulo 4 group in which there is one two-group subgroup, and the Klein 4 group consisting of two two-group subgroups. The rule structures for these groups having four elements are shown in Figure 8. In the abstract it is difficult to formulate logical rules for these complex groups to predict the outcome of different combinations of state and play.

1. Cyclic group

Play \ State	a	b	c	d
a	a	b	c	d
b	b	a	d	c
c	c	d	b	a
d	d	c	a	b

2. Klein group

Play \ State	a	b	c	d
a	a	b	c	d
b	b	a	d	c
c	c	d	a	b
d	d	c	b	a

FIGURE 8. Four element groups

This difficulty is due to the existence of four possible outcomes instead of the conventional two in conceptual studies. However, when the different elements are represented by symbols which suggest the subgroups embedded within the group, these rules can be identified more easily. For example, consider a four group where <u>a</u> is represented by a yellow triangle, <u>b</u> by a yellow circle, <u>c</u> by a blue triangle, and <u>d</u> by a blue circle. For the Klein group two biconditionals may be identified, one involving colours, the other involving shapes. Thus a yellow outcome is produced when the same colour is present on both variables, i.e., *S* (Y) $\leftrightarrow$ *P* (Y); similarly, a triangular shape is the outcome whenever the same shape is evident on both variables, i.e., *S* (△) $\leftrightarrow$ *P* (△). The finding that the use of such symbols facilitated performance on this task therefore is not surprising. No comparable improvement was observed with the cyclic group, for which the biconditional provides only a partial description of the solution to the group. A counting rule gives a more adequate account. Thus, *a* leaves the state unchanged, *b* changes the state by one position, *c* by two positions, and *d* by three positions. This form of solution resembles the operations conceived by mathematical group theory. It is understandable that for the cyclic group subjects performed best who utilized an "operator" strategy, where one variable was held constant whilst the effects of changing the other were observed.

On this argument performance on the Klein 4 group should show some similarities to the acquisition of a four-category concept involving two biconditional rules. A task of this nature may be expected to prove relatively difficult for at least two reasons: (1) because the biconditional has been found to be a comparatively difficult concept to learn, and (2) because the problem necessitates the learning of two such concepts concurrently. With regard to the latter, Restle and Emmerich (1966) have shown that this requirement places a greater load on memory and consequently produces a decrement in performance on each concept. Offsetting these effects is the fact that there are no irrelevant stimuli, that is, subjects have only to learn the correct response to each combination of stimuli. Moreover, the subject is informed of the true outcome as well as the correctness of his response on each trial.

The procedure employed by Dienes and Jeeves for presenting instances was somewhat unusual in two ways. Firstly, their task was so arranged that one of the variables, the state, could not be controlled by the subject unless he knew the rules that he was supposed to be learning. Only the second variable, the play, could be manipulated by the subject. Thus, the method of presentation of instances was neither completely selection nor reception, but actually a mixture of both. The relative difficulty of this type of mixed procedure in comparison with the conventional selection and reception procedures has not been investigated in concept learning. An argument could be made, however, that such a procedure where the subject has only partial control over incoming information approximates more closely to the conditions of learning in everyday situations. It is suggested therefore that selection and reception procedures should be regarded as end-points of a continuum representing the degree of control that a subject has over the sequence of instances encountered. An experiment to investigate concept identification as a function of this variable needs to be done. A second feature of Dienes and Jeeves's procedure was that before the outcome of a particular combination of state and play appeared on the display, the state disappeared. Thus an additional requirement of their task as presented was for subjects to remember the value of one of the two variables combined. Clearly such a requirement must increase the difficulty of the problem.

The particular apparatus utilized by Dienes and Jeeves (1970), whilst possessing the capability of a wide range of mathematical groups of different complexity, is comparatively restricted in the methods which can be employed in presenting the problem. The nature of their display is illustrated in Figure 9. To examine some of the questions raised in the preceding discussion an apparatus was developed which was more versatile with respect to the variety of methods of presentation available. With this apparatus it was possible to present problems with up to four elements using any one of four procedures - complete selection, state selection and play reception, state reception and play selection, and complete reception. Furthermore this apparatus enabled subjects to observe the combination of state and play and outcome together, unlike Dienes and Jeeves's machine where the same display panel was used to indicate the state and

FIGURE 9. Display of apparatus used by Dienes and Jeeves (1970).

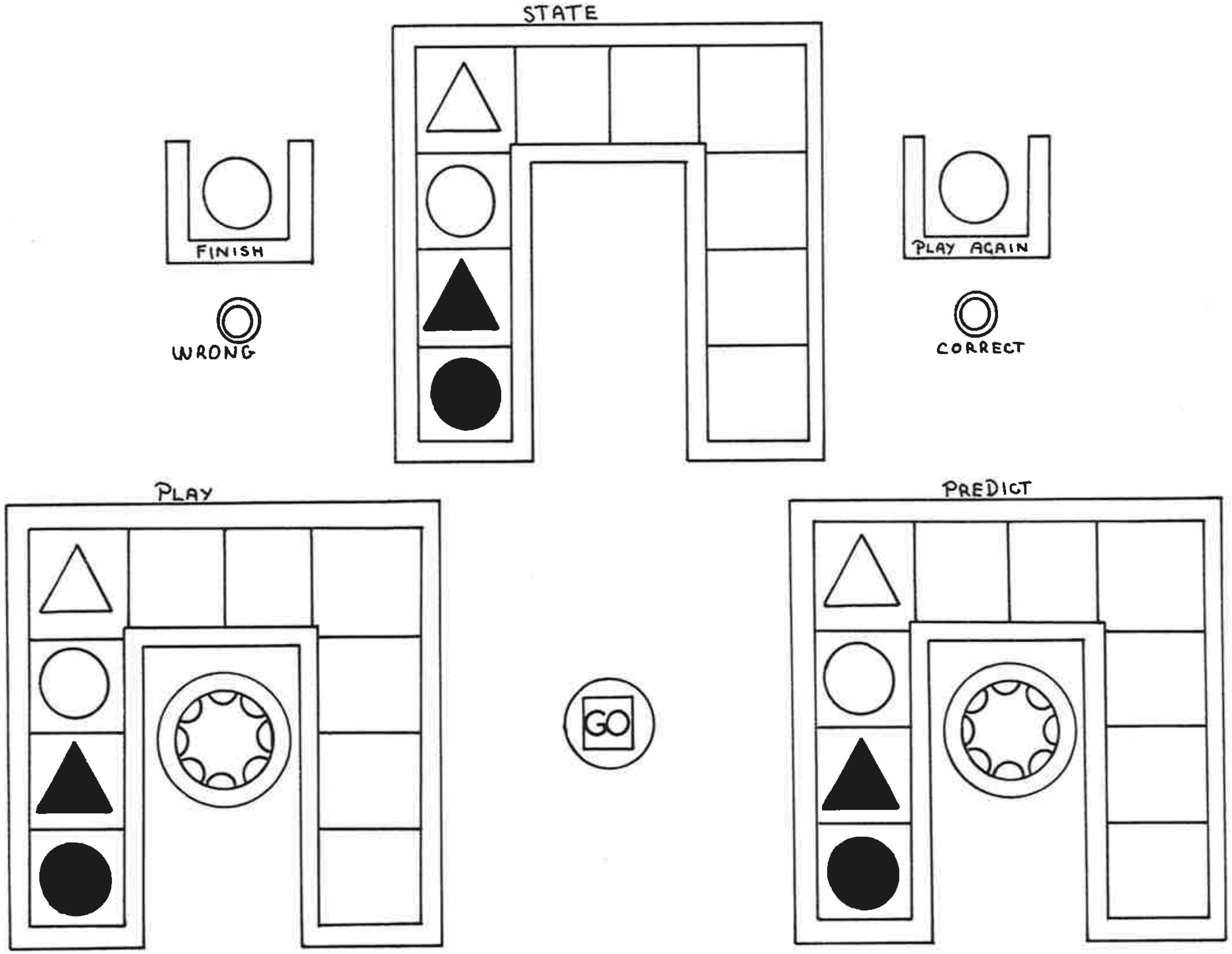
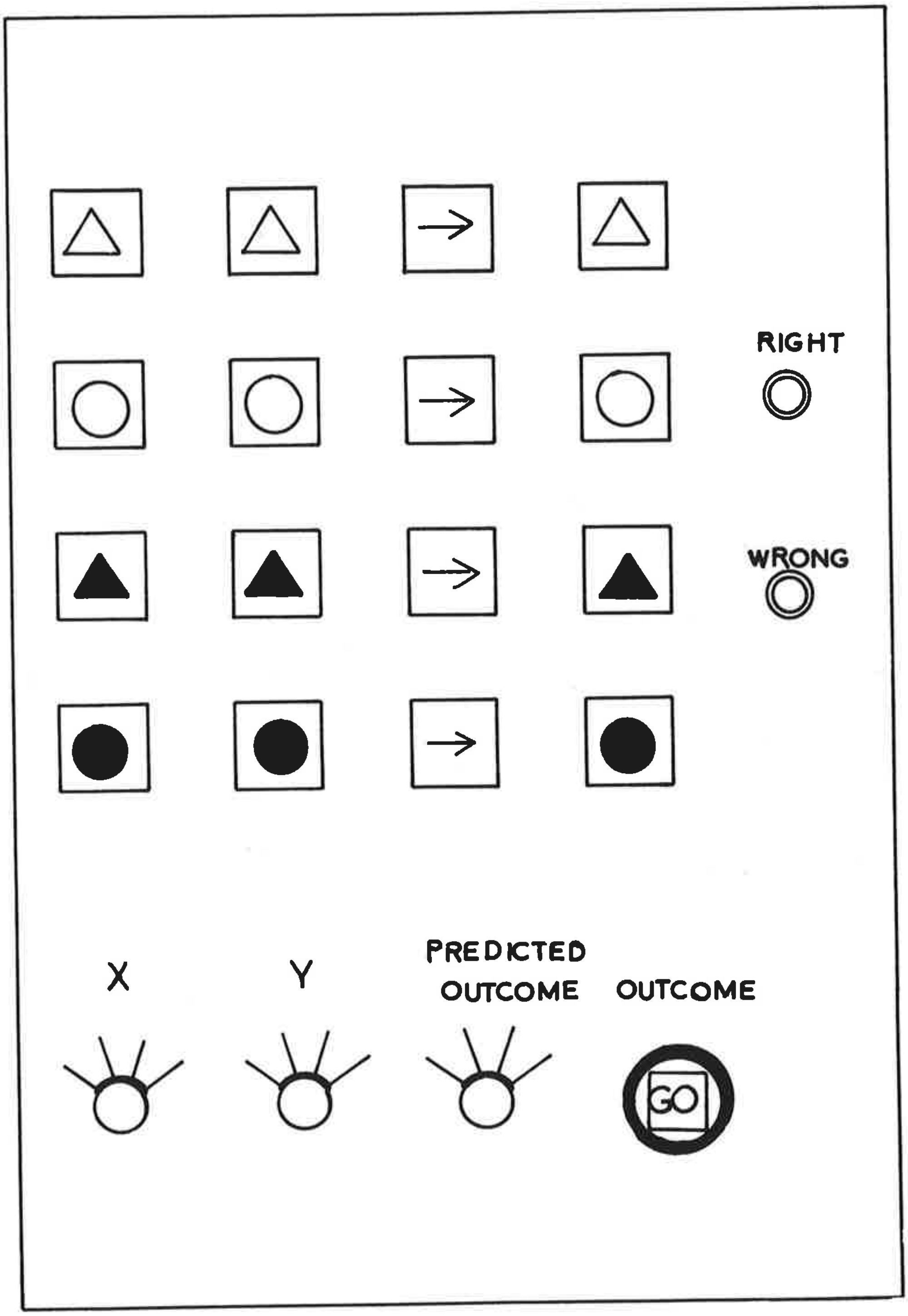

STATE
FINISH
WRONG
PLAY AGAIN
CORRECT
PLAY
GO
PREDICT

FIGURE 10. Display of apparatus employed in Experiments XIII and XIV


RIGHT
WRONG
X
Y
PREDICTED OUTCOME
OUTCOME
GO

the outcome. As in the apparatus developed by Dienes and Jeeves subjects could be presented both with the true outcome and with feedback on their prediction of the outcome. Alternatively the prediction panel could be blocked so that the subject passively observed the outcome to each combination of state and play, or the outcome panel could be omitted so that the subject received information only on whether his prediction was right or wrong. A diagram of the full display panel is shown in Figure 10.

Several exploratory experiments were run using the new apparatus. One experiment sought to examine the effect of providing subjects with preliminary experience in identifying concepts involving the biconditional on subsequent performance on the task of solving the Klein 4 group. Symbols were used to suggest the two subgroups or biconditionals in the group. If subjects do solve this group as two biconditionals, then the effect of pretraining on this rule should facilitate their performance relative to untrained subjects.

Experiment XIII: Effect of preliminary training on the biconditional on the solving of a Klein 4 mathematical group.[1]

1. Method.

a. Subjects. - The subjects were 20 undergraduate students

---

[1] The assistance of Mr. Peter Young in the running of subjects for this experiment is gratefully acknowledged.

from the University of Adelaide. Ten of these subjects had previously taken part in Experiment III in which they were given extensive practice (25 problems) at identifying biconditional concepts. It was assumed that participation in this earlier experiment constituted sufficient pretraining on the biconditional. The remaining 10 subjects had not had any previous experience in handling the biconditional and were taken as the untrained group.

b. Stimuli. - The stimulus display was the same as that illustrated in Figure 10.

c. Apparatus and Procedure. - The apparatus was the one described above and the display panel was identical to that shown in Figure 10. The complete selection procedure was employed. Thus on each trial the subject could choose a particular symbol for the state and for the play which were labelled "X" and "Y" respectively. He then made his prediction of the outcome of this combination, after which he activated the "go" button whereupon the true outcome of the combination lit up as did one of two lights labelled "right" and "wrong", indicating the correctness of the prediction.

d. Instructions. - The instructions were read to the subject who was seated in front of the machine. The subject was told that he was about to play a game against the machine. Basically there were three columns of symbols, two labelled "X" and "Y" which could be selected by turning knobs (demonstrate) and one labelled "outcome" which was produced by the combination of X and Y. When the "go" switch was pressed, one of the four outcome lights would come on as a result of the two symbols which were lit up for X and Y. It was explained to the subject that he should play various combinations of X and Y and see what the respective answers or outcomes were. When he knew all the answers the problem was solved. So that some idea could be gained of the degree to which the subject had learned the task, he was asked to make a prediction of the outcome by lighting up one of the arrows in the "predicted outcome" column so that it pointed to the answer he thought would be obtained, before operating the go switch. Solution of the problem was ascertained when the subject was able to predict correctly for any particular pair of symbols.

2. Results and Discussion

The results showed that the median number of trials to solution was 25.5 for the group given pretraining on the biconditional and 38 for the untrained subjects. Clearly the difference between the medians is in the predicted direction, but a Mann Whitney U test failed to show that this difference was significant at the 5% level. The failure to find a significant difference seems to be due to the poor performance of two subjects in the pretrained group who required 78 and

89 trials to solution. Possibly a significant result may have been obtained if more subjects were run.

Apart from any statistical reasons, the pretrained group may not have been superior to the untrained group as expected because of the type of preliminary training they received. For the Klein group in this experiment there were no irrelevant attributes. The subject's task therefore consisted of learning the appropriate outcome to each combination of the attributes of shape and of colour, a task which appears more akin to a rule-learning situation than the attribute identification problems on which preliminary training was given. A subject who has acquired the truth-table strategy described earlier in this chapter should require a maximum of 8 trials to solve the group. In other words, more marked transfer may have occurred if pretraining had been on rule learning.

Another experiment sought to assess the effect on performance of the partial selection procedure of Dienes and Jeeves in comparison to the complete selection procedure for the present apparatus. It will be recalled that two differences exist between the two techniques for presenting the task. Firstly, subjects are not permitted on the Dienes and Jeeves apparatus to control both the state and the play, only the play. Secondly, subjects on the Dienes and Jeeves machine cannot observe simultaneously the state, the play and the outcome, unlike subjects on the present apparatus. It was argued previously that these differences, especially the latter, should lead to poorer performance in Dienes and Jeeves' experiments than on the machine described here.

Experiment XIV: Effect of method of presenting instances on the solving of a Klein 4 mathematical group.[1]

1. Method.

a. Subjects. - The subjects were 40 undergraduate students from the first-year psychology subject pool at the University of Adelaide. No subject had previously taken part in an experiment on concept learning.

b. Design. - Half the subjects were required to solve the Klein 4 group using the apparatus described in this chapter. The other subjects were presented with the same problem but on the machine used by Dienes and Jeeves (1970).

c. Stimuli. - The stimuli for each apparatus were the same as those presented in the previous experiment.

d. Instructions and Procedure. - The instructions to subjects were similar for both groups of subjects. They began by explaining the operation of the machine, and then defined

---

[1] The assistance of Mr. Peter Young in the running of subjects for this experiment is gratefully acknowledged. The author also wishes to thank Professor M. A. Jeeves for permission to use his apparatus.

the object of the task. The details of the instructions were similar to those given in the previous experiment. Problem solution was ascertained when the subject could predict correctly for any pair of symbols in the state and play. The task was concluded after 120 trials if solution had not occurred before this time.

2. Results and Discussion

The results are summarized in Table 40 in terms of the means and standard deviations of the number of trials to solution for the two groups of subjects. The data for the Dienes and Jeeves machine group include those of three subjects who had not solved the problem by 120 trials. A unrelated samples $\underline{t}$ test was used to compare the performance of the two machines. This showed that subjects on Taplin's machine required significantly fewer trials to solution than subjects on the Dienes and Jeeves machine, $\underline{t}\ (38) = 4.69,\ \underline{p} < .001$.

	Dienes and Jeeves' apparatus	Taplin's apparatus
Mean	72.8	37.7
S.D.	25.9	19.9

TABLE 40. Number of trials to solution of a Klein 4 group using two different procedures for presenting the problem.

This finding indicates that the solving of a mathematical group is a function of the method by which the problem is presented. The present experiment, however, does not enable us to say specifically whether the improvement in performance is due to either one or both of the two procedural differences noted in the introduction. Further experiments using the Taplin apparatus and comparing the complete selection procedure with a partial selection procedure where the subject can observe simultaneously the state, the play, and the outcome should permit the role of each procedural difference to be determined.

Other pilot studies have been carried out on the apparatus devised. For example, the task was made a little more difficult by allowing the subject feedback only on whether his prediction was right or wrong; no information was provided on the true outcome in the event of a wrong prediction. As expected, subjects required more trials to solution; for 12 subjects the median number of trials to solution was 55. It was noticeable that some subjects under these conditions commonly did not attempt to find the correct outcome after an incorrect prediction had been made. This may be indicative of an hypothesis-testing approach to the problem where failure of a prediction results in rejection of the hypothesis and the sampling of another consistent with the existing information. Where the subject did seek the correct outcome, he employed a direct test of every possible outcome until a correct prediction had been made. This occurred even when 3 of the 4 predicted outcomes had been found to be wrong. In general, subjects displayed a tendency to change one variable at a time, except in those cases where they were examining the outcome when X and Y had the same values, or where they were testing whether X and Y were commutative. From the comments made during the task it was evident that the problem was solved as a series of sub-problems, each of which had its own rule for solution. For example, a number of subjects discovered early in the task that when X and Y had the same symbol the outcome was a yellow triangle. They did not infer from this that the same colour yields yellow, and the same shape produces triangle. Instead they went on trying combinations to learn that a different shape and a different colour on X and Y gave a blue circle, and so on for the other outcomes. The failure to see the connection between the various sub-problems was further seen by the difficulty that subjects had in retaining rules which they had previously learned.

Clearly a more systematic program of research needs to be pursued in order to investigate adequately the approach to problem solving advocated here. What it is hoped has been demonstrated is the potential of this approach in relating some of the apparently unrelated studies in this area to the work on concept learning. At the same time the attempt to account for problem solving in this way may raise questions for investigation in concept learning which have hitherto not been considered.

CHAPTER 7

LOGIC AND REASONING

Inductive and deductive reasoning

In previous chapters we have been concerned with the acquisition of concepts by a process of inductive reasoning. An inductive argument is one which aims at establishing a true conclusion on the basis of true premises, where the content of the conclusion goes beyond the content of the premises. That is to say, we have been examining the attainment of a concept from evidence in the form of instances, where the concept is supposed to apply not only to the particular instances selected but also to all the other instances in the stimulus population.

The conclusion of an inductive argument is acceptable if (1) the premises are true, (2) the argument has a correct form, and (3) the premises of the argument embody all available relevant evidence (Salmon, 1963). Examples of the effects of the failure to fulfil the first and third conditions on concept learning can be readily seen in the experiments described here. Thus, the role of memory may be understood in terms of the first condition. Concepts are more easily identified when the instances (premises) are correctly retained. Confusion in either the particular attribute values or the truth value of the instance may lead to a wrong hypothesis of the concept being proposed. The importance of the third condition is apparent in at least two ways. Firstly, an incorrect conclusion may be due to a failure to sample sufficient instances in the stimulus population. As a consequence there may be more than one hypothesis which can logically describe the instances chosen. Selection of one of these hypotheses as the concept therefore carries a certain probability or risk of error. Secondly, an incorrect conclusion may be due to a bias in the sampling of instances in the stimulus population. An obvious example of this is seen in the work on rule learning. Unless instances are selected from all four stimulus classes, TT, TF, FT, and FF, there is some likelihood that the hypothesis which the subject proposes on this task will be wrong.

Little, however, can be said on the basis of the present experiments about varying the form of the argument. On the surface it would seem that all inferences should be of the universal enumerative induction type: all observed A's are B's; therefore all A's are B's. The instructions were partly intended to cause subjects to assume this form of inductive reasoning. However, other forms are also possible. For example, there may be a tendency arising from the fact that very few universal empirical truths are encountered in everyday thinking, for subjects to reason in the form of a prevalence generalization. A prevalence generalization expresses the inference, A's are B's, with such qualifiers as "generally", "usually", "probably", "tendency", "mostly", and so on. The likelihood that this type of reasoning is involved, particularly during the early stages of the learning process, can be shown by an illustration. Consider a subject set the task of identifying the conjunctive concept, red and square. Initial selections made by the subject may inform him that red cards are sometimes members of the concept. Further selections then indicate that square cards are also sometimes members of the concept. When these two prevalence observations have been made, the subject is then in the position of being able to combine them to form the universal generalization that all red and square cards are members of the concept. More complex illustrations can be similarly conceived for other conceptual rules. Discrimination between universal and prevalence inductions, even from subjective reports, is likely to prove difficult since both forms of generalization may sometimes be expressed without any explicit qualifier (Ennis, 1969). Another type of inductive argument, probabilistic generalization, is also possible but does not seem generally applicable to the types of concepts investigated here. A probabilistic generalization expresses the inference, A's are

B's, as a probability m/n. This form of reasoning seems more appropriate in tasks involving statistical or probabilistic concepts. Once again, however, prevalence generalization is probably involved, at least in the early stages, of such tasks. Some studies of probabilistic concept attainment have been performed (e.g., Bruner, Goodnow, and Austin, 1956; Mandler, Cowan, and Gold, 1964; Bourne, 1963), but no attempt will be made to review them here.

Another sort of reasoning, involved perhaps to a lesser extent in concept learning yet nonetheless important in thinking, is deduction. In deduction, unlike induction, the conclusion is supposed to follow necessarily from what is given. There is no intention to extrapolate beyond the content of the premises. Hence for a deductive argument the conclusion is valid if (1) the premises are true, and (2) the argument has a correct form (Salmon, 1963). In contrast to an inductive argument the validity of a deductive argument is unaffected by any additional evidence. A logically correct inductive argument may have true premises and a false conclusion; a logically correct deductive argument with true premises must have a true conclusion.

One instance of deductive reasoning considered in the foregoing investigation of concept learning is to be found in the chapter on hypothesis testing. Here subjects were viewed as evaluating an hypothesis by the testing of a prediction made by that hypothesis about a particular stimulus. If the hypothesis is true, then the prediction about the state of the stimulus is true, represents the form of reasoning. Finding that the prediction is true or false has certain implications for the truth of the hypothesis. In particular, if the prediction is false, then the hypothesis is also false. On the other hand, if the prediction is true, the hypothesis thereby achieves a greater degree of credibility. If and when subjects exhibit behaviour of this sort, they are employing a form of deductive reasoning known as conditional reasoning.

In this chapter we shall examine the nature of human deductive reasoning and some of the variables which can affect the logicality of this process. To begin, however, it is necessary to describe the two major types of deductive reasoning (Ennis, 1969).

Types of deductive reasoning

1. Class reasoning

Before discussing class reasoning, we must first explain what is meant by a categorical statement or proposition. A categorical statement contains two terms, a subject term and a predicate term. Each term stands for a class of things. This, "A's are B's" is a categorical statement. The content of a categorical statement depends on the classes referred to by the subject and predicate terms. Four forms of categorical statements can be produced, and these have been traditionally denoted by one of the first four vowels. Type A is the universal affirmative, and is of the form, "all A's are B's"; type E is the universal negative which has the form, "no A's are B's"; type I is the particular affirmative and reads, "some A's are B's"; and type O, the particular negative, is expressed "some A's are not B's". Categorical propositions then may be universal or particular, and affirmative or negative. It should be realized that there is more than one way of expressing each statement. The above forms are merely those adopted conventionally in discussing each type; a list of equivalent expressions of A, E, I, and O propositions may be found in Salmon (1963).

We may note here that the converse of a categorical proposition is not necessarily true. By converse is meant the interchanging of the subject and the predicate. The converse of E is necessarily true: if no A's are B's, then obviously no B's are A's. The converse of I is also necessarily true. But the converses of A and O propositions are false. A simple method of finding the truth or falsity of the converse is to use the Euler's circles technique. This approach will not be described here, but can be found in many elementary logic texts (Ennis, 1969).

Class or categorical reasoning occurs in arguments composed entirely of categorical statements. Apart from the converse, the simplest categorical argument is the categorical syllogism. A syllogism is an argument with two premises and a conclusion. In a categorical syllogism the premises and the conclusion are each composed of statements connecting two terms of classes, yet the whole syllogism has only three different terms. One of these terms occurs in both premises and is called the "middle term"; the other two terms, "end terms", occur in only one premise and in the conclusion. Thus, a general illustration of a categorical syllogism is the following: A's are B's; B's are C's; therefore A's are C's. In this case the conclusion is valid. However, other arguments may also be considered where the conclusion does not logically follow: for example, A's are C's; B's are C's; therefore, A's are B's. Thus, many forms of syllogisms can be conceived, some of which are valid, others invalid. The validity of a syllogism depends on two things; firstly, which of the four types of categorical statements are involved, and secondly, the positions of the middle term and the end terms. Two systems have been developed by logicians to assist in determining the validity of a syllogism. The first technique is Euler's circles and a discussion of this can be found in many introductory logic texts (e.g., Ennis, 1969). The second is a system of three rules based on a concept of distribution (Salmon, 1963). A term is said to be distributed in a categorical statement if that statement says something about each and every member of the class that term designates. It is outside the scope of this discussion to explain further the meaning of this concept or how it is used to derive the rules for testing the validity of a syllogism. Let us simply assume that the subject term of a universal statement is distributed, the predicate term of a negative statement is distributed, and all other terms are undistributed. Then the three rules for a valid syllogism are: (1) the middle term must be distributed exactly once; (2) no end term may be distributed only once; and (3) the number of negative premises must equal the number of negative conclusions. Any syllogism which violates one or more of these rules is invalid.

More complex types of class reasoning may also be found. Such arguments involve the drawing of a conclusion from more than two premises. In many cases these arguments may proceed in a series of steps where each step constitutes a syllogism. Except for the first step, one of the premises in the syllogism is the conclusion from the previous syllogism. Of course, such complicated arguments need not be linear but can involve branching, for example, when the conclusions of two separate arguments are used as the premises for a third argument.

2. <u>Sentence reasoning</u>

This type of reasoning differs from class reasoning in that the basic units are not classes but sentences. The important factor in sentence reasoning is the connective linking the sentence units. Several types of connectives may be utilized. If p and q are taken to represent the sentence units, examples of connectives include conjunction, p and q, disjunction, either p or q, conditional, if p then q, biconditional, p if and only if q, conjunctive absence, not both p and q, and so on. In short, they are the same logical operators described in Chapter 1.

Perhaps the most important kind of sentence reasoning is conditional reasoning. Before discussing the various forms of conditional arguments, let us examine firstly the nature of conditional sentences. Consider the conditional sentence, if p then q. In this sentence p is called the antecedent and q the consequent. The content of the sentence is determined by p and q, but the form of the sentence is due to the conditional connective which places p and q in definite relation to one another. There are many ways of expressing conditionals besides "if p, then q". Alternative forms include "q, if p", "never p without q", "unless q, $\overline{p}$", "either $\overline{p}$ or q", and "all p's are q". The lastmentioned form, it will be realized, is actually the universal affirmative. One other alternative expression has been given a special name because of its fundamental relationship to the standard form; this is the contrapositive, "if $\overline{q}$, then $\overline{p}$." For the conditional, the converse, if q, then p, is logically invalid.

Four basic forms of conditional arguments exist. Two of these are valid and the other two invalid. In each case there are two premises and a conclusion, where one of the premises is a conditional statement. The first type of argument is called "affirming the antecedent" or "modus ponens", and has the following form: if p, then q; p; therefore, q. This is a valid argument form. The second type of argument is "affirming the consequent" and is given in the following: if p, then q; q; therefore, p. This form of argument is logically invalid. The third argument form is "denying the antecedent": if p, then q; $\overline{p}$; therefore, $\overline{q}$. Again the conclusion is logically incorrect. The fourth argument form is called "denying the consequent" or "modus tollens": if p, then q; $\overline{q}$; therefore, $\overline{p}$. In this case, the conclusion does necessarily follow, and is therefore valid. Further discussion of these arguments can be found in any logic textbook (e.g., Ennis, 1969).

<u>Logic: a model of human reasoning</u>

The preceding section provides an outline of deductive reasoning according to a logician. The question arises from this as to whether logic is an accurate representation or

---

The expression, $\overline{p}$, is equivalent to not p.

description of human reasoning.

It is clear that many early logicians regarded logic as the science of the laws of thought, as "deducible from a consideration of the operations of the mind in reasoning" (e.g., Boole, 1854; J. S. Mill, 1874; Kant, 1885). More recent logicians, however, have rejected this view. For example, Russell (1904) contended that there is no necessary relationship between human thinking and logic, that "throughout logic and mathematics, the existence of the human or any other mind is totally irrelevant". Schiller (1930) maintained that syllogistic reasoning "has nothing whatever to do with actual reasoning, and can make nothing of it".

Some psychologists have also tended to place little emphasis on the role of logic in human thinking. For example, Bruner et al. (1956) suggested that "much of human reasoning is supported by a kind of thematic process rather than by an abstract logic. The principal feature of this thematic process is its pragmatic rather than its logical structure". Similar suggestions have also been made by Morgan and Morton (1944) and Lefford (1946). Other psychologists, however, have chosen to accept logic as describing the fundamental operations of thought. Thus, Hunt (1962) and Haygood and Bourne (1965) used the propositional calculus to define the rules which combine stimuli to form concepts. Likewise, Piaget (Piaget, 1957; Inhelder and Piaget, 1958) assumed that the various connectives between two sentences have the same truth functions in thought as they do in logic.

The view to be taken here is that propositional logic should be regarded as a model of thought. With respect to deductive reasoning an important feature of such a model is the clear predictions it makes about responses to problems. Comparisons between these predictions and the actual responses or products of human reasoning may then enable us to suggest modifications to the model which approximate more closely to human functioning. To this end therefore we shall examine some of the experiments which have investigated the logicality of human deductive reasoning.

Variables affecting deductive reasoning.

1. Concrete versus abstract material

Wilkins (1928) presented adult subjects with categorical syllogisms using materials which differed in the degree of abstractness or concreteness. Four types of materials for reasoning were tested: syllogisms with (1) familiar and concrete terms, (2) symbolic terms in the form of letters, (3) unfamiliar terms in the form of long words, and (4) familiar and concrete terms where the logical conclusion was contrary to everyday experience. Twenty items involving each type of material were randomly presented. Each item consisted of two premises and three different conclusions. The subject's task was to indicate which, if any, of the conclusions necessarily followed from the given premises.

The results showed that Type 1 material was the least difficult. More difficult were syllogisms presented in Type 4 material. Most difficult were the syllogisms involving symbolic and unfamiliar (long words) material, that is, Types 2 and 3 respectively. It is plausible to consider Types 1 and 4 as involving the more concrete terms and Types 2 and 3 as the more abstract terms. Wilkins' finding, therefore, indicates that categorical reasoning is better with concrete terms than with abstract terms. In general, subjects who did well on one type of material also did well on the other types. Nevertheless, there were some subjects for whom a change in material did produce a significant change in performance in relation to other subjects. Analyzing the difficulty of particular items it was evident that all three rules for a valid syllogism were disobeyed by subjects. The most common fallacy involved the first rule which says that the middle term must be distributed exactly once. Most subjects, however, detected the fallacy due to violation of the third rule which says that the number of negative premises must equal the number of negative conclusions. Typically these fallacies were more common in abstract than in concrete material.

Morgan and Morton (1944) compared the use of abstract symbols and concrete words ("troubles," "unpleasant", and "insults") in syllogistic reasoning. Thirty-two forms of categorical syllogisms were presented using each type of material and on each syllogism the subject had to choose between five alternative conclusions - A, E, I, O, and "None of the conclusions seems to follow logically". The results indicated that on 18 out of the 32 syllogisms different conclusions were accepted for abstract and concrete terms. In contrast to Wilkins' finding, however, performance with concrete material did not appear to be logically superior to that with symbolic material.

There seems to be some doubt, therefore, that concrete material necessarily leads to logically better performance compared with abstract material. At the same time the type of conclusion deduced for the two types of material does appear to differ, particularly for some forms of syllogisms. Two factors at least seem likely therefore to influence performance: (1) the form of the syllogism, and (2) the nature of the concrete material. The second factor is included because Morgan and Morton utilized one set of concrete terms; perhaps another set of concrete terms may yield a different result.

2. Attitude or belief held by subject

In addition to her findings on the effects of concreteness and abstractness on the validity of reasoning, Wilkins' data also suggested that fallacious conclusions were more likely to be accepted when they were compatible with the experience of the subject and also that valid conclusions were less likely to be accepted when they were inconsistent with the subject's experience. Her result, however, was not decisive for two reasons. The first concerns the ambiguity of the notion of the "suggestiveness" of the conclusion. The second is that only two-thirds of Type 4 items were designed to be suggestive since it was feared that if all items were of this type then subjects might learn the basis on which conclusions were either true or false. Later experiments have overcome these deficiencies and clearly demonstrate that subjects are likely to be satisfied with a fallacious argument if they agree with the conclusion to which the argument leads, and conversely, that they are less likely to accept a valid argument if they disagree with the conclusion.

Janis and Frick (1943) tested two hypotheses derived from the above proposition: (1) there are more errors in judgments of the logical validity of syllogisms on "agree-invalid" items than on "agree-valid" items; and (2) there are more errors in judgments of the logical validity of syllogisms on "disagree-valid" items than on "disagree-invalid" items. Subjects were presented with 16 concrete categorical syllogisms, 8 of which were valid and 8 invalid. Syllogisms were selected whose conclusion were anticipated to produce either general agreement or general disagreement, with a minimum of neutral attitude responses. After the syllogism test was completed, an attitude test was given in which the conclusions of all the syllogisms were presented in random order and the subject was asked whether he agreed or disagreed with them. The results confirmed the two hypotheses.

Morgan and Morton (1944) investigated syllogistic reasoning when the premises and the conclusions concerned vital issues under discussion in the mass media. Most of the problem material was related to aspects of the war on which it was expected subjects held strong emotional prejudices. This experiment employed the same procedure on each syllogism as was described in the previous section for Morgan and Morton (1944). Fifteen syllogisms were presented and each syllogism was examined using symbolic material (letters) and the material for which subjects were assumed to have strong convictions, wishes, or fears. The results showed a clear bias in responding towards conclusions favouring the anticipated personal prejudices or hopes of the subject.

Gorden (1953) conducted a similar type of experiment to that performed by Morgan and Morton to investigate the effect of attitude toward Russia on syllogistic reasoning. In this study, however, attitude was measured on a Likert scale. In the reasoning test 24 items were presented, 12 of which were on Russia, the other 12 being items of a "non-social and non-controversial nature". For each Russian item, two of the five conclusions were pro-Russian, two were anti-Russian, and one read "None of the above conclusions are necessarily logical". The last-mentioned conclusion was the logically appropriate one. A bias effect toward the attitude indicated on the Likert scale was obtained but this effect was not as marked as was found by Morgan and Morton. This was interpreted to mean that the definition of the task as a test of logical ability had reduced the influence of personal feeling, even when the subject had a definite attitude to Russia. In other words, attitude did bias syllogistic reasoning, but there was some evidence that subjects would accept conclusions which appeared logical even though they were not consonant with their personal opinion.

The studies described thus far have investigated the effect of attitude on categorical syllogistic reasoning. One study by Lefford (1946) presented subjects with a reasoning test involving categorical and conditional arguments; most arguments were syllogisms but the test also included conversions of the universal affirmative. Forty items were presented in two groups of 20 each; one group of 20 contained material of a controversial nature on which subjects were likely to hold some opinion, the other contained material of a neutral nature.

The subjects were required firstly to judge the validity of the item, i.e., "to indicate whether they thought that the conclusion could adequately be drawn from the premises given in support of them"; and secondly, to state whether they agreed or disagreed with the conclusion. The neutral items were equated with emotional items for structure and length. Lefford's results showed that neutral items were judged more often logically correctly than emotional items. For emotional items evaluation was influenced in the direction of the attitude or belief or extraneous knowledge held by the subject about the truth or falsity of the conclusion.

Another study by Thistlewaite (1950) employed conditional syllogisms of the form; if p then q; if p then r; therefore if q then r. He presented 72 arguments similar to the above consisting of 36 with relatively neutral content and 36 with statements and conclusions on Negroes, Jews, women, and patriotism, on which highly ethnocentric subjects should have strong attitudes and beliefs. Subjects therefore were sampled from regions in which ethnocentric attitudes and beliefs were common and from regions in which such attitudes were rare. The results indicated that distortions from logical reasoning were greater for subjects from ethnocentric regions on items dealing with ethnocentrism than from subjects from anti-ethnocentric areas.

Recent experiments which confirm the trend of the results described above include Thouless (1959), Feather (1964), and Suppes (1965). Kaufmann and Goldstein (1967) have pointed to several methodological deficiencies present in some of these earlier studies. For example, some studies (Janis and Frick, 1943; Thouless, 1959) did not employ a control where neutral syllogisms were evaluated for validity. Moreover, other studies (Morgan and Morton, 1944; Lefford, 1946; Thistlewaite, 1950) did not include in the test emotional statements with which subjects might disagree. Even more importantly, in most experiments the subject's attitude toward the conclusion was tested before or after the syllogism test. It is possible that this procedure may have led to commitment for the second task such that, either the attitude expressed affected subsequent evaluation of the syllogism, or the conclusion derived from the syllogism influenced the attitude towards the conclusion.

Kaufmann and Goldstein (1967) performed an experiment designed to control for these confounding effects. They presented 32 female subjects with a syllogism test containing 36 syllogisms. Twenty-four of the conclusions to these syllogisms had been previously judged by 112 other subjects to be statements with which they agreed or disagreed; the remaining 12 conclusions were "presumably neutral". Conclusions were further divided according to degree of quantification and validity so that a 3 x 2 x 2 analysis of variance design was employed. Analysis of the number of errors made on each condition revealed significant main effects for all three factors, affective loading, quantification, and validity, and also significant interaction effects between factors. Interpretation of this analysis suggested that more errors were made in reasoning with emotional material, but this applied only to acceptance of invalid conclusions and not to the rejection of valid conclusions, and also mainly on positively affected universal conclusions.

Implications for logic as a model of reasoning

Whilst the evidence bearing on concreteness and abstractness of material as a variable affecting deductive reasoning is inconclusive, there appears to be considerable evidence suggesting that reasoning is in part influenced by the attitude maintained by the subject towards the conclusion of an argument. This evidence has been supposed by some workers to indicate that reasoning is not adequately represented by a logical model. Indeed Morgan and Morton (1944) contend that "the only circumstance under which we can be relatively sure that the inferences of a person will be logical is when they lead to a conclusion which he has already accepted."

Henle (1962), on the other hand, has argued that the presence of such errors in reasoning is not sufficient to refute logic as a model. To support her argument she attempts to conceive of a process which obeys the rules of logic which can account for the errors obtained in studies such as those reported in the previous section. Firstly, she suggests the errors in dealing with presented material may be due to a failure to accept the logical task, that is, to distinguish between a conclusion which is logically valid and one which is factually correct or one with which the subject agrees. This possibility, it will be recalled, was also entertained by Gorden (1953). Secondly, errors may be due to the restatement of a premise or conclusion so that the intended meaning is changed. For example, Henle reports a case where a subject modified two particular premises to universals, before making an evaluation of the conclusion. In such cases, therefore, the validity of the argument should be judged in relation to the syllogism actually employed, not the one intended. Thirdly, errors may be due to the omission of a premise, and fourthly, to the intrusion of additional premises. Henle suggests that when the intended premises are replaced by the actual propositions utilized, the judgements of the subject may be found to be logically correct. In other words, errors may be accounted for not in terms of a failure of the logical model, but rather in terms of changes in the material from which reasoning proceeds.

Consistent with Henle's position are some recent approaches to the formation of beliefs and attitudes (Jones and Gerard, 1967; Bem, 1970). Here the attempt has been made to characterize higher-order beliefs in terms of syllogisms where the premises are composed of lower-order or more primitive beliefs. Underlying this attempt has been the assumption that individuals do not merely subscribe to random collections of beliefs but rather they maintain coherent systems of beliefs which are internally consistent.

The above discussion seems to account for the effect of beliefs and attitudes in producing errors in deductive reasoning. It does not, however, explain why errors should occur when, the material used is abstract, for example, in the form of letters. In this situation some other process not relevant according to a model based on propositional logic seems to be operating.

The atmosphere effect

In attempting to account for erroneous reasoning with symbolic material, Woodworth and Sells (1935) took as a starting-point the observation that some logical fallacies are more common than others. For example, a syllogism with an undistributed middle term is more difficult than one with a universal conclusion from particular premises. From this observation they formulated three hypotheses. One of these has become especially associated with their work and is called the "atmosphere effect" hypothesis.

The concept of "atmosphere" refers to the global impression conveyed by a categorical statement. The atmosphere of a premise can be universal or particular, and affirmative or negative. Whatever it is, Woodworth and Sells postulated that it creates a sense of validity for the corresponding conclusion. Thus, "an affirmative atmosphere in the premises makes it easy to accept an affirmative conclusion", and so on. This hypothesis correctly predicts, therefore, that it is easy to accept the converse of a proposition. It also correctly predicts that a universal conclusion tends to be accepted fallaciously to follow from two universal premises. Similar predictions for syllogisms with two particular premises, two affirmative premises, and two negative premises are also confirmed. Two supplementary hypotheses were proposed for the case when the two premises in a syllogism differ in their predicted atmosphere. These were: (1) a combination of a universal and a particular premise produces a particular atmosphere; and (2) a combination of an affirmative premise and a negative premise creates a negative atmosphere.

The atmosphere hypothesis was tested in an experiment by Sells (1936). Items were presented consisting of the 16 possible paired combinations of the four types of premises, A, E, I, and O. Letters were used as material for the premises to avoid the influence of factual knowledge and attitudes. To each of the 16 premise pairs was added one of four possible conclusions - A, E, I, or O. There were two replications of the 64 different syllogisms in the test. The subject was required to evaluate the syllogism as "absolutely true", "probably true", "indeterminate", or "absolutely false". The atmosphere hypothesis correctly predicted the conclusion most commonly accepted for 13 out of the 16 sets of premises. The three exceptions were: AA where A was predicted yet I was most frequently accepted; AE where E was predicted but O was the most frequently accepted conclusion; and EA where E was favoured by the hypothesis but O was obtained in the experiment. These findings were considered to support the hypothesis of an atmosphere effect.

An even better fit to the data was reported by Sells when an additional hypothesis was assumed. This hypothesis has two parts. Firstly, it postulates that subjects are more cautious in accepting a universal conclusion than a particular conclusion. Thus, I conclusions should be more readily accepted than A conclusions, and O conclusions more readily than E conclusions. For Sells' data this prediction was confirmed on all 16 sets of premises for I - A, and on all but one of the 16 sets for O - E. Secondly, it was suggested that subjects are more cautious in accepting an affirmative conclusion than a negative conclusion. Thus, E conclusions should be more readily accepted than A conclusions, and O more than I. There were four exceptions in the 16 premise pairs to this prediction for both E - A and O - I.

Chapman and Chapman (1959) have criticised the procedure used by Sells to test the atmosphere and caution hypotheses. On Sells' test format they argue it is not surprising that more I conclusions were accepted than A conclusions or that more O conclusions were accepted than E conclusions. The substance of their argument is as follows; If an A conclusion is accepted, then, logically, an I conclusion must also be accepted, since the particular implies "at least some" without excluding the possibility of "all". Similarly, if an E conclusion is accepted, then so must an O. On the other hand, acceptance of an I or an O conclusion does not necessarily imply the logical acceptance of an A or an E conclusion respectively. Therefore the frequency of acceptance of I and O should never be less and should probably be more than the frequency of acceptance of A and E, assuming the subject is consistent in his responding. This situation, Chapman and Chapman point out, is a consequence of presenting a single conclusion on each item.

Adopting a multiple-choice format where subjects must choose from among 5 possible conclusions - A, E, I, O, and "none of these" - overcomes this objection. Using this procedure, they obtained results which they claimed did not substantiate Woodworth and Sells' hypotheses. Detailed comparison between their data and the predictions made by these hypotheses, however, suggests that Chapman and Chapman were rather hasty in rejecting these proposals. In their experiment the conclusion most frequently accepted by subjects was correctly predicted by the atmosphere hypothesis in 12 out of the 14 types of syllogisms examined. The exceptions were the IE and OE items where the preferred conclusion for both was E, but the predicted conclusion O. These two exceptions, however, do not constitute sufficient evidence to eliminate an atmosphere account of categorical syllogistic reasoning. A better approach would be to ask why the exceptions to this account occur. Considering a combination of atmosphere and caution hypotheses for these two syllogisms, it is found that the predicted relative frequencies of acceptance of I > A, O > E, O > I, and E > A are all confirmed except O > E. The addition of a caution hypothesis therefore does seem to provide, at least, a partially satisfactory account of reasoning on the IE and OE items. Evaluation of this hypothesis on other items is rather difficult since very few conclusions other than the one predicted by the atmosphere hypothesis and the logically valid "none of these", were accepted. However, if we take only those items where there was more than one conclusion with a frequency of acceptance greater than 10%, that is, EO, OI, IO, AO, and OO, the four predictions made by the caution hypothesis are confirmed in each case. On this analysis Woodworth and Sells' account of categorical syllogistic reasoning appears to be generally supported rather than refuted by Chapman and Chapmans' data.

More recent studies by Simpson and Johnson (1966) and Begg and Denny (1969) are also consistent with the atmosphere hypothesis. Both of these studies note the similarity between the predictions of error by the atmosphere hypothesis and by the account suggested by Chapman and Chapman in terms of illogical conversion and probabilistic inference. Illogical conversion refers to the acceptance of the converse of either A or O premises, while probabilistic inference is the acceptance by subjects of invalid conclusions which have some probability of being correct. It would appear that what Chapman and Chapman have done is not to reject the atmosphere hypothesis, but to present an alternative account of the same set of predictions. The merits of the two explanations, however, cannot be assessed by considering the end-products of reasoning but by examining the process involved in arriving at these products. In either case, nevertheless, the account requires a modificationto a logical model of deductive reasoning.

Meaning of quantifiers and connectives

1. "Some"

In addition to the atmosphere and caution hypotheses described in the previous section, Woodworth and Sells (1935) also proposed a third explanation of erroneous reasoning with abstract material. This proposal was that the meaning of the quantifier, "some", is ambiguous. In propositional logic this quantifier is taken to mean literally "at least some". In language, however, "some" may have a different denotation, namely, "some but not all". Wilkins (1928) likewise noted that subjects seemed to regard it as invalid to draw a particular conclusion when the premises were universal. To offset any ambiguity, subjects in the experiments by Sells (1936) and Chapman and Chapman (1959) were instructed that "some" meant "at least some".

The effect of the two usages of "some" on reasoning may be seen by an illustration. Consider the syllogism: all A's are B's; some C's are A's; therefore some C's are B's. In logic this is a valid argument. In logic "some" means "at least some". If, however, "some" means "some but not all", then this is an invalid argument because it might be the case that all C's are B's. Thus, the two meanings of the quantifier lead to two different predictions about the validity of this argument. Similar effects can also be shown in other arguments. Thus, a subject's performance may vary from the expectations of the logical model if he assumes the quantifier to denote "some but not all".

Johnson-Laird (1969a, 1969b) has suggested that "some" tends to be given the meaning of "some but not all" when it appears at the beginning of the sentence in the grammatical subject, whereas the "at least some" interpretation tends to be given when it occurs at the end of the sentence in the grammatical object. In one experiment inferences from eight doubly-quantified sentences were judged by subjects as correct or incorrect. Each inference involved an active and its corresponding passive sentence, e.g.: "Some medicine cures every disease; therefore, every disease is cured by some medicine". Inferences were from active to passive, and passive to active. As predicted, an inference from a premise with "some" in the grammatical subject to a conclusion with "some" in the grammatical object tended to be judged as valid, but an inference in the converse direction tended to be judged as invalid.

2. "If...then..."

In the previous chapter the suggestion was made that the logical meaning of the connective, if...then..., is not the same as that given in propositional logic. Rather the truth-table function of the sentence, if p then q, appeared to be pq true, $p\bar{q}$ false, $\bar{p}q$ irrelevant, and $\overline{pq}$ irrelevant. This truth function for a conditional sentence is the same as that put forward by Wason (1966).

In an experiment in which subjects were set the task of selecting instances of p, $\bar{p}$, q, and $\bar{q}$, necessary to determine whether a given conditional sentence was true or false, Wason (1968b) has obtained evidence which supports the suggested truth function for the pq, $p\bar{q}$, and $\bar{p}\bar{q}$ contingencies, but the majority of his subjects evaluated the $\bar{p}q$ contingency as false. Few subjects made the contrapositive inference, that is, that $\bar{p}$ follows from $\bar{q}$, when trying to show that the conditional sentence was true or false. This latter finding was not so evident when the conditional sentence was expressed in the form, "either $\bar{p}$ or q" (Wason and Johnson-Laird, 1969), but no difference was found for "$\bar{p}$ if $\bar{q}$", "never p without q" (Johnson-Laird and Tagart, 1969), "all p's are q's" (Johnson-Laird and Wason, 1970a) and "every p is a q" (Wason, 1969). Wason (1968) postulated that the infrequency of the contrapositive inference is due to a set towards truth or correspondence between sentences and state of affairs. The necessity of this postulate is not clear, and a more parsimonious account would be that, if the truth value of the $\bar{p}\bar{q}$ contingency for a conditional sentence is irrelevant, then the contrapositive inference does not convey any information about the truth of the conditional sentence, and thus is also irrelevant.

In studies using children as subjects Matalon (1962) and Peel (1967) suggested an alternative truth function for the conditional to that proposed by Wason. Matalon conducted two experiments which differed in the nature of the relationship between the antecedent and the consequent in the conditional sentence. In the first experiment the connection was arbitrary, involving the switching on of two different coloured lights; this was referred to as "material implication". In the second experiment, a village setting was displayed in front of the subject. The design of the village was such that, if a man went to the post office, then he passed a certain house. In this case there was a clear connection between the actions of the man in the antecedent and the consequent and this was referred to as "natural implication". Peel investigated children's understanding of the conditional by a game technique. Each game was played between the experimenter and the child according to a given rule, which connected the responses of the child to the responses of the experimenter. Both Matalon and Peel found that the conditional tends to be interpreted as a biconditional where the truth function reads pq true, $p\bar{q}$ false, $\bar{p}q$ false, and $\overline{pq}$ true.

Matalon and Peel's conclusion is consistent with the finding in Experiment XII that subjects could not discriminate between conditional and biconditional sentences. In this experiment, however, subjects were also unable to discriminate between conjunction and the conditional and biconditional. It seems therefore that these results conform more closely with the truth function suggested by Wason. Clearly the difference between Wason, on the one hand, and Matalon and Peel, on the other, centres on the truth value assigned to the $\overline{pq}$ category. Both, however, agree that the meaning of the conditional, if p then q, is not that predicted by propositional calculus.

Two things may be responsible for the differences between the truth functions of Wason, and Matalon and Peel. Firstly, Wason used adult subjects whereas Matalon and Peel used children. Therefore, as a result of development the meaning of the conditional may undergo some modification from one form to the other. Secondly, there is some difference in the nature of the task in the two sets of experiments. In Wason's studies the subject was required to select instances in order to determine whether the conditional was true or false. Essentially the same requirement was involved in Experiment XII. In the studies of Matalon and Peel, however, the conditional was given to be true and the subject had to evaluate the validity of a conclusion derived from the conditional. Wason (personal communication; Wason and Johnson-Laird, 1970; Johnson-Laird and Wason, 1970b) has suggested that a distinction needs to be made between selection and evaluation tasks and perhaps the truth function inferred concerning a conditional sentence depends on which task is used.

From this it seems desirable to perform an experiment to examine the nature of the truth function for conditional sentences when adult subjects are given the task of evaluating various conditional arguments. Four types of conditional arguments were presented - affirming the antecedent, affirming the consequent, denying the antecedent, and denying the consequent. The truth function for the conditional derived by Wason (1968b) would predict that affirming the antecedent and affirming the consequent are regarded as valid arguments, while denying the antecedent and denying the consequent are invalid. Matalon (1962) and Peel's (1967) truth function would predict that valid responses are given on all four arguments. A logical model, on the other hand, would predict that affirming the antecedent and denying the consequent are judged as valid arguments, while denying the antecedent and affirming the consequent are invalid.

From the evaluations made by a subject of the validity of these four arguments it is possible to infer whether or not the subject was responding in a truth-functional manner and, if so, what truth function he was employing. Inferences about the truth function are based on the assumption that, if it is known that a given conditional sentence is true, and the truth value of either the antecedent or the consequent is also known, then the truth valuesfor the given conditional sentence may be derived from judgments of the validity of a conclusion involving the consequent or antecedent respectively. For affirming the antecedent, a "valid" response would suggest that the subject has a pq truth-table category which is true, and a $p\bar{q}$ category which is false; an invalid response, on the other hand, might indicate either pq is false or $p\bar{q}$ is true or both. For denying the antecedent, an "invalid" response might mean that both $\bar{p}q$ and $\bar{p}\bar{q}$ have the same truth value, either true or false, or, that $\bar{p}\bar{q}$ is false and $\bar{p}q$ is true; a "valid" response, however, would indicate that $\bar{p}\bar{q}$ is true and $\bar{p}q$ is false. For affirming the consequent, an "invalid" response would suggest that pq and $\bar{p}q$ have the same truth value, be it true or false, or, that pq is false and $\bar{p}q$ is true; a "valid" response, on the other hand, would mean that pq is true and $\bar{p}q$ is false. For denying the consequent, a "valid" response would indicate that $p\bar{q}$ is false and $\bar{p}\bar{q}$ is true; however, an "invalid" response might suggest that either $p\bar{q}$ is true or $\bar{p}\bar{q}$ is false or both. A summary of these inferences can be found in Table 41.

It can be seen that two inferences are drawn about the truth values of each truth-table contingency. Thus, if it is the case that conditional sentences are considered in a truth-functional manner, then we can expect significant correlations in performance between at least one and possibly both of affirming the antecedent and affirming the consequent, and affirming the antecedent and denying the consequent. For the same reason, significant correlations may also be expected between one or both of denying the antecedent and affirming the consequent, and denying the antecedent and denying the consequent.

Argument	Response	
	Valid	Invalid
Affirming the antecedent	pq(T) and $p\overline{q}$(F)	pq(F) or $p\overline{q}$(T) or both
Denying the antecedent	$\overline{p}q$(F) and $\overline{pq}$(T)	either $\overline{p}q$(T/F) and $\overline{pq}$(T/F), or pq(T) and $\overline{pq}$ (F)
Affirming the consequent	pq(T) and $\overline{p}q$(F)	either pq(T/F) and $\overline{p}q$ (T/F) or pq(F) and $\overline{p}q$ (T)
Denying the consequent	$p\overline{q}$(F) and $\overline{pq}$(T)	$p\overline{q}$(T) or $\overline{pq}$ (F) or both

TABLE 41. Inferences to be drawn about truth-table values for a conditional sentence, if p then q, from responses made to four deductive arguments.

Experiment XV: Reasoning with conditional sentences

1. Method

a. Subjects. - The subjects were 43 undergraduates from the subject pool of first-year psychology students at the University of Adelaide. Selection of subjects was random with the one exception that none of them, according to their student record cards, was currently or had previously been enrolled in a course in which instruction in logic was given.

b. Problems. - The problems were of four basic types. In the first type, affirming the antecedent, the subject was given: (1) if p then q, and (2) p, and was asked whether q necessarily followed. In the second type, denying the antecedent, the subject was given: (1) if p then q, and (2) $\overline{p}$, and was asked whether $\overline{q}$ necessarily followed. In the third type, affirming the consequent, the subject was given: (1) if p then q, and (2) q, and was asked whether p necessarily followed. In the fourth type, denying the consequent, the subject was given: (1) if p then q, and (2) $\overline{q}$, and was asked whether $\overline{p}$ necessarily followed. The subject gave his answer to each problem by checking one of "yes" or "no".

Verbal material was used to form the sentences. An attempt was made to eliminate any sentences whose conclusions were obviously plausible or implausible or such that individual subjects may have strong prejudices about truth or falsity.

Each type of problem was replicated four times to allow for any residual effects due to these factors and also for any variations in the nature of the relationship between antecedent and consequent. Thus, 16 problems were presented to each subject. The order of problems was randomized with the aid of a latin square so that each problem type occurred once in each block of 4 problems. The complete list of problems presented may be seen in Appendix A.

c. Procedure and Instructions. - The procedure was given in the instructions on the front of the test booklet and read in part as follows:

> "Inside this cover are 16 sets of three statements. Each statement within a set is labelled a, b, or c. In this test you are to assume that statements a and b are known to be true. The question you are required to answer is whether statement c necessarily follows from statements a and b, assuming that these two statements (a and b) are known to be true. Indicate your answer by crossing out either 'yes' or 'no', whichever is incorrect, at the end of each question."

When the subject finished reading the instructions, he was allowed to begin. Subjects were reminded that they should answer all questions, and that they could work at their own pace since there was no time limit.

2. Results

The results are summarized in Table 42 in terms of the frequency of the number of affirmative responses to each type of problem or argument. These results were first analyzed to determine whether subjects were behaving in a random fashion in producing their responses. A single-sample chi-square was used to compare the distribution of the frequency of affirmative responses obtained with a binomial distribution where the probability of an affirmative response is equal to the probability of a negative response is equal to ½. Significant differences from random responding were found for each problem type ($p < .01$).

Argument	Number of affirmative responses					
	0	1	2	3	4	Total
Affirming the antecedent	0	1	1	7	34	43
Denying the antecedent	7	6	12	6	12	43
Affirming the consequent	7	14	8	9	5	43
Denying the consequent	4	11	10	10	8	43
Binomial distribution	2.69	10.75	16.13	10.75	2.69	43

TABLE 42. Frequency of affirmative responses to each argument.

In the case of problems involving affirming the antecedent, there is a clear tendency to respond affirmatively which gives rise to the significant chi-square. For the remaining arguments, however, the significant chi-square is not attributable to a skewness in the frequency distribution in a single direction. Rather it seems to be due to a flattening and spreading of the distribution such that the observed frequencies at the extremes are higher than expected for a random distribution, while the frequencies at the centre of the distribution are lower than expected. This finding would seem to be due to there being two or maybe three groups of subjects. The first group of subjects displays a consistent tendency to answer "yes", and the second group a consistent tendency to answer "no"; a third small group may also be present who respond equally "yes" or "no".

Correlations between responses to the different arguments were estimated using Spearman's rank technique. A significant correlation was obtained between denying the antecedent and affirming the consequent, rho (41) = 0.37, $\underline{p} < .01$, but no other correlations were significant at the 5% level of significance. Marginally significant was the correlation between affirming the antecedent and denying the consequent, rho (41) = 0.23, $\underline{p} < .10$.

Looking at the judgments of individual subjects, only 5 of the 43 subjects were completely consistent in their responses to all four arguments. Of these, 2 evaluated all four arguments as valid; 2 evaluated affirming the antecedent and denying the antecedent as valid, but denying the antecedent and denying the consequent as invalid; and 1 subject affirmed the conclusion for affirming the antecedent and denying the consequent, and denied the conclusions of the other two arguments.

3. Discussion

Whilst it is clear that subjects were not responding randomly in evaluating these arguments, the present results do not provide any straight-forward indication of the basis for evaluation. A considerable degree of variation in responding is evident, both between subjects and within subjects. The only exception to this is the overall consistency with which subjects evaluate affirming the antecedent as a valid argument. This would suggest that most subjects possess a truth function where pq is true, and $p\bar{q}$ is false. Assigning truth values to the remaining truth-table categories is possible with the present data only for three subjects. Two subjects apparently utilized a truth function which reads pq true, $p\bar{q}$ false, $\bar{p}q$ false, and $\overline{pq}$ true; a third subject had a truth function where pq was true, $p\bar{q}$ false, $\bar{p}q$ true, and $\overline{pq}$ true. That is, two subjects possessed a truth function identical to that of the biconditional in propositional logic, while one subject reasoned in the manner predicted for the conditional by logic.

The lack of consistent responding within subjects makes it impossible to determine a complete truth function for the remaining subjects. Even to say that they are utilizing a truth function may be "jumping to conclusions". It was predicted earlier that, if subjects do evaluate conditional arguments in a truth-functional manner, then some correlation should be found between certain pairs of arguments. The results indicate a significant correlation in one case and a suggestion of a significant correlation in a second. This is admittedly not very strong evidence that there was any logical basis to responding at all for many subjects. Just as probable is an account simply in terms of a response bias towards affirmation[1], in view of the tendency to affirm the validity of the conclusion for all arguments.

Because interpretation of these results is obscure, this experiment was replicated with three modifications. Firstly, so that subjects may be more easily identified as tending to respond affirmatively or negatively, the number of examples of each type of argument presented was increased from four to 12. From a binomial test then we can say that subjects responding either affirmatively or negatively on 9 or more

---

[1] The author is indebted to Dr. D. McNicol for this suggestion.

problems of the same type display a significant tendency to respond in that direction. Secondly, to reduce the likelihood of a simple response bias explanation, a fifth type of argument was included in the test. This type had the same form as affirming the antecedent except for a negated conclusion, thus making it logically invalid. A consistent denial of the validity of the conclusion to this argument would suggest that an explanation other than affirmative bias is more likely. Thirdly, more information on each evaluation was sought by requiring subjects to indicate not only their decision on whether the conclusion followed from the premises, but also their confidence in this decision.

Experiment XVI: Reasoning with conditional sentences - a replication and extension.

1. Method

a. Subjects. - The subjects were 56 undergraduates from the subject pool of first-year psychology students at the University of Adelaide. Selection of subjects was once again made with the provision that, according to student record cards, none was currently or had been previously enrolled in a course in which instruction in logic was given.

b. Problems. - The problems consisted of the four types employed in Experiment XV plus a further type, affirming the antecedent with a negated conclusion. Here the subject was given: (1) if p then q, and (2) p, and was asked whether $\bar{q}$ necessarily followed. Subjects gave their answer to each problem by checking one of "yes" or "no", and then showed their confidence in their answer immediately below according to one of 3 categories: very confident, moderately confident, not confident.

The majority of conditional sentences were taken from various logic textbooks. A wide range of topics was involved and an independent judge[1] was used to eliminate any sentences whose conclusions were obviously plausible or implausible. The sentences contained a mixture of instances where the two propositions (p, q) were causally related, and where the connection between the propositions was either arbitrary or ambiguous (Peel, 1967). For example, an instance of a conditional sentence in which the two propositions were causally related was: "If food is constantly supplied to them, then the very fiercest creatures live peacably together". An example of a conditional sentence where the connection between the propositions was arbitrary was: "If Mrs. Elton is a snob, then Mrs. Bates is a bore". The complete list of problems presented may be found in Appendix B.

c. Design. - Sixty problems were presented, 12 problems for

---

[1] The author expresses his gratitude to Dr. D. McNicol.

each problem type. The order of problems was randomized with the aid of a latin square so that each problem type occurred once in each block of 5 problems. All subjects did all problems in the same order at their own pace. The task took approximately 30 minutes for each subject.

d. Procedure and Instructions. - The procedure and instructions were the same as in the previous experiment. In addition to giving their answer by crossing out either "yes" or "no" at the end of each question, subjects had to indicate their confidence in their answer by checking one of the 3 categories underneath: very confident? moderately confident? not confident?

2. Results

The number of logically correct responses made by each subject over 12 problems for each problem type was calculated and a frequency distribution of this data is recorded in Table 43. A 2-way (problem type x subjects) analysis of variance was performed on this data and a significant effect found for problem type, F (4,220) = 48.84, $p < .001$. A post hoc analysis using Newman-Keuls multiple range comparisons revealed no significant differences between Problem Types 2 and 3, and between 1 and 5, with all other comparisons significant ($p < .01$). Thus, the order of problem types in terms of increasing number of correct responses was 2 and 3, 4, and 1 and 5. A summary of the analysis of variance may be found in Table 44.

Argument	Number of correct responses												
	0	1	2	3	4	5	6	7	8	9	10	11	12
1. Affirming the antecedent	0	0	0	0	1	1	0	2	1	3	2	15	31
2. Denying the antecedent	14	8	3	2	0	4	3	9	3	0	1	3	6
3. Affirming the consequent	10	11	0	3	3	4	1	3	5	3	5	4	4
4. Denying the consequent	2	3	0	3	4	6	8	2	2	1	8	7	10
5. Affirming the antecedent with negated conclusion	0	0	0	0	0	0	1	0	1	2	4	17	32
Binomial distribution	4.1				6.8	10.8	12.7	10.8	6.8	4.1			
Probability	<.05				.12	.19	.22	.19	.12	<.05			

TABLE 43. Observed frequency of correct responses for each problem type together with the expected frequency and associated probability of correct responses for a binomial distribution where the probability of a correct response is equal to the probability of an incorrect response equals ½.

Source	S.S.	D.F.	M.S.	F	P
Between subjects	528.74	55	9.61		
Within subjects	4665.21	224			
Problem Type	2194.34	4	548.58	48.84	< .001
Residual	2470.87	220	11.23		
Total	5193.95	279			

TABLE 44. Analysis of variance on number of correct responses.

The results for each problem type were analyzed to determine whether subjects were behaving in a random fashion in producing their responses. A single-sample chi-square was used to compare the distribution of the frequency of correct responses with a binomial distribution where the probability of a correct response is equal to the probability of an incorrect response is equal to ½. Significant differences were obtained between the observed distribution and the random distribution for all problem types ($p < .001$). From a binomial test, 91.1%, 17.9%, 28.6%, 48.2%, and 96.4% of the subjects displayed a consistent tendency to respond correctly, and 0%, 48.2%, 42.9%, 14.3%, and 0% of subjects consistently responded incorrectly for Problem Types 1, 2, 3, 4, and 5 respectively.

It has been predicted earlier that performance may show some correlation between certain problem types and this was assessed by a Spearman rank correlation. This analysis indicated that subjects who were good performers on Problem Type 1 were also good performers on Type 3, rho (54) = 0.27, $p < .05$, Type 4, rho (54) = 0.28, $p < .05$, and Type 5, rho (54) = 0.49, $p < .01$, and subjects who were good performers on Type 2 were good performers on Type 3, rho (54) = 0.78, $p < .01$, but poor performers on Type 4, rho (54) = - 0.62, $p < .01$.

An examination was also made of the problems used for each problem type in order to determine whether some problems were more difficult than others. A chi square test yielded no significant departures from a rectangular distribution for each type, indicating that there were no real differences between problems for number of correct responses.

The confidence that the subject had in his answer to each problem was measured by assigning the scores 1, 2, and 3 to the confidence categories, very confident, moderately confident, and not confident respectively. An analysis of variance on the confidence ratings totalled over the 12 problems in each problem type for each subject revealed a significant difference in confidence for different types,

$\underline{F}$ (4,220) = 38.59, $\underline{p} < .001$. Differences between individual types were tested by the Newman-Keuls multiple-range comparisons method. This post hoc analysis showed that subjects were least confident on problem Type 4; higher confidence ratings were obtained on Type 3 than Type 4 ($\underline{p} < .05$), Type 2 than Type 3 ($\underline{p} < .05$), and Types 1 and 5 than Type 2 ($\underline{p} < .01$). An examination was also made of the confidence ratings for different problems within each type, and a chi-square single-sample test revealed no significant difference between problems within each problem type.

The degree of correspondence between performance and confidence was investigated by a Spearman rank correlation test. It was found that the more confident the subject was, the more correct responses he made on Type 1, rho (54) = 0.38, $\underline{p} < .01$, and on Type 5, rho (54) = 0.47, $\underline{p} < .01$, but no significant correlations were obtained for other problem types.

3. Discussion

From these results it can be seen that all but 3.6% of the subjects tended to consistently reject the conclusion as invalid (i.e., respond correctly) for affirming the antecedent with a negated conclusion (Type 5). This finding enables us to tentatively disregard an explanation of these results given merely in terms of an affirmative response bias, although it should be admitted that this test is insufficient to completely eliminate this possibility from consideration on problems other than affirming the antecedent. It is also evident that most subjects were not making their judgments at random for any of the four types of arguments examined.

The results indicate that performance on affirming the antecedent is logically better than on denying the consequent which is better than on denying the antecedent and affirming the consequent. They also suggest that the distribution of scores for subjects on affirming the antecedent is uni-modal, that subjects are uniform in their judgments of the validity of this argument. This convergence is lacking, however, on denying the antecedent, affirming the consequent, and denying the consequent, where the distribution of responses seems at least bimodal, and possibly trimodal. On denying the antecedent, 48.2% of the subjects consistently affirmed the validity of this argument, while 17.9% denied its validity; on affirming the consequent, 42.9% of subjects affirmed the validity of the argument, and 28.6% denied its validity; and on denying the consequent, 48.2% of the subjects regarded this argument as valid, and 14.3% judged it as invalid. Related to this is the observation of a correlation between performance and confidence on affirming the antecedent, while no such relationship was obtained for the other arguments, probably as a result of the divergence of responding on these arguments.

This finding suggests that no truth-table function exists for conditional sentences which is common to all individuals. For example, Wason's (1968b) derived truth function of pq true, $p\bar{q}$ false, $\bar{p}q$ false, and $\bar{p}\bar{q}$ irrelevant, would predict valid responses to affirming the antecedent and affirming the consequent, and invalid responses to denying the antecedent and denying the consequent. On only one argument, affirming the antecedent, is this prediction confirmed for the majority of subjects. Nearer to describing the present results is Matalon's (1962) and Peel's (1967) suggestion of a truth function in which pq is true, $p\bar{q}$ false, $\bar{p}q$ false, and $\bar{p}\bar{q}$ true, since this function predicts valid responses on all four arguments.

It therefore appears that either a number of truth functions have been employed, or, alternatively, some other criterion was used which was not truth-functional in judging the validity of arguments. The prediction was made earlier that, if subjects were evaluating these arguments in a truth-functional manner, correlations could be expected between one or both of affirming the antecedent and affirming the consequent, and affirming the antecedent and denying the consequent. Since the predicted significant, although in some cases small, correlations have been found on three out of four of these comparisons, it seems that perhaps subjects are behaving according to some truth-function when reasoning with conditional sentences.

Examination of the results of individual subjects indicates that 37.5% of the subjects consistently produced affirmative judgments of the validity of all four arguments; 3.6% affirmed the validity of affirming the antecedent and denying the consequent, and denied the validity of denying the antecedent and affirming the consequent; 3.6% said "yes" to affirming the antecedent, and "no" to the other arguments; 3.6% gave affirmative responses to affirming the antecedent and denying the antecedent, and negative responses to the remaining arguments; while the rest (51.7%) of the subjects were not consistent in their responding on one or more of the arguments. From this analysis we may infer that only 44.7% of the subjects were behaving consistently in a truth-functional fashion when reasoning with the conditional. The last two groups of subjects listed above cannot be classified by a truth-function using the kind of inferences given in Table 41. According to the inferences summarized in this Table it would appear that 37.5% of the subjects have a truth function of pq true, $p\bar{q}$ false, $\bar{p}q$ false, and $\bar{p}\bar{q}$ true; 3.6% possess a truth function where pq is true, $p\bar{q}$ is false, $\bar{p}q$ is true, and $\bar{p}\bar{q}$ is true; and 3.6% of the subjects have a truth function which reads pq true, $p\bar{q}$ false, $\bar{p}q$ true, and $\bar{p}\bar{q}$ false. It will be realized that the most common truth function obtained is the same as the truth function given in propositional calculus for the biconditional. It will also be noted that a few subjects are in fact employing the truth function for a conditional sentence as found in the propositional calculus.

The finding that the most common truth function utilized by these adult subjects, average age 18.7 years, is the same as that found by Matalon (1962) with children between 9.25 and 11.5 years, median 10.17 years, and Peel (1967) with children between 5 + and 11 + years, raises important implications for Piaget's theory of intellectual development. According to Piaget, from the age of 11 to 12 years onward, subjects develop formal operations. By this it is meant that subjects use the logical forms of propositional calculus and understand the logical relations formalized in the propositional calculus, including the conditional. Doubts that propositional logic provides the essential structure of the final stage of logical development and that Piaget's model completely describes the difference between the formal operational period and its predecessor, the concrete operational stage, have already been expressed (Parsons, 1960), and the present results can only add further weight to these objections with respect to conditional reasoning.

It would seem that if the truth functions inferred in this experiment are to be reconciled with that derived by Wason (1968b) and suggested by Experiment XII, some account must be taken of the task. In selection tasks such as that typically employed by Wason, other processes may also be involved in addition to the logical function determined in evaluation tasks such as the present. An initial attempt to distinguish selection and evaluation processes has been made by Johnson-Laird and Wason (1970b).

The present experiment serves to show that propositional logic does not account for human reasoning with conditional sentences. It would be useful to be able to suggest ways in which the logical model may be modified to approximate more closely to human performance. The best approximation that can be offered here is that the conditional sentence, if p then q, is treated as having a truth-functional meaning the same as the biconditional, p if and only if q, in propositional logic. Unfortunately, this modification is able to describe the behaviour of only about two-fifths of the subjects. The majority of subjects cannot be classified as functioning in a logical sense. On what basis are these subjects responding? The immediate suggestion is that they were evaluating arguments in terms of the plausibility of the conclusions. Not only were attempts made to minimize the role of this factor in designing the test, however, but also no difference was found between problems either in terms of the logicality of response or confidence in this response for any argument. This finding also eliminates an effect due to the nature of the connection between propositions (Peel, 1967). Clearly some other strategy for responding must have been utilized, but the present experiment does not permit its identification. The desired modifications to the model of human reasoning cannot therefore be satisfactorily specified.

Other quantifiers and connectives

The foregoing discussion has suggested that errors are evident in deductive reasoning because the meanings of the quantifier, "some", and the connective, "if...then...", are not the same in language as they are in logic. Similar claims can be made about other quantifiers and connectives.

Johnson-Laird (1969b) refers to the ambiguity of universal negative statements. For example, consider the sentence, "All rectangles are not squares". This may mean either that there are no rectangles which are square, or that not all rectangles are squares. Again, "John does not love any child" may mean either that John loves no children or that John does not love _just_ any child.

In sentence reasoning, connectives such as "and" and "or" can have more than one meaning. The conjunction, "and", for instance, may possess either a commutative or an ordered property, depending on the semantic properties of the conjoined verbs (Fillenbaum, 1971). The disjunction, "or", can be inclusive or exclusive, so that unless the sentence specified which is tended, the subject may interpret the sentence in different ways. Moreover, the findings of Experiment XII raised doubts that the connective, "if and only if", is equivalent to the biconditional in symbolic logic. The comments of some subjects in this experiment suggested that the sentence, "Green cards are members of the class if and only if they are square" was rephrased to read "Cards are members of the class if and only if they are green and square".

Further work is necessary to investigate the semantics of these quantifiers and connectives. In addition, it is desirable to determine whether the various expressions which were suggested earlier as alternatives for the standard expressions for the basic quantifiers and connectives in logic are in fact identical in their connotations and denotations.

Comment

The problem with a logical model of reasoning emphasized in this section is that of finding linguistic expressions which are equivalent to the operations described symbolically in logic. Underlying the discussion is a proposal which bears on the controversy not only between logic and thought, but also between language and thought. Since all problems in reasoning considered hitherto have involved verbal material, it is to be expected that language may play an important role in the nature of the reasoning. Much has been written on the question of the relationship between language and thought. Some theorists have maintained that thought is dependent on language (e.g., Whorf, 1956), others have contended that language is dependent on thought (e.g., Piaget, 1926), and others that thought is language (e.g., Skinner, 1957), while Jenkins (1969) has taken the eclectic position of accepting all three points of view. What is suggested here is that, to the extent that deductive reasoning is dependent on quantifiers and connectives such as those examined in this chapter, the meaning which is assigned to these operations in language has a commensurate role in determining the outcome of the reasoning process. A similar proposal has been advanced by Clark (1969) in a study of a form of deductive reasoning not examined in this chapter, three-term series problems.

An inductive tendency

A third possible source of fallacious reasoning in deductive problems may be a tendency for subjects to make inferences inductively rather than deductively. In deduction conclusions are either true or false. Possibly lack of familiarity with propositions which fit so conveniently into one or other of these categories leads subjects to seek intermediate categories, for example, probably true and probably false. Such categories, however, exist only in induction.

This type of explanation of errors in deductive reasoning is consistent with the "caution" hypothesis of Woodworth and Sells (1935), who postulated that particular conclusions are more readily accepted than universals. It is also in agreement with the accounts given by Chapman and Chapman (1959) and Wason (1964).

Regrettably, no experiments have directly compared inductive and deductive approaches to problems. One method that might be used in some future experiments on syllogistic reasoning is to require subjects to assess the conclusion on a 5-point rating scale representing a true-false continuum ranging from absolutely true through probably true, indeterminate, probably false, to absolutely false. If subjects are reasoning deductively, then they should assign extreme values to the conclusion. On the other hand, if subjects are reasoning inductively, a more central tendency should be evident.

From Experiment XVI it seems likely that the tendency to reason either inductively or deductively may vary with the form of the syllogism. In this experiment subjects were typically vary confident in evaluating the conclusion to affirming the antecedent as true, yet with other arguments they were less confident. On these arguments, if given the opportunity, subjects may have judged the conclusion as probably true or probably false or even indeterminate. It also seems possible that this tendency may differ for concrete and abstract material, and between selection and evaluation tasks. Individual differences are also of interest, to see whether individuals consistently vary from one another in the degree to which they exhibit a deductive or inductive tendency. Clearly, there is scope here for further work.

CHAPTER 8

IMPLICATIONS FOR FUTURE RESEARCH

The starting point for this research camefrom a consideration of the types of solutions required of subjects in various problem-solving tasks. One of the difficulties with interpretation in this area is related to the question of what it is that the subject is learning. Reference has already been made in Chapter 6 to these difficulties. It occurred to the author that one approach which might eventually lead to the integration of problem-solving studies was to attempt to construct a system of rules which could by various combinations of rules produce any of the solutions to the tasks considered. In this scheme a rule was conceived of simply as any specifiable relationship between sets of entities or events. The obvious problem in this approach was to determine which rules were basic to these tasks, so in Chapter 1 a taxonomy of rules was suggested. Logical rules were then chosen for detailed investigation to begin with because they comprised the most adequately classified set of rules. As an experimental paradigm for investigating the relationship between such rules and behaviour, concept learning was chosen because it seemed in many respects to be similar to problem-solving, and moreover, an extensive amount of data was already available on the performance of subjects under varying conditions, including a few studies on the effects of rules.

These were the origins of the research reported in this thesis. If inferences about problem-solving behaviour are to be made, then this research needs to be extended in several ways. Some of these ways have already been discussed in earlier chapters. One of the ways not mentioned is the need to examine the attainment of more complex concepts formed by combinations of rules. The present experiments have been limited to concepts having two relevant attribute values linked by a single rule. In some cases this rule has been suggested to consist of more fundamental rules, for example, the biconditional, $p \leftrightarrow q$, may also be expressed as $(p \cap q) \cup (\bar{p} \cap \bar{q})$. More work, however, should be done to determine the effects of chaining of rules, to see to what extent the difficulty of the chain is a function of the difficulties of the individual links in the chain. Some experiments could be conducted, for example, to compare conceptual performance for concepts like $(p \cap q) \cap r$, $(p \cap q) \cup r$, $p \cap (q \cup r)$, etc..

A second direction for future research is to investigate the attainment and utilization of other types of concepts. For example, an approach to sequential pattern learning can now be established on the foundations laid by Simon and Kotovsky (1963) and Restle (1970a). In this area the conditional rule seems likely to be an important tool in analyzing relationships between various components in the sequence, and indeed this seems a more natural situation for this rule to be operating than the conventional nominal form of classification employed in the experiments here. Regrettably little assistance in these investigations seems likely to come from traditional studies of serial learning (Battig, 1969). Through systematic experimentation in the manner advocated here, however, it may be found that many of the variables which affect the formation of logical rules have a similar effect on the acquisition of sequential relationships. Other studies should be devoted to an examination of mathematical rules. One experiment exemplifying the type of research envisaged is Eifermann and Etzion's (1964) attempts to account for why the operation of addition is easier than that of subtraction.

The notion of rule which has been considered here is one of a means for relating objects and events in the environment. It has been assumed that individuals are constantly seeking to represent the information in their environment in a more economical form and that to achieve this objective they look for regularities or consistencies in the relationships between stimuli. Rules have been taken to provide a way of describing these consistencies and regularities existing between stimuli. An important distinction is necessary here. Because subjects are found to respond in a manner predicted by a rule analysis, this is not sufficient evidence to indicate that the subject himself is utilizing the same rule as the experimenter, unless it can be conclusively shown that there are no other rules which could be used to produce the response. In the latter case it may be deduced that the particular rule used to relate the stimuli must be part of the subject's behavioural repertoire. With very few exceptions, however, we do not seem to have reached this point where we can make strong assertions about the equivalence of the rules used in task analysis and the rules which give rise to regularities and consistencies in responding. In some cases, in fact, as for the conditional, we have been able to show that the logical rule and the psychological rule are not equivalent.

This distinction needs to be made in view of some apparent confusion in certain theories of intellectual behaviour. Gagne (1970), for example, suggests that a rule "must be something that accounts for regularity in behaviour in the face of virtually infinite variations in specific stimulation. A rule, then, is an inferred capability that enables the individual to respond to a class of stimulus situations with a class of performances". Piaget likewise sees the logical operations used to characterize or describe the thinking of his subjects as intrinsic to their cognitive structure. The position suggested here is that rules are a useful descriptive tool for research on human reasoning and conceptual behaviour, but it seems rather premature to infer that these are the rules which subjects themselves utilize to transform information encountered during a task. In this sense, more evidence is required before we can say, for example, that children have acquired certain logical operations which are contained in cognitive structures (Inhelder and Piaget, 1958). This is not intended to suggest that no attempt should be made to determine the nature of cognitive structures, only that such attempts should be definitive before claiming evidence for the presence of a particular rule governing behaviour. The precision required for producing such evidence represents one of the major challenges facing this approach to cognition.

APPENDIX A

Sentences presented in Experiment XV

1. a. If John catches the bus on time, then he misses his appointment.
   b. He does not miss his appointment.
   c. John does not catch the bus on time.

   Yes? No?

2. a. If the President is going to veto this bill, then the Senate will stand by him, in his efforts to get his tax legislation passed.
   b. The President is not going to veto this bill.
   c. The Senate will not stand by him in his efforts to get his tax legislation passed.

   Yes? No?

3. a. If Mary knows the rules of punctuation, then she did well on the test today.
   b. Mary did well on the test today.
   c. Mary knows the rules of punctuation.

   Yes? No?

4. a. If Mark Twain intended to satirize local customs, then he was a foolish man.
   b. Mark Twain intended to satirize local customs.
   c. He was a foolish man.

   Yes? No?

5. a. If the Board of Education suspends young Brown from school, then it will be acting unconstitutionally.
   b. The Board of Education does not suspend young Brown from school.
   c. It will not be acting unconstitutionally.

   Yes? No?

6. a. If Peter is nearsighted, then his eyes are defective.
   b. Peter is nearsighted.
   c. His eyes are defective.

   Yes? No?

7. a. If the kitchen has dark-coloured walls, then it is well-lighted.
   b. It is not well-lighted.
   c. The kitchen does not have dark-coloured walls.

   Yes? No?

8. a. If the word "going" in that sentence is a gerund, then it functions like a noun.
   b. It functions like a noun.
   c. The word "going" in that sentence is a gerund.

   Yes? No?

9. a. If the Attorney-General intervenes, then the mayor-elect will be a machine politician.
   b. The Attorney-General intervenes.
   c. The mayor-elect will be a machine politician.

   Yes? No?

10. a. If the commissioner is an independent, then he is a night club owner.
    b. The commissioner is not a night club owner.
    c. The commissioner is not an independent.

    Yes? No?

11. a. If this painting has true artistic worth, then it is a pure study in form.
    b. This painting has not true artistic worth.
    c. It is not a pure study in form.

    Yes? No?

12. a. If that man burns our flag, then he is a Communist.
    b. He is a Communist.
    c. That man burns our flag.

Yes? No?

13. a. If there is going to be a storm tonight, then the barometer is falling.
    b. The barometer is not falling.
    c. There is not going to be a storm tonight.

Yes? No?

14. a. If anything is a spider, then it is an insect.
    b. Anything is a spider.
    c. It is an insect.

Yes? No?

15. a. If Henry has moral scruples against drinking, then he never drinks.
    b. He never drinks.
    c. Henry has moral scruples against drinking.

Yes? No?

16. a. If the sum of the digits of 298 is evenly divisible by nine, then 298 is evenly divisible by nine.
    b. The sum of the digits of 298 is not evenly divisible by nine.
    c. 298 is not evenly divisible by nine.

Yes? No?

APPENDIX B

Sentences presented in Experiment XVI

1. a) If the stimulus of examination is unnecessary for the ambitious student, then examinations should be dropped from all courses.

   b) The stimulus of examinations is unnecessary for the ambitious student.

   c) Examinations should be dropped from all courses.

   ANSWER: Yes? No?
   CONFIDENCE: Very confident? Moderately confident? Not confident?

2. a) If the doctrine of the peace movement is applied all around, then there will be an end to military discipline.

   b) The doctrine of the peace movement is not applied all around.

   c) There will not be an end to military discipline.

   ANSWER: Yes? No?
   CONFIDENCE: Very confident? Moderately confident Not confident?

3. a) If education becomes universal, then we will all be white-collar workers.

   b) We will not all be white-collar workers.

   c) Education does not become universal.

   ANSWER: Yes? No?
   CONFIDENCE: Very confident? Moderately confident? Not confident?

4. a) If each one of us can have direct acquaintance only with a single mind, namely, his own, then a scientific psychology is impossible.

   b) A scientific psychology is impossible.

   c) Each one of us can have direct acquaintance only with a single mind, namely, his own.

   ANSWER: Yes? No?
   CONFIDENCE: Very confident? Moderately confident? Not confident?

5. a) If Mary is a true friend, then John is telling the truth.

   b) Mary is a true friend.

   c) John is not telling the truth.

   ANSWER: Yes? No?
   CONFIDENCE: Very confident? Moderately confident? Not confident?

6. a) If his utterance has multiple senses, then it is ambiguous.

   b) His utterance does not have multiple senses.

   c) It is not ambiguous.

   ANSWER: Yes? No?
   CONFIDENCE: Very confident? Moderately confident? Not confident?

7. a) If I play poker, then people will take me as a member of the upper classes.

   b) I play poker.

   c) People will take me as a member of the upper classes.

   ANSWER: Yes? No?
   CONFIDENCE: Very confident? Moderately confident? Not confident?

8. a) If the contract is valid, then Horatio is liable.

b) The contract is valid.

c) Horatio is not liable.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

9. a) If the patient is late for her appointment, then she has lost confidence in the doctor.

b) She has not lost confidence in the doctor.

c) The patient is not late for her appointment.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

10. a) If Marianne believed that Colonel Brandon was too old to marry, then her conduct was not consistent with her beliefs.

b) Her conduct was not consistent with her beliefs.

c) Marianne believed that Colonel Brandon was too old to marry.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

11. a) If the prisoner was tricked into pleading guilty, then the prisoner was denied due process of law.

b) The prisoner was not denied due process of law.

c) The prisoner was not tricked into pleading guilty.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

12. a) If a just peace is not achieved by democratic means, then it is necessary to use violence.

b) A just peace is not achieved by democratic means.

c) It is not necessary to use violence.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

13. a) If the murderer did not want to cover up a crime of theft, then he committed a crime of passion.

b) He committed a crime of passion.

c) The murderer did not want to cover up a crime of theft.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

14. a) If happiness consists in peace of mind, then psychology is necessary for true happiness.

b) Happiness consists in peace of mind.

c) Psychology is necessary for true happiness.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

15. a) If children are paid for all the jobs they do around the house, then the family will be bankrupt.

b) Children are not paid for all the jobs they do around the house.

c) The family will not be bankrupt.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

16. a) If Jones does not appear before the draft board, then he has failed his medical examination.

b) He has failed his medical examination.

c) Jones does not appear before the draft board.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

17. a) If economic behaviour is the lifeless phenomenon portrayed in economic models, then the significant attributes of occupations are their skill and the supply and demand of them.

b) The significant attributes of occupations are not their skill and the supply and demand of them.

c) Economic behaviour is not the lifeless phenomenon portrayed in economic models.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

18. a) If the two Great Powers come to an agreement, then we may hope for peace for at least twenty years.

b) The two Great Powers do not come to an agreement.

c) We may not hope for peace for at least twenty years.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

19. a) If all men are mortal, then the human race will some day come to an end.

b) All men are mortal.

c) The human race will not some day come to an end.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

20. a) If some responsible union official will meet with the Federal conciliation official, then the management proposes that the present investigation be continued.

b) Some responsible union official will meet with the Federal conciliation official.

c) The management proposes that the present investigation be continued.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

21. a) If water has a higher latent heat than air, then more calories are needed to warm a given amount of water than are needed to warm an equal amount of air.

b) Water has a higher latent heat than air.

c) More calories are not needed to warm a given amount of water than are needed to warm an equal amount of air.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

22. a) If inheritance is an innate institution of a capitalist economy, then its abolition would lead to a progressive socialization of the ownership of producers' goods.

b) Its abolition would lead to a progressive socialization of the ownership of producers' goods.

c) Inheritance is an innate institution of capitalist economy.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

23. a) If that citizen enrolls in a political party, then he is forfeiting his independence.

b) That citizen enrolls in a political party.

c) He is forfeiting his independence.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

24. a) If all other circumstances are equal, then wages are higher in new than in old trades.

    b) All other circumstances are not equal.

    c) Wages are not higher in new than in old trades.

    ANSWER: Yes? No?
    CONFIDENCE: Very confident? Moderately confident? Not confident?

25. a) If the Roman Empire crumbled to dust, then it lacked the spirit of liberalism and free enterprise.

    b) It did not lack the spirit of liberalism and free enterprise.

    c) The Roman Empire did not crumble to dust.

    ANSWER: Yes? No?
    CONFIDENCE: Very confident? Moderately confident? Not confident?

26. a) If society is without sympathy and love, then I am not responsible for my actions.

    b) Society is without sympathy and love.

    c) I am responsible for my actions.

    ANSWER: Yes? No?
    CONFIDENCE: Very confident? Moderately confident? Not confident?

27. a) If objects of art are expressive, then they are a language.

    b) They are a language.

    c) Objects of art are expressive.

    ANSWER: Yes? No?
    CONFIDENCE: Very confident? Moderately confident? Not confident?

28. a) If the high price of corn is the effect of an increasing demand, then it is preceded by an increase in wages.

b) It is not preceded by an increase in wages.

c) The high price of corn is not the effect of an increasing demand.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

29. a) If that man is idle, then he is an apprentice.

b) That man is not idle.

c) He is not an apprentice.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

30. a) If Smith is allowed to be a judge in his own cause, then his interest will corrupt his integrity.

b) Smith is allowed to be a judge in his own cause.

c) His interest will corrupt his integrity.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

31. a) If we can take eternity to mean not infinite temporal duration but timelessness, then eternal life belongs to those who live in the present.

b) Eternal life belongs to those who live in the present.

c) We can take eternity to mean not infinite temporal duration but timelessness.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

32. a) If that woman is wealthy, then she has a weak desire to produce children.

b) That woman is not wealthy.

c) She does not have a weak desire to produce children.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

33. a) If extra-sensory perception is accepted as a fact, then the doctrine of clairvoyance must be considered seriously.

b) Extra-sensory perception is accepted as a fact.

c) The doctrine of clairvoyance must be considered seriously.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

34. a) If Manpertius proposed to institute psychological investigations by means of opium, then he was an ingenious man.

b) He was not an ingenious man.

c) Manpertius did not propose to institute psychological investigations by means of opium.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

35. a) If she looksbeautiful in the photograph, then the photographer was very skilful in his use of lighting.

b) She looks beautiful in the photograph.

c) The photographer was not very skilful in his use of lighting.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

36. a) If Pluto has a diameter of more than 4,200 miles, then an occultation occurred at McDonald Observatory.

b) An occultation did not occur at McDonald Observatory.

c) Pluto does not have a diameter of more than 4,200 miles.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

37. a) If starving Hindus are sensible, they will slaughter their cows for food.

b) Starving Hindus are sensible.

c) They will not slaughter their cows for food.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

38. a) If food is constantly supplied to them, then the very fiercest creatures live peacably together.

b) Food is not constantly supplied to them.

c) The very fiercest creatures do not live peacably together.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

39. a) If people who are tolerably fortunate in their outward lot do not find in life sufficient enjoyment to make it valuable to them, then the cause is, caring for nobody but themselves.

b) People who are tolerably fortunate in their outward lot do not find in life sufficient enjoyment to make it valuable to them.

c) The cause is, caring for nobody but themselves.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

40. a) If these laws burden the petitioners' right to worship God in their own fashion, then the laws should be struck down.

b) The laws should be struck down.

c) These laws burden the petitioners' right to worship God in their own fashion.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

41. a) If the license tax is laid specifically on the privilege of disseminating ideas, then the license tax would infringe the right of free speech.

b) The license tax is not laid specifically on the privilege of disseminating ideas.

c) The license tax would not infringe on the right of free speech.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

42. a) If these children have a few more years of compulsory education, then they will all end up gibbering idiots.

b) These children have a few more years of compulsory education.

c) They will all end up gibbering idiots.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

43. a) If Hamilton held a considerable number of mining shares, then he left a large estate at his death.

b) Hamilton held a considerable number of mining shares.

c) He did not leave a large estate at his death.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

44. a) If Kingsley gives the name "fair sex" to that under-sized, narrow-shouldered, broad-hipped, short-legged race, then his intellect is clouded by his sexual impulse.

b) His intellect is clouded by his sexual impulse.

c) Kingsley gives the name "fair sex" to that under-sized, narrow-shouldered, broad-hipped, short-legged race.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

45. a) If that question is discussed freely, then it is settled rightly.

b) It is not settled rightly.

c) That question is not discussed freely.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

46. a) If Mr. Jones is the brakeman's next-door neighbour, then Mr. Jones' annual earnings are exactly divisible by three.

b) Mr. Jones is the brakeman's next-door neighbour.

c) Mr. Jones' annual earnings are exactly divisible by three.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

47. a) If we interfere with the publication of false and harmful doctrines, then we shall be guilty of suppressing the liberties of others.

b) We shall not be guilty of suppressing the liberties of others.

c) We do not interfere with the publication of false and harmful doctrines.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

48. a) If the market is perfectly free, then a single supplier cannot affect prices.

b) A single supplier cannot affect prices.

c) The market is perfectly free.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

49. a) If the car was drifting all over the road, then the driver had a breathalyser reading above .08.

b) The car was drifting all over the road.

c) The driver did not have a breathalyser reading above .08.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

50. a) If the Mosaic account of the cosmogony is strictly correct, then the sun was created on the fourth day.

b) The Mosaic account of the cosmogony is not strictly correct.

c) The sun was not created on the fourth day.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

51. a) If Harrison was a cigarette smoker, then he died of lung cancer.

b) Harrison was a cigarette smoker.

c) He did not die of lung cancer.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

52. a) If Harold entered into the contract, then the bank will lend him money.

b) The bank will lend him money.

c) Harold entered into the contract.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

53. a) If people like to live with their in-laws, then housing is scarce.

b) People like to live with their in-laws.

c) Housing is scarce.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

54. a) If prices are low, then sales are high.

b) Sales are not high.

c) Prices are not low.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

55. a) If every witness is telling the truth, then Bluenose will be found guilty.

b) Every witness is not telling the truth.

c) Bluenose will not be found guilty.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

56. a) If either wages or prices are raised, there will be inflation.

b) There will be inflation.

c) Either wages or prices are raised.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

57. a) If there are government subsidies of agriculture, then there are government controls of agriculture.

b) There are not government subsidies of agriculture.

c) There are not government controls of agriculture.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

58. a) If Henry graduated from medical school, then his income is very high.

b) Henry graduated from medical school.

c) His income is not very high.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

59. a) If the governor favours public housing, then he is in favour of restricting the scope of private enterprise.

b) The governor favours public housing.

c) He is in favour of restricting the scope of private enterprise.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

60. a) If Mrs. Elton is a snob, then Mrs. Bates is a bore.

b) Mrs. Bates is not a bore.

c) Mrs. Elton is not a snob.

ANSWER: Yes? No?
CONFIDENCE: Very confident? Moderately confident? Not confident?

REFERENCES

ADAMS, J. Concepts as operators. Psychological Review, 1953, 60, 241-251.

ADAMS, J. A. Motor behaviour, In M. H. Marx (Ed.) Learning: Processes. London: Macmillan, 1969.

ANDERSON, R. C., and GUTHRIE, J. T. Effects of some sequential manipulations of relevant and irrelevant stimulus dimensions on concept learning. Journal of Experimental Psychology, 1966, 72, 501-504.

ARCHER, E. J. Concept identification as a function of obviousness of relevant and irrelevant information. Journal of Experimental Psychology, 1962, 63, 616-620.

ARCHER, E. J., BOURNE, L. E.,Jr., and BROWN, F. G. Concept identification as a function of irrelevant information and instruction. Journal of Experimental Psychology, 1955, 49, 153-164

ATTNEAVE, F. Some informational aspects of visual perception. Psychological Review, 1954, 61, 183-193.

BADDELEY, A. D. The capacity for generating information by randomization. Quarterly Journal of Experimental Psychology, 1966, 18, 119-129.

BANKS, W. P. Signal detection theory and human memory. Psychological Bulletin, 1970, 74, 81-99.

BATTIG, W. F. Sequential learning and thought: An overlooked communality. In J. F. Voss (Ed.), Approaches to thought. Columbus: Merrill, 1969.

BATTIG, W. F., and BOURNE, L. E., Jr., Concept identification as a function of intradimensional and interdimensional variation. Journal of Experimental Psychology, 1961, 61, 329-333.

BAUM, M. Simple concept learning as a function of intralist generalization. Journal of Experimental Psychology, 1954, 47, 89-94.

BEGG, I., and DENNY, J. P. Empirical reconciliation of atmosphere and conversion interpretations of syllogistic reasoning. Journal of Experimental Psychology, 1969, 81, 351-354.

BEM, D. J. Beliefs, attitudes, and human affairs. Belmont, California: Brooks/Cole, 1970.

BERLYNE, D. E. Structure and direction in thinking. New York: Wiley, 1965.

BINDRA, D., WILLIAMS, J., and WISE, S. S. Judgements of sameness and difference: Experiments on reaction time. Science, 1965, 150, 1625-1626.

BOOLE, G. An investigation of the laws of thought. London: Macmillan, 1854.

BOURNE, L. E., Jr. Effects of delay of information feedback and task complexity on the identification of concepts. Journal of Experimental Psychology, 1957, 47, 89-94.

BOURNE, L. E., Jr. Long term effects of misinformative feedback on concept identification. Journal of Experimental Psychology, 1963, 65, 139-147.

BOURNE, L. E., Jr. Hypotheses and hypothesis shifts in classification learning. Journal of General Psychology, 1965, 72, 251-261.

BOURNE, L. E., Jr. Human conceptual behaviour. Boston: Allyn and Bacon, 1966.

BOURNE, L. E., Jr. Learning and utilization of conceptual rules. In B. Kleinmuntz (Ed.), Concepts and the structure of memory. New York: Wiley, 1967.

BOURNE, L. E., Jr. Concept learning and thought: Behaviour, not process. In J. F. Voss (Ed.), Approaches to thought. Columbus: Merrill, 1969.

BOURNE, L. E., Jr. Knowing and using concepts. Psychological Review, 1970, 77, 546-556.

BOURNE, L. E., Jr., and BATTIG, W. F. Complex processes. In J. B. Sidowski (Ed.), Experimental methods and instrumentation in psychology. New York: McGraw-Hill, 1966.

BOURNE, L. E., Jr., and BUNDERSON, C. V. Effects of delay of informative feedback and length of post feedback interval on concept identification. Journal of Experimental Psychology, 1963, 65, 1-5.

BOURNE, L. E. Jr., EKSTRAND, B., and MONTGOMERY, B. Concept learning as a function of the conceptual rule and the availability of positive and negative instances. Journal of Experimental Psychology, 1969, 82, 538-544.

BOURNE, L. E., Jr., GOLDSTEIN, S., and LINK, W. E. Concept learning as a function of availability of previously presented information. Journal of Experimental Psychology, 1964, 67, 439-448.

BOURNE, L. E., Jr., and GUY, D. E. Learning conceptual rules: I. Some inter-rule transfer effect. Journal of Experimental Psychology, 1968, 76, 423-429. (a)

BOURNE, L. E., Jr., and GUY, D. E. Learning conceptual rules: II. The role of positive and negative instances. Journal of Experimental Psychology, 1968, 77, 488-494. (b)

BOURNE, L. E., Jr., GUY, D. E., DODD, D. H., and JUSTESEN, D. R. Concept identification: The effects of varying length and informational components of the intertrial interval. Journal of Experimental Psychology, 1965, 69, 624-629.

BOURNE, L. E. Jr., GUY, D. E., and WADSWORTH, N. Verbal-reinforcement combinations and the relative frequency of informative feedback in a card-sorting task. Journal of Experimental Psychology, 1967, 73, 220-226.

BOURNE, L. E., Jr., and HAYGOOD, R. C. The role of stimulus redundancy in the identification of concepts. Journal of Experimental Psychology, 1959, 58, 232-238.

BOURNE, L. E., Jr., and HAYGOOD, R. C. Supplementary report: Effect of redundant relevant information upon the identification of concepts. Journal of Experimental Psychology, 1961, 61, 259-260.

BOURNE, L. E., Jr., and JENNINGS, P. C. The relationship between contiguity and classification learning. Journal of General Psychology, 1963, 69, 335-338.

BOURNE, L. E., Jr., and O'BANION, K. Memory for individual events in concept identification. Psychonomic Science, 1969, 16, 101-103.

BOURNE, L. E., Jr., and O'BANION, K. Conceptual rule learning and Chronological age. Program on Concept Learning Report No. 11, Institute for the Study of Intellectual Behaviour, University of Colorado, 1970.

BOURNE, L. E., Jr., and PENDLETON, R. B. Concept identification as a function of completeness and probability of information feedback. Journal of Experimental Psychology, 1958, 56, 413-420.

BOURNE, L. E., Jr., and RESTLE, F. Mathematical theory of concept identification. Psychological Review, 1959, 66, 278-296.

BOWER, A. C., and KING, W. L. The effect of number of stimulus dimensions, verbalization, and sex on learning biconditional classification rules. Psychonomic Science, 1967, 8, 453-454.

BOWER, G. H., and TRABASSO, T. Reversals prior to solution in concept identification. Journal of Experimental Psychology, 1963, 66, 409-418.

BOWER, G. H., and TRABASSO, T. Concept identification. In R. C. Atkinson (Ed.), Studies in mathematical psychology. Stanford: Stanford University Press, 1964.

BROWN, F. G., and ARCHER, F. J. Concept identification as a function of task complexity and distribution of practice. Journal of Experimental Psychology, 1956, 52, 316-321.

BRUNER, J. S., GOODNOW, J. J., and AUSTIN, G. A. A study of thinking. New York: Wiley, 1956.

BRUSH, F. R. Stimulus uncertainty, response uncertainty, and problem solving. Canadian Journal of Psychology, 1956, 10, 239-247.

BUCHWALD, A. M. Experimental alterations in the effectiveness of verbal reinforcement combinations. Journal of Experimental Psychology, 1959, 57, 351-361.

BULGARELLA, R., and ARCHER, E. J. Concept identification of auditory stimuli as a function of amount of relevant and irrelevant information. Journal of Experimental Psychology, 1962, 63, 254-257.

BUSS, A. H., BRADEN, W., ORGEL, A., and BUSS, E. H. Acquisition and extinction with different verbal reinforcement combinations. Journal of Experimental Psychology, 1956, 52, 288-295.

BUSS, A. H., and BUSS, E. H. The effect of verbal reinforcement combinations on conceptual learning. Journal of Experimental Psychology, 1956, 52, 282-287.

BYERS, J. L. Strategies and learning set in concept attainment. *Psychological Reports*, 1963, *12*, 623-634.

BYERS, J. L., and DAVIDSON, R. E. The role of hypothesizing in the facilitation of concept attainment. *Journal of Verbal Learning and Verbal Behaviour*, 1967, *6*, 595-600.

BYERS, J. L., and DAVIDSON, R. E. Relevant and irrelevant information in concept attainment. *Journal of Experimental Psychology*, 1968, *76*, 277-281.

BYERS, J. L., DAVIDSON, R. E., and ROHWER, W. D. The effects of strategy instructions and memory on concept attainment. *Journal of Verbal Learning and Verbal Behaviour*, 1968, *7*, 831-837.

CAHILL, H. E., and HOVLAND, C. I. The role of memory in the acquisition of concepts. *Journal of Experimental Psychology*, 1960, *59*, 137-144.

CAMPBELL, A. C. On the solving of code items demanding the use of indirect procedures. *British Journal of Psychology*, 1965, *56*, 45-51.

CHAPANIS, A. Random-number guessing behaviour. *American Psychologist*, 1953, *8*, 332.

CHAPMAN, L. J., and CHAPMAN, J. P. Atmosphere effect re-examined. *Journal of Experimental Psychology*, 1959, *58*, 220-226.

CLARK, H. H. Linguistic processes in deductive reasoning. *Psychological Review*, 1969, *76*, 387-404.

COLEMAN, E. B. Verbal concept learning as a function of instructions and dominance level. *Journal of Experimental Psychology*, 1964, *68*, 213-214.

CONANT, M. B., and TRABASSO, T. Conjunctive and disjunctive concept formation under equal-information conditions. *Journal of Experimental Psychology*, 1964, *67*, 250-255.

D'AMATO, M. R. Instrumental conditioning. In M. H. Marx (Ed.) *Learning: Processes.* London: Macmillan, 1969.

DEFFENBACHER, K. An extension of four single-cue process models to conjunctive concept learning. *Psychological Reports*, 1970, *26*, 95-102.

DENNY, N. R. Memory and transformations in concept learning. Journal of Experimental Psychology, 1969, 79, 63-68. (a)

DENNY, N. R. Memory load and concept-rule difficulty. Journal of Verbal Learning and Verbal Behaviour, 1969, 8, 202-205. (b)

DeSOTO, C., LONDON, M., and HANDEL, S. Social reasoning and spatial paralogic. Journal of Personality and Social Psychology, 1965, 2, 513-521.

DETAMBEL, M. H., and STOLUROW, L. M. Stimulus sequence and concept learning. Journal of Experimental Psychology, 1956, 51, 34-40.

DIENES, Z. P., and JEEVES, M. A. Thinking in structures. London: Hutchinson, 1965.

DIENES, Z. P., and JEEVES, M. A. The effect of structural relations upon transfer. London: Hutchinson, 1970.

DI VESTA, F. J., and WALLS, R. T. Rule and attribute identification in children's attainment of disjunctive and conjunctive concepts. Journal of Experimental Psychology, 1969, 80, 498-504.

DODD, D. H., and BOURNE, L. E., Jr. Test of some assumptions of a hypothesis-testing model of concept identification. Journal of Experimental Psychology, 1969, 80, 69-72.

DOMINOWSKI, R. L. Role of memory in concept learning. Psychological Bulletin, 1965, 63, 271-280.

DOMINOWSKI, R. L. Concept attainment as a function of instance contiguity and number of irrelevant dimensions. Proceedings of the 76th Annual Convention of the American Psychological Association, 1968, 3, 41-42.

DONALDSON, M. Positive and negative information in matching problems. British Journal of Psychology, 1959, 50, 235-262.

DUNCAN, C. P. Recent research in problem solving. Psychological Bulletin, 1959, 56, 397-429.

DUNCAN, D. B. Multiple range and multiple F tests. Biometrics, 1955, 11, 1 - 42.

EIFERMANN, R. R. Negation: A linguistic variable. Acta Psychologica, 1961, 18, 258-273.

EIFERMANN, R. R. Selection strategies in concept attainment: A re-examination. In R. R. Eifermann (Ed.) Scripta hierosolymitana: Studies in psychology. The Hebrew University of Jerusalem: Magnes Press, 1965 (a)

EIFERMANN, R. R. Response patterns and strategies in the dynamics of concept attainment behaviour. British Journal of Psychology, 1965, 56, 217-222. (b)

EIFERMANN, R. R., and ETZION, D. Awareness of reversibility: Its effect on performance of converse arithmetical operations. British Journal of Educational Psychology, 1964, 34, 151-157.

ENNIS, R. H. Logic in teaching. Englewood cliffs, N.J.: Prentice-Hall, 1969.

ENTUS, A., and BINDRA, D. Common features of the "repetition" and "same-different" effects in reaction time experiments. Perception and Psychophysics, 1970, 7, 143-148.

ERICKSON, J. R. Hypothesis sampling in concept identification. Journal of Experimental Psychology, 1968, 76, 12-18.

ERICKSON, J. R., and ZAJKOWSKI, M. M. Learning several concept identification problems concurrently. Journal of Experimental Psychology, 1967, 74, 212-218.

ERICKSON, J. R., ZAJKOWSKI, M. M., and EHMANN, E. D. All-or-none assumptions in concept identification. Journal of Experimental Psychology, 1966, 72, 690-697.

FALMAGNE, R. Construction of a hypothesis model for concept identification. Journal of Mathematical Psychology, 1970, 7, 60-96.

FEATHER, N. T. Acceptance and rejection of arguments in relation to attitude strength, critical ability, and intolerance of inconsistency. Journal of Abnormal and Social Psychology, 1964, 69, 127-136.

FILLENBAUM, S. On coping with ordered and unordered conjunctive sentences. Journal of Experimental Psychology, 1971, 87, 93-98.

FRANKL, F., LEVINE, M., and KARPF, D. Human discrimination learning: a test of the blank-trials assumption. Journal of Experimental Psychology, 1970, 85, 342-348.

FREIBERGS, V., and TULVING, E. The effect of practice on utilization of information from positive and negative instances in concept identification. Canadian Journal of Psychology, 1961, 15, 101-106.

GAGNE, R. M. The conditions of learning, 2nd edition. New York: Holt, Rinehart, and Winston, 1970.

GARNER, W. R. Uncertainty and structure as psychological concepts. New York: Wiley, 1962.

GELFORD, S. Effects of prior associations and task complexity upon the identification of concepts. Psychological Reports, 1958, 4, 567-574.

GIAMBRA, L. M. Effect of number of irrelevant dimensions with ten concept types on the attribute identification task in the selection mode with exemplar and non-exemplar start cards. Psychonomic Science, 1969, 14, 75-76.

GIAMBRA, L. M. Conditional and biconditional rule difficulty with attribute identification, rule learning, and complete learning tasks. Journal of Experimental Psychology, 1970, 86, 250-254.

GIAMBRA, L. M. Selection strategies for eight concept rules with non-exemplar start cards. Journal of Experimental Psychology, 1971, 87, 78-92 (a)

GIAMBRA, L. M. Selection strategies for eight concept rules with exemplar and nonexemplar start cards: A within-subjects replication. Journal of Experimental Psychology, 1971, 87 143-145. (b)

GLANZER, M., HUTTENLOCHER, J., and CLARK, W. H. Systematic operations in solving concept problems: A parametric study of class of problems. Psychological Monographs, 1963, 77, No. 1.

GORDEN, R. L. The effect of attitude toward Russia on logical reasoning. Journal of Social Psychology, 1953, 37, 103-111.

GREENE, J. M. The semantic function of negatives and passives. British Journal of Psychology, 1970, 61, 17-22.

GREGG, L. W., and SIMON, H. A. Process models and stochastic theories of simple concept formation. Journal of Mathematical Psychology, 1967, 4, 246-276.

GUY, D. E., VAN FLEET, F.M., and BOURNE, L. E., Jr. Effects of adding a stimulus dimension prior to a nonreversal shift. Journal of Experimental Psychology, 1966, 72, 161-168.

HARRE, R. The formal analysis of concepts. In H. J. Klausmeier and C. W. Harris (Eds.), Analyses of concept learning. New York: Academic Press, 1966.

HAYGOOD, R. C., and BOURNE, L. E., Jr. Forms of relevant stimulus redundancy in concept identification. Journal of Experimental Psychology, 1964, 67, 392-397.

HAYGOOD, R. C., and BOURNE, L. E., Jr. Attribute and rule learning aspects of conceptual behaviour. Psychological Review, 1965, 72, 175-195.

HAYGOOD, R. C., and DEVINE, J. V. Effects of composition of the positive category on concept learning. Journal of Experimental Psychology, 1967, 74, 230-235.

HAYGOOD, R. C., HARBERT, T. L., and OMLOR, J. A. Intradimensional variability and concept identification. Journal of Experimental Psychology, 1970, 83, 216-219.

HAYGOOD, R. C.,SANDLIN, J., YODER, D. J., and DODD, D. H. Instance contiguity in disjunctive concept learning. Journal of Experimental Psychology, 1969, 81, 605-607.

HAYGOOD, R. C., and STEVENSON, M. Effects of number of irrelevant dimensions in nonconjunctive concept learning. Journal of Experimental Psychology, 1967, 74, 302-304.

HAYS, W. Statistics for psychologists. New York: Holt, Rinehart, and Winston, 1963.

HENLE, M. On the relation between logic and thinking. Psychological Review, 1962, 69, 366-378.

HOLSTEIN, S. B., and PREMACK, D. On the different effects of random reinforcement and presolution reversal on human concept identification. Journal of Experimental Psychology, 1965, 70, 335-337.

HOVLAND, C. I. A "communication analysis" of concept learning. Psychological Review, 1952, 59, 461-472.

HOVLAND, C. I., and WEISS, W. Transmission of information concerning concepts through positive and negative instances. Journal of Experimental Psychology, 1953, 45, 165-182.

HULL, C.L. Quantitative aspects of the evolution of concepts. Psychological Monographs, 1920, 28, No. 1.

HUNT, E. B. Concept learning: An information processing problem. New York: Wiley, 1962.

HUNT, E. B. Selection and reception conditions in grammar and concept learning. Journal of Verbal Learning and Verbal Behaviour, 1965, 4, 211-215.

HUNT, E. B., and HOVLAND, C. I. Order of consideration of different types of concepts. Journal of Experimental Psychology, 1960, 59, 220-225.

HUNT, E. B., MARIN, J., and STONE, P. J. Experiments in induction. New York: Academic Press, 1966.

HUNTER, I. M. L. The solving of three-term series problems. British Journal of Psychology, 1957, 48, 286-298.

HUTTENLOCHER, J. Effects of manipulation of attributes on efficiency of concept formation. Psychological Reports, 1962, 10, 503-509. (a)

HUTTENLOCHER, J. Some effects of negative instances on the formation of simple concepts. Psychological Reports, 1962, 11, 35-42. (b)

HUTTENLOCHER, J. Constructing spatial images: A strategy in reasoning. Psychological Review, 1968, 75, 550-560.

INHELDER, B., and PIAGET, J. The growth of logical thinking from childhood to adolescence. New York: Basic Books, 1958.

JACOBY, L. L., and RADTKE, R. C. Effect of contiguity and meaningfulness of relevant and irrelevant attributes on concept formation. Journal of Experimental Psychology, 1969, 81, 454-459.

JANIS, I. L., and FRICK, F. The relationship between attitudes toward conclusions and errors in judging logical validity of syllogisms. Journal of Experimental Psychology, 1943, 33, 73-77.

JEEVES, M. A. Experiments on structural thinking. Australian Journal of Psychology, 1968, 20, 93-109.

JEFFREY, R. C. Formal logic: Its scope and limits. New York: McGraw-Hill, 1967.

JENKINS, J. J. Language and thought. In J. F. Voss (Ed.), Approaches to thought. Columbus, Ohio: Merrill, 1969.

JENSEN, A. R. Individual differences in concept learning. In H. J. Klausmeier and C. W. Harris (Eds.), Analyses of concept learning. New York: Academic Press, 1966.

JOHNSON-LAIRD, P. N. Reasoning with ambiguous sentences. British Journal of Psychology, 1969, 60, 17-23. (a)

JOHNSON-LAIRD, P. N. On understanding logically complex sentences. Quarterly Journal of Experimental Psychology, 1969, 21, 1-13. (b)

JOHNSON-LAIRD, P. N., and TAGART, J. How implication is understood. American Journal of Psychology, 1969, 82, 367-373.

JOHNSON-LAIRD, P. N., and WASON, P. C. Insight into a logic relation. Quarterly Journal of Experimental Psychology, 1970, 22, 49-61. (a)

JOHNSON-LAIRD, P. N., and WASON, P. C. A theoretical analysis of insight into a reasoning task. Cognitive Psychology, 1970, 1, 134-148. (b)

JONES, E. E., and GERARD, H. B. Foundations of social psychology. New York: Wiley, 1967.

JONES, S. Decoding a deceptive instruction. British Journal of Psychology, 1966, 57, 405-411.

KANT, I. Introduction to logic and essay on the mistaken subtlity of the four figures (translated by T. K. Abbott). London: Longmans, 1885.

KAUFMAN, H., and GOLDSTEIN, S. The effects of emotional value of conclusions upon distortions in syllogistic reasoning. Psychonomic Science, 1967, 7, 367-368.

KAY, H. The effects of position in a display on problem solving. Quarterly Journal of Experimental Psychology, 1954, 6, 155-169.

KEELE, S. W., and ARCHER, E. J. A comparison of two types of information in concept identification. Journal of Verbal Learning and Verbal Behaviour, 1967, 6, 185-192.

KENDLER, H. H., and KENDLER, T. S. Vertical and horizontal processes in problem solving. Psychological Review, 1962, 69, 1 - 16.

KINTSCH, W. Learning, memory, and conceptual processes. New York: Wiley, 1970.

KURTZ, K. H., and HOVLAND, C. I. Concept learning with differing sequences of instances. Journal of Experimental Psychology, 1956, 51, 239-243.

LAUGHLIN, P. R. Selection strategies in concept attainment as a function of number of persons and stimulus display. Journal of Experimental Psychology, 1965, 70, 323-327.

LAUGHLIN, P. R. Conditional concept attainment as a function of if factor complexity and then factor complexity. Journal of Experimental Psychology, 1968, 77, 212-222. (a)

LAUGHLIN, P. R. Focusing strategy for eight concept rules. Journal of Experimental Psychology, 1968, 77, 661-669. (b)

LAUGHLIN, P. R. Information specification in the attainment of conditional concepts. Journal of Experimental Psychology, 1969, 79, 370-372. (a)

LAUGHLIN, P. R. Selection versus reception concept-attainment paradigms as a function of memory, concept rule, and concept universe. Journal of Educational Psychology, 1969, 60, 267-273. (b)

LAUGHLIN, P. R., and JORDAN, R. M. Selection strategies in conjunctive, disjunctive, and biconditional concept attainment. Journal of Experimental Psychology, 1967, 75, 188-193.

LEFFORD, A. The influence of emotional subject matter on logical reasoning. Journal of General Psychology, 1946, 34, 127-151.

LEVINE, M. Cue neutralization: The effects of random reinforcements upon discrimination learning. Journal of Experimental Psychology, 1962, 63, 438-443.

LEVINE, M. Mediating processes in humans at the outset of discrimination learning. Psychological Review, 1963, 70, 254-276.

LEVINE, M. Hypothesis behaviour by humans during discrimination learning. Journal of Experimental Psychology, 1966, 71, 331-338.

LEVINE, M. The size of the hypothesis set during discrimination learning. Psychological Review, 1967, 74, 428-430.

LEVINE, M. Neo-noncontinuity theory. In G. Bower and J. T. Spence (Eds.), The psychology of learning and motivation. Vol. 3. New York: Academic Press, 1969.

LEVINE, M. Human discrimination learning: the subset-sampling assumption. Psychological Bulletin, 1970, 74, 397-404.

LEVINE, M., LEITENBERG, H., and RICHTER, M. The blank trials law: the equivalence of positive reinforcement and nonreinforcement. Psychological Review, 1964, 71, 94-103.

LEVINE, M., MILLER, P., and STEINMEYER, C. H. The none-to-all theorem of human discrimination learning. Journal of Experimental Psychology, 1967, 73, 568-573.

LEVINE, M., YODER, R.M., KLEINBERG, J., and ROSENBERG, J. The presolution paradox in discrimination learning. Journal of Experimental Psychology, 1968, 77, 602-608.

LOONEY, N. J., and HAYGOOD, R. C. Effects of number of relevant dimensions in disjunctive concept learning. Journal of Experimental Psychology, 1968, 78, 169-171.

LORDAHL, D. S. Concept identification using simultaneous auditory and visual signals. Journal of Experimental Psychology, 1961, 62, 282-290.

LORGE, I., TUCKMAN, J., AIKMAN, L., SPIEGEL, J., and MOSS G. Solutions by teams and by individuals to a field problem at different levels of reality. Journal of Educational Psychology, 1955, 46, 17-24. (a)

LORGE, I., TUCKMAN, J., AIKMAN, L., SPIEGEL, J., and MOSS G. Problem-solving by teams and by individuals in a field setting. Journal of Educational Psychology, 1955, 46, 160-166. (b)

LUCHINS, A. S. Classroom experiments on mental set. American Journal of Psychology, 1946, 59, 295-298.

LUCHINS, A. S. Primacy-recency in impression formation. In C. I. Hovland (Eds.), The order of presentation in persuasion. New Haven: Yale University Press, 1957.

MANDLER, G., COWAN, P. A., and GOLD, C. Concept learning and probability matching. Journal of Experimental Psychology, 1964, 67, 514-522.

MATALON, B. Etude genetique de l'implication. Etudes d'epistemologie genetique: XVI, Implication, formalisation, et logique naturelle, 1962, 69-95.

MILL, J. S. A system of logic, 8th edition. London: Longmans, 1874.

MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 1956, 63, 81-97.

MILLER, S. A. Bidimensional solution shifts in children and adults. Program on Concept Learning Report No. 10, Institute for the Study of Intellectual Behaviour, University of Colorado, 1970.

MOORE, O. K., and ANDERSON, S. B. Modern logic and tasks for experiments on problem solving behaviour. Journal of Psychology, 1954, 38, 151-160.

MORGAN, J. J. B., and MORTON, J. T. The distortion of syllogistic reasoning produced by personal convictions. Journal of Social Psychology, 1944, 20, 39-59.

MURRAY, F. S., and GREGG, R. E. Reception versus selection procedures in concept learning. Journal of Experimental Psychology, 1969, 82, 571-572.

NAHINSKY, I. D. A test of axioms of all-or-none concept identification models. Journal of Verbal Learning and Verbal Behaviour, 1968, 7, 593-601.

NAHINSKY, I. D. A hypothesis sampling model for conjunctive concept identification. Journal of Mathematical Psychology, 1970, 7, 293-316.

NAHINSKY, I. D., PENROD, W. C., and SLAYMAKER, F. L. Relationship of component cues to hypotheses in conjunctive concept learning. Journal of Experimental Psychology, 1970, 83, 351-353.

NAHINSKY, I. D., and SLAYMAKER, F. L. Sampling without replacement and information processing following correct responses in concept identification. Journal of Experimental Psychology, 1969, 80, 475-482.

NAHINSKY, I. D., and SLAYMAKER, F. L. Use of negative instances in conjunctive concept identification. Journal of Experimental Psychology, 1970, 84, 64-68.

NEIMARK, E. D. Development of comprehension of logical connectives: Understanding of "or". Psychonomic Science, 1970, 21, 217-219.

NEIMARK, E. D., and SLOTNICK, N. S. Development of the understanding of logical connectives. Journal of Educational Psychology, 1970, 61, 451-460.

NEISSER, U., and WEENE, P. Hierarchies in concept attainment, Journal of Experimental Psychology, 1962, 64, 644-645.

NEWMAN, S. E. Effects of contiguity and similarity on the learning of concepts. Journal of Experimental Psychology, 1956, 52, 349-353.

NICKERSON, R. S. Response times for "same" - "different" judgments. Perceptual and Motor Skills, 1965, 20, 15-18.

NICKERSON, R. S. Categorization time with categories defined by disjunctions and conjunctions of stimulus attributes. Journal of Experimental Psychology, 1967, 73, 211-219.

NICKERSON, R. S., and FEEHRER, C. E. Stimulus categorization and response time. Perceptual and Motor Skills, 1964, 18, 785-793.

OLDFIELD, R. C. Memory mechanisms and the theory of schemata. British Journal of Psychology, 1954, 45, 14-23.

OSGOOD, C. E. Method and theory in experimental psychology. New York: Oxford, 1953.

PARSON, C. Inhelder and Piaget's, The growth of logical thinking. II, A logician's viewpoint. British Journal of Psychology, 1960, 51, 75-84.

PEEL, E. A. A method for investigating children's understanding of certain logical connectives used in binary propositional thinking. British Journal of Mathematical and Statistical Psychology, 1967, 20, 81-92.

PETERSON, M. J. Some effects of the percentage of relevant cues and presentation methods on concept identification. Journal of Experimental Psychology, 1962, 64, 623-627.

PETERSON, M. J. Concept identification as a function of type of training series. Journal of Experimental Psychology, 1968, 78, 128-136.

PIAGET, J. The language and thought of the child. New York: Harcourt, Brace, 1926.

PIAGET, J. Logic and Psychology. New York: Basic Books, 1957.

PISHKIN, V. Availability of feedback - corrected error instances in concept learning. Journal of Experimental Psychology, 1967, 73, 318-319.

PISHKIN, V., and WOLFGANG, A. Number and type of available instances in concept learning. Journal of Experimental Psychology, 1965, 69, 5-8.

PISHKIN, V., WOLFGANG, A., and RASMUSSEN, E. Age, sex, amount, and type of memory information in concept learning. Journal of Experimental Psychology, 1967, 73, 121-124.

POSNER, M. I. Information reduction in the analysis of sequential tasks. Psychological Review, 1964, 71, 491-504.

POSNER, M. I. Memory and thought in human intellectual performance. British Journal of Psychology, 1965, 56, 197-215.

POSNER, M. I., and MITCHELL, R. F. Chronometric analysis of classification. Psychological Review, 1967, 74, 392-409.

POSNER, M. I., and ROSSMAN, E. Effect of size and location of informational transforms upon short-term retention. Journal of Experimental Psychology, 1965, 70, 496-505.

RESTLE, F. The selection of strategies in cue learning. Psychological Review, 1962, 69, 329-343.

RESTLE, F. Speed of adding and comparing numbers. Journal of Experimental Psychology, 1970, 83, 274-278.

RESTLE, F. Theory of serial pattern learning: Structural trees. Psychological Review, 1970, 77, 481-495. (a)

RESTLE, F. and EMMERICH, D. Memory in concept attainment. Journal of Experimental Psychology, 1966, 71, 794-799.

RUSSELL, B. The axiom of infinity. Hibbert Journal, 1904, 2, 809-812.

SALMON, W. Logic. Englewood Cliffs, N. J.: Prentice-Hall, 1963.

SAUGSTAD, P., and RAAHEIM, K. Problem-solving, past experience and availability of functions. British Journal of Psychology, 1960, 51, 97-104.

SCANDURA, J. M. Role of rules in behaviour: Toward an operational definition of what (rule) is learned. Psychological Review, 1970, 77, 516-533.

SCHILLER, F. C. S. Logic for use. London: Bell, 1930.

SCHULZ, R. W., MILLER, R. L., and RADTKE, R. C. The role of instance contiguity and dominance in concept attainment. Journal of Verbal Learning and Verbal Behaviour, 1963, 1, 432-435.

SCHWARTZ, S. N. Trial-by-trial analysis of processes in simple and disjunctive concept-attainment tasks. Journal of Experimental Psychology, 1966, 72, 456-465.

SEGGIE, J. L. Levels of Learning involved in conjunctive and disjunctive concepts. Australian Journal of Psychology, 1969, 21, 325-333.

SELLS, S. B. The atmosphere effect: An experimental study of reasoning. Archives of Psychology, 1936, No. 200.

SHEPARD, R. N., HOVLAND, C. I., and JENKINS, H. N. Learning and memorization of classifications. Psychological Monographs, 1961, 75, No. 13.

SHEPP, B. E., and ZEAMAN, D. Discrimination learning of size and brightness by retardates. Journal of Comparative and Physiological Psychology, 1966, 62, 55-59.

SIEGEL, S. Non-parametric statistics for the behavioural sciences. New York: McGraw-Hill, 1956.

SIMON, H. A., and KOTOVSKY, K. Human acquisition of concepts for sequential patterns. Psychological Review, 1963, 70, 534-546.

SIMPSON, M. E., and JOHNSON, D. M. Atmosphere and conversion errors in syllogistic reasoning. Journal of Experimental Psychology, 1966, 72, 197-200.

SKINNER, B. F. Verbal behaviour. New York: Appleton-Century-Crofts, 1957.

SMOKE, K. L. Negative instances in concept learning. Journal of Experimental Psychology, 1933, 16, 583-588.

STEVENS, S. S. Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), Handbook of experimental psychology. New York: Wiley, 1951.

SUPPES, P. On the behavioural foundations of mathematical concepts. In L. N. Morrisett and J. Vinsonhaler (Eds.), Mathematical learning, Monographs of the Society for Research in Child Development, 1965, 30, No. 1, 60-96.

SUPPES, P., and SCHLAG-REY, M. Observable changes of hypotheses under positive reinforcement. Science, 1965, 148, 661-662.

THISTLEWAITE, D. Attitude and structure as factors in the distortion of reasoning. Journal of Abnormal and Social Psychology, 1950, 45, 442-458.

THORNDIKE, E. L., and LORGE, I. The teacher's wordbook of 30,000 words. New York: Teachers College, Columbia University, 1944.

THOULESS, R. H. Effect of prejudice of reasoning. British Journal of Psychology, 1959, 50, 289-293.

TRABASSO, T., and BOWER, G. H. Component learning in the four-category concept problem. Journal of Mathematical Psychology, 1964, 1, 143-169. (a)

TRABASSO, T., and BOWER, G. H. Memory in concept identification. Psychonomic Science, 1964, 1, 133-134. (b)

TRABASSO, T., and BOWER, G. H. Presolution dimensional shifts in concept identification: A test of the sampling with replacement axiom in all-or-none models. Journal of Mathematical Psychology, 1966, 3, 163-173.

TRABASSO, T., and BOWER, G. H. Attention in learning. New York: Wiley, 1968.

TUNE, G. S. A brief survey of variables that influence random generation. Perceptual and Motor Skills, 1964, 18, 705-710.

UHL, C. N. Learning of interval concepts: I. Effects of differences in stimulus weights. Journal of Experimental Psychology, 1963, 66, 264-273.

UNDERWOOD, B. J. An orientation for research on thinking. Psychological Review, 1952, 59, 209-220.

UNDERWOOD, B. J., and RICHARDSON, J. Some verbal materials for the study of concept formation. Psychological Bulletin, 1956, 53, 84-95. (a)

UNDERWOOD, B. J., and RICHARDSON, J. Verbal concept learning as a function of instructions and dominance level. Journal of Experimental Psychology, 1956, 51, 229-238. (b)

WALKER, C. M., and BOURNE, L. E., Jr. Concept identification as a function of amounts of relevant and irrelevant information. American Journal of Psychology, 1961, 74, 410-417.

WASON, P. C. The processing of positive and negative information. Quarterly Journal of Experimental Psychology, 1959, 11, 92-107.

WASON, P. C. On the failure to eliminate hypotheses in a conceptual task. Quarterly Journal of Experimental Psychology, 1960, 12, 129-140.

WASON, P. C. Response to affirmative and negative binary statements. <u>British Journal of Psychology</u>, 1961, <u>52</u>, 133-142.

WASON, P. C. The effect of self-contradiction on fallacious reasoning. <u>Quarterly Journal of Experimental Psychology</u>, 1964, <u>16</u>, 30-34.

WASON, P. C. The contexts of plausible denial. <u>Journal of Verbal Learning and Verbal Behaviour</u>, 1965, <u>4</u>, 7 - 11.

WASON, P. C. Reasoning. In B. Foss (Ed.), <u>New horizons in psychology</u>. Harmondsworth: Penguin, 1966.

WASON, P. C. "On the failure to eliminate hypotheses..." - a second look. In P. C. Wason and P. N. Johnson-Laird (Eds.), <u>Thinking and Reasoning</u>. Harmondsworth: Penguin, 1968. (a)

WASON, P. C. Reasoning about a rule. <u>Quarterly Journal of Experimental Psychology</u>, 1968, <u>20</u>, 273-281. (b)

WASON, P. C. Regression in reasoning? <u>British Journal of Psychology</u>, 1969, <u>60</u>, 471-480.

WASON, P. C., and JOHNSON-LAIRD, P. N. Proving a disjunctive rule. <u>Quarterly Journal of Experimental Psychology</u>, 1969, <u>21</u>, 14-20.

WASON, P. C., and JOHNSON-LAIRD, P. N. A conflict between selecting and evaluating information in an inferential task. <u>British Journal of Psychology</u>, 1970, <u>61</u>, 509-515.

WASON, P. C., and JONES, S. Negatives: Denotation and connotation. <u>British Journal of Psychology</u>, 1963, <u>54</u>, 299-307.

WEISS, R. L. On producing random responses. <u>Psychological Reports</u>, 1964, <u>14</u>, 931-941.

WELFORD, A. T. <u>Fundamentals of Skill</u>. London: Methuen, 1968.

WHITMAN, J. R., and GARNER, W. R. Concept learning as a function of form of internal structure. <u>Journal of Verbal Learning and Verbal Behaviour</u>, 1963, <u>2</u>, 195-202.

WHORF, B. L. <u>Language, thought, and reality</u>. New York: Wiley, 1956.

WILKINS, M. C. The effect of changed material on ability to do formal syllogistic reasoning. Archives of Psychology, 1928, No. 102.

WINER, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

WOLFF, J. L. Concept-shift and discrimination-reversal learning in humans. Psychological Bulletin, 1967, 68, 369-408.

WOODRUFF, P. N. Logic and truth value gaps. In K. Lambert (Ed.), Philosophical problems in logic. Leidel: Dordrecht, 1970.

WOODWORTH, R. S., and SELLS, S. B. An atmosphere effect in formal syllogistic reasoning. Journal of Experimental Psychology, 1935, 18, 451-460.