SOME HYPOTHESIS THEROY MODELS FOR PERFORMANCE IN CONCEPT LEARNING TASKS

Thesis for the Degree of Ph. D.
MICHIGAN STATE UNIVERSITY
Charles Ernest Kenoyer
1970

This is to certify that the

thesis entitled

SOME HYPOTHESIS THEORY MODELS
FOR PERFORMANCE IN CONCEPT
LEARNING TASKS

presented by

Charles Ernest Kenoyer

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Psychology

James L. Phillips
Major professor

Date August 14, 1970

O-169

ABSTRACT

SOME HYPOTHESIS THEORY MODELS
FOR PERFORMANCE IN CONCEPT
LEARNING TASKS

By

Charles Ernest Kenoyer

Recent research (Levine, 1966) has led to rejection of the sampling-with-replacement axiom. The procedure of the Levine study differed from that of the typical concept identification study in that blank trials were administered and the feedback that was provided on other trials was predetermined (fixed). A modified procedure was subsequently developed (Kenoyer and Phillips, 1968), in which feedback was fixed for early trials and no blank trials were used. Further evidence against sampling with replacement and for multiple-hypothesis processing was obtained with this modified procedure, which is like that of the typical concept identification study from the subject's point of view. The present study replicated the Kenoyer and Phillips study and extended it by including all combinations of fixed feedback over the first three trials. Several implications of the Restle and Bower-Trabasso models were tested in Experiment 1 by means of this procedure.

Levine's (1966) hypothesis theory assumes memory for the current hypothesis set following an error. A detailed model (Chumbley, 1969) within the framework of Levine's theory was tested in the present study, against data from Levine's study. Inadequate fit suggested a need for

additional models. Five models were presented, in which individual hypotheses are eliminated independently.

For Model 1, on each trial each hypothesis is eliminated with a probability that is determined by the trial outcome, "right" (R) or "wrong" (W), and the set of hypotheses that have been eliminated are retained perfectly. Models 2 and 3 assume the same hypothesis-elimination process assumed in Model 1, but also assume fallible memory for eliminated hypotheses. In both models, an elimination operator is applied to the probability that each hypothesis is in the current set, then a memory operator is applied to the probability that each hypothesis remains in the eliminated set. The memory operator is the same for every trial for Model 2, but depends upon the trial outcome (R or W) for Model 3. For Model 4, the operators of Model 3 are applied in opposite order, and Model 5 is obtained by reversing the order of the operators of Model 2.

All five models were tested against Levine's data. Models 1 and 3 were inadequate and were not tested further. Model 5 yielded acceptable fit by a chi-square criterion. Models 2 and 4 failed to meet the same criterion, but were retained for further testing against data from Experiment 2. It was conjectured that the most important form of loss from memory might differ for the two studies. The best-fitting model for Experiment 2 was Model 4. A suggested explanation for the difference was that Levine's use of blank trials introduces a long interval during which a constant forgetting process is important, while cognitive strain due to information processing should be the major cause of forgetting in the present study. Here the process should be affected by trial outcome.

Model 4, while clearly superior to Models 2 and 5 for these data, did not satisfy a chi-square goodness-of-fit criterion ($p < .001$). This measure of fit was computed for points on the mean learning curve and the trial-of-last-error (TLE) curve for eight experimental conditions, for a total of 104 data points, and so was extremely sensitive to deviations from fit. Although this test indicates that the model is not true, it was also found that the model accounts for 91 per cent of the variance among TLE points and 97 per cent of the variance among the mean learning curve points, over all eight experimental conditions.

SOME HYPOTHESIS THEORY MODELS FOR PERFORMANCE IN CONCEPT LEARNING TASKS

By

Charles Ernest Kenoyer

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Psychology

1970

G-65437
1-20-71

To Jan, Danny, Kathy, and Timmy, whose patience, love, and prayers were a crucial part of the total effort.

ACKNOWLEDGMENTS

The help and direction provided by the members of my guidance committee, Drs. A. M. Barch, J. F. Hanna, J. E. Hunter, and J. L. Phillips are gratefully acknowledged. I am grateful to Dr. Phillips for providing guidance without curtailing my freedom to make the thesis an expression of my own research interests. I wish to thank Dr. Hunter for the refinement and stimulation resulting from his incisive criticisms of theoretical concepts.

Also, I wish to express my appreciation for the use of the research facilities of the Human Learning Research Institute and those of the Michigan State University Computer Center.

Finally, to my wife, Jan, not only for her understanding and moral support through particularly trying times, but also for her very tangible help in proofing, my warmest thanks.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Several recent models of concept learning have described processes by which characteristics of the problem are assumed to be abstracted and used as a basis for classifying stimulus objects. Restle's (1962) cue learning model accounts for acquisition of such classification behavior in terms of random sampling of strategies from a hypothetical pool of strategies available to the subject, and subsequent testing and rejection of the selected strategies as classification information is provided by feedback on each trial. Bower and Trabasso's (1964) concept identification model explains acquisition in terms of random selection and testing of _cues,_ and is otherwise very similar to Restle's model. Later models (Trabasso and Bower, 1966, 1968; Levine, 1966) assume a process in which _hypotheses_ are manipulated. Levine (1966) and Richter (1965) have pointed out that the terms _strategy, cue,_ and _hypothesis_ are used in these models to refer to similar elements, and Levine (1967) has discussed the models under the more general heading, "hypothesis theory."

These models are applicable to situations in which the subject is required to learn to classify stimulus objects on the basis of characteristics that are already discriminable by the subject. The models are applicable to concept attainment, (Cf. Bruner, Goodnow, and Austin, 1956), concept utilization (Cf. Martin, 1965), or concept identification (Cf. Bower and Trabasso, 1964), but not to concept _formation._ The distinction between concept formation and the other terms listed above is that a new concept, i.e., one based on a characteristic of the stimulus not previously discriminated, is involved in concept formation, while previously discriminated characteristics are the basis of concept identification, utilization, or attainment tasks (Cf. Bourne, 1966, p. 3).

Concept identification differs from simple discrimination in that there are several stimulus characteristics that could serve as bases for classifying the stimuli, but only one characteristic leads to correct classification responses for a given problem. When it is of interest to establish the set of hypotheses from which samples are drawn, a list of the characteristics on which stimuli vary is sometimes provided for the subject. (Cf. Trabasso and Bower, 1966.) The stimulus qualities (e.g., color, size, etc.) that constitute potential bases for correct classification responses are called dimensions, cues, or attributes. The description of an individual stimulus in such problems comprises a value for each attribute (e.g., red, large, etc.). Solution of such problems can be indicated by a criterion run of correct responses or by a statement of the attribute value or combination of attribute values that determine correct classification.

Hypothesis theory provides a framework within which questions about concept identification may be formulated and tested in the kind of experiment described above. Variations in experimental procedure may therefore lead to new predictions within the theoretical framework, and so the theoretical framework serves to suggest a variety of ways of examining the process. The framework also provides a way to produce various specific models. By changing assumptions about memory, sampling of hypotheses, and response rules, it is possible to generate models that differ substantially although they are all formulated within the general framework. Such models can be compared to the data and to each other and the theory can be elaborated by choosing among models on the basis of these comparisons.

The mathematical models cited have typically been tested in a restricted class of experiments. In these experiments each cue takes on two values, the set of stimuli is presented in several random orders, the response set consists of two classification responses, problems continue to a learning criterion, classification is based upon a single dimension, and the classification rule is predetermined by the experimenter. Important questions of a preliminary nature have been examined in this rather restricted situation, but it is clearly desirable that a theory of concept identification be applicable to a broader class of situations. As Levine (1967) has pointed out, hypothesis theory is applicable to complex concepts (e.g., conjunctive or relational) as well as to the simple one-dimension concept.

Deviations from the constraints listed above have appeared recently in experiments designed to test the hypothesis model. Levine (1966) introduced a procedure in which subjects were informed of outcomes ("right" or "wrong") only on every fifth trial, beginning with trial 1. The subjects' hypotheses were inferred from sets of responses on the intervening "blank" trials (trials on which subjects were not informed of outcomes). The outcomes were determined arbitrarily, and the outcome sequence was used as an independent variable. Since it was necessary to control the amount of information provided on each trial, Levine did not randomize the stimuli, but organized them in a highly constrained sequence. In this situation the solution that is consistent with the information provided to the subject on outcome trials is jointly determined by the stimuli, responses, and outcomes on the outcome trials. Since the responses are not under the experimenter's control, neither is the solution, and so the solution is a dependent variable. Levine also ran subjects for a fixed number of trials, rather than to criterion.

A procedure that represents a compromise between Levine's paradigm and the more common experiment in concept identification was used by Kenoyer and Phillips (1968) to test assumptions of the hypothesis models. Arbitrary outcomes were administered on the first three trials. The solution that was consistent with the information provided on those trials was then the basis for outcomes on later trials. Trabasso and Bower (1966) also used a procedure in which the classification rule was determined jointly by the stimuli, responses, and outcomes, in order to test an assumption of their concept identification model. Although these experiments differ considerably in procedure, they are all relevant to assumptions about the processing of hypotheses in a problem requiring the identification of a classification rule. Emphasis on different aspects of the theory, hypothesized process, or experimental paradigm has led investigators to refer to experiments of this type as discrimination (e.g., Levine, 1966) concept identification, (e.g., Bower and Trabasso, 1964) cue learning (Restle, 1962), or concept attainment (e.g., Haygood and Bourne, 1965). No attempt will be made here to review the work in all these areas, since many studies would not deal with the theoretical issues of interest in the present study. A review of work in any one of the areas would both include irrelevant studies and exclude relevant ones.

ISSUES IN HYPOTHESIS THEORY

The Hypothesis as a Construct

Krechevsky (1932) reported that rats performing in a discrimination experiment displayed strong positional preferences at the outset of the experiment, and referred to such preferences as hypotheses (Hs). This designation amounted to a behavioral or operational definition of a word that had already acquired meaning in everyday English. It was perhaps on this account that Spence (1940) objected to this use of the term. He argued that such perseverative tendencies were not adaptive and that they would, in fact, retard learning. His objection to applying the term "hypotheses" to such tendencies thus seems to have been based upon positive connotative meaning already associated with the term.

Harlow (1959) has subsequently developed a theory in which learning is taken to be a process of inhibiting error factors, which are the same kind of maladaptive behavioral tendencies as Krechevsky's Hs. The nonrandom nature of the naive subject's behavior at the outset of a discrimination problem (error factor) and the nonrandom choice behavior at the outset of a transfer problem (learning set) are quite different in terms of their adaptiveness, but may be considered as hypotheses which happen to vary in their appropriateness to the performance criteria defined by the experimenter. Harlow and his associates have investigated these phenomena extensively in primates.

Levine (1963) studied hypotheses (in Krechevsky's sense) in human subjects. In the first of two related experiments, he distinguished two kinds of response tendencies. One kind was uncorrelated

with cues, and consisted of identifiable patterns of responding, such as alternation. Levine designated these response tendencies "Response-sets." The other kind of patterns were called "Predictions." These patterns displayed regularity with respect to the stimulus set. One prediction pattern is "win stay, lose shift." A subject displays such a pattern with respect to a given cue, such as color. If the subject displayed a strong tendency to shift his choice to the opposite color after an error and to repeat his color choice after a correct response, he was said to follow this prediction pattern. Four cues were varied in the experiment, and each subject performed in 90 two-trial problems. Within each problem, either of the two possible responses would be a repetition with respect to some cues and a shift with respect to others. Levine performed an involved analysis of conditional response probabilities over the whole problem set, however, and found reliable prediction patterns. He also concluded from this analysis that response sets contributed little or nothing to performance.

In Experiment II, therefore, he directed his attention to further analysis of prediction behavior. He administered 24 multiple-cue discrimination problems to two groups. Color hypotheses were correct for the first 12 problems and letter hypotheses for the last 12 problems. Every fourth problem (problems 2, 6, ..., 22 for one group and problems 4, 8, ..., 24 for the other) was a test series of four trials. Subjects were not informed of outcomes on these trials, and stimuli were organized so that every possible response sequence on the four trials was inconsistent with all but one of the eight hypotheses. Half of the possible patterns were not consistent with any hypothesis. All problems other than the test problems were 14 trials long, and subjects were informed of outcomes.

Levine combined response sequences corresponding to each value of a cue. For example, responses consistent with the hypothesis "large" and those consistent with "small" were combined and called size hypotheses. He plotted the proportion of each of these cue hypotheses over test problems. The graphs of probabilities of all hypotheses showed that the probability of a color hypothesis increased over the first twelve problems, on which color was correct, then suddenly decreased after the thirteenth problem, on which the solution was changed to letter. The probability of a letter hypothesis remained low until after the thirteenth problem, then increased quickly to an asymptote around .5. It was clear that hypotheses were being held over from one problem to the next, and were therefore involved in trials at the outset of some of the problems.

Supporting evidence for hypotheses at the outset of learning was provided by a later experiment (Kenoyer and Phillips, 1968), in which outcomes ("right" or "wrong") were arbitrarily set for the first three trials, rather than depending upon the subject's response as is usually the case. There were eight possible hypotheses (classification rules) which were listed for the subjects. The stimuli for the first three trials were so related that, for a given string of outcomes, a unique hypothesis was consistent with each possible sequence of three choice responses. On subsequent trials, outcomes were consistent with the hypothesis determined on the first three trials. In one of the treatments, subjects were told "right" after each of the first three responses (the RRR treatment condition). If a subject began with a hypothesis, therefore, errorless performance was to be expected, since the procedure "tracked" any such hypothesis over trials 1-3. If subjects were responding randomly until an error occurred, however, as Bower and Trabasso's (1964) model specifies, the probability of errorless performance on the remaining six trials of a problem would be $(1/2)^6 = 1/64$. The observed proportion of correct responses for the RRR treatment condition was .976. It is clear that subjects were processing information at the outset of the problems, and Levine's conclusion that human subjects employ hypotheses at this stage of a problem was supported by this result.

It is important to note, however, that the behavioral indicator in this case was performance on subsequent trials, and so the conclusion pertains only to <u>information gain</u> on trials 1-3, rather than to hypothesis <u>behavior</u> on those trials. The term "hypothesis" in the study cited just above, refers to a theoretical construct rather than to a response sequence, as in Levine's (1963) use of the term.

Levine's (1963) results suggest that subjects display hypothesis behavior at the outset of a problem. In both of the experiments reported in that study, however, early trials constituted the whole test series. In the first experiment there were 90 problems of two trials each, and in the second the hypothesis data were obtained on test problems of four trials each. The test problems were identified as tests in the instructions to the subjects. More recent evidence indicates that these special circumstances may have caused subjects to behave somewhat differently than they would have done in a more extended task. Chumbley (1969) gave subjects four initial training trials, on which outcome information was provided, followed by seven test trials without outcomes. Chumbley obtained a good fit of his Hypothesis Manipulation (HM) model to the test-trial data, but stated that it could not be fitted to the training-trial data. He found that the probability of a right-hand button-press was higher than chance. This result is consistent with the assumption that subjects in this situation behave according to response sets, in the sense defined above. Another finding prevents this conclusion, however. The tasks were experimenter-paced, and so subjects who did not respond on schedule simply had a trial without a response. Some of the subjects did not respond at all on training trials but responded without error on test trials. Chumbley concluded that his instructions had led subjects to emphasize test-trial behavior to the exclusion of meaningful choice behavior on training trials. Although some kind of effective problem-solving process during training trials was indicated by test-trial performance, hypotheses were not evident from training-trial data. The definition of "hypothesis" as a theoretical construct used to explain organization of subsequent behavior is therefore not consistent, in the context of Chumbley's study, with the definition of the term as a pattern of responses. Throughout the remainder of this paper, "hypothesis" will refer to the theoretical construct unless otherwise specified. The usefulness of such a notion in organizing findings about concept identification and discrimination, both within and between problems, is evident from the above discussion. The models to be discussed below represent the problem-solving process as generation (or selection) and testing of hypotheses. They differ, however, in their assumptions about the nature of the selection process.

Memory

An important characteristic of a hypothesis model is the amount of memory that is assumed. Restle (1962) developed three alternative models. The alternative processes for selecting hypotheses were selection of one at a time, all at once, and *n* at a time. Restle showed that the three models were alike in their predictions on error data. The memory assumption of each model was pivotal in the derivations of the error predictions, however, and so Restle's proof did not establish that single-hypothesis models and multiple-hypothesis models are indiscriminable in general, even with respect to error data. The equivalence was established for Restle's three specific models, with their assumptions of severely limited memory.

Restle assumed sampling of hypotheses with replacement in the one-at-a-time model. For the all-at-once model the subject was assumed to consider all hypotheses at the beginning of the task. This model assumes that response probabilities are determined by the proportion of the strategies consistent with each response. The hypothesis set is assumed to be partitioned on the basis of consistency with the classification response and those in the inconsistent set are assumed to be dropped (forgotten) from the set being considered. A correct response

occurs on those occasions when the correct hypothesis is in the consistent set, i.e., the set that is retained. Occurrence of an error is possible only when the correct hypothesis is in the discarded set.

Since the subject is not assumed to be able to retrieve these hypotheses without starting again with the total hypothesis set, an error implies that the subject has the full set to work with, just as at the outset of the problem. The _n_-at-a-time model requires an additional sampling assumption, and Restle chose to assume that all subsets of size _n_ were equally likely to be selected. The multiple-hypothesis models are similar in all other respects. The restarting property, which implies that the subject is in the same state of ignorance after each error as at the beginning of the problem, is therefore common to both multiple-hypothesis models as well as to the single-hypothesis model.

Bower and Trabasso (1964) developed a model that was mathematically equivalent to Restle's one-strategy model, except for their added assumption that subjects begin problems in a guessing state and continue to guess until an error occurs. The selection process assumed in this model operates upon cue values rather than strategies, however. Since the model assumes that the subject deals with only one cue at a time, hypotheses based upon two or more cues are excluded from consideration.

A later model (Trabasso and Bower, 1968) assumes multiple hypotheses, and is quite similar to Restle's (1962) _n_-at-a-time model. This model assumes that a "focus sample" of size _s_ is selected from the stimulus array. Sampling probabilities are assumed to be controlled by

cue salience. The focus sample is assumed to be reduced after each correct response, as in the Restle model, and after each error a new focus sample is assumed to be selected with replacement. Response probabilities are generated as in the Restle model.

Levine (1962) and Holstein and Premack (1965) provided random outcome information to subjects for a given number of trials, where the number of trials varied over experimental conditions. Random outcome trials were followed by a discrimination problem. The finding that random outcomes retarded solution of the discrimination problem is inconsistent with the sampling-with-replacement assumption. The amount of retardation was constant over variations in the number of random-outcome trials.

Restle and Emmerich (1966) performed three related experiments in which they investigated memory in a concept identification situation. In the first experiment, four groups of subjects were given one problem at a time or two, three, or six problems concurrently, i.e., with trials for one problem interspersed with trials from another problem or problems. Learning was faster in the groups that had one or two concurrent problems than in the groups with three or six problems. They pointed out that this result was in conflict with two hypothesis models (Restle, 1962; Bower and Trabasso, 1964). The break between two and three problems was interpreted as evidence that the memory span was overloaded with nine stimulus dimensions (three per problem) but not with six. The authors argued that it must be memory for stimuli, rather than hypotheses, that was breaking down in the multiple-problem condition, thus indicating that they could remember the correct hypothesis. This conclusion does not follow from the data, however. The process described by Levine (1966) would imply a considerable memory load on early trials, but less as the hypothesis sets were reduced, and finally only one hypothesis per problem.

Experiment 2 of the Restle and Emmerich study compared a condition in which the stimulus remained available to subjects after feedback with a condition in which the stimulus disappeared before feedback. The two levels of the stimulus availability variable were arranged factorially with number of problems. Subjects solved either one problem or six problems concurrently. Stimulus availability reduced errors for the one-problem group, but not for the six-problem group. Restle and Emmerich pointed out that the effect on the one-problem group was consistent with stimulus memory, but also with hypothesis memory, since the presence of the stimulus could be used to limit the hypothesis set from which the subject sampled. They offered no explanation for the lack of effect on the six-problem group. Erickson and Zajkowski (1967), however, suggested that concurrent problems lead to interference with short-term memory of hypotheses that have been tested but rejected. If this were the case, it would be reasonable to expect subjects to adopt a strategy requiring no memory for rejected hypotheses when performing in the concurrent problem condition. If subjects conformed to the Restle (1962) model or, equivalently, to the Bower and Trabasso (1964) model in that situation, they would need to remember only the current hypothesis for each problem and in the group without stimulus availability, the stimulus. Five hypotheses would then have to be remembered while the subject processed information leading to selection of a hypothesis in the current problem. The cue values for the different problems were quite dissimilar, however, and so interference should not be great. Furthermore, the error probabilities would be unaffected by such interference unless the whole set of cue values were forgotten, since only one hypothesis is assumed to be retained. Selecting (i.e., remembering) one cue value randomly and basing the hypothesis on it is equivalent to remembering two or three cue values and randomly selecting one of these as the basis for a single hypothesis. It is reasonable to assume that subjects in a single-problem situation retain information from past trials (about either stimuli or hypotheses), but have too little available memory to do so in the six-problem condition.

Further information about memory in concept identification was reported by Trabasso and Bower (1966), who tested the sampling-with-replacement assumption of their previously published model (Bower and Trabasso, 1964) with a rather complex experimental procedure. For one group, the correct choice responses could be based on either of two characteristics of the stimulus. For example, size and color could be redundant, so that choosing the large object would be behaviorally equivalent to choosing the red object, and either would be correct. For a second group, the same two cues (e.g., size and color) were treated as follows. A problem began with only one of these cues relevant. When the subject made an error, he was informed of it, and if he made no further errors he proceeded quickly to criterion and solved the problem. A second error, however, was treated differently. The subject was not informed of the error. Instead the criterion for a correct response was changed, e.g., from large to red. In shifting the criterion from size to color, the specific color to be associated with the correct response was selected so as to be consistent with the trial on which the subject was _informed_ of an error. This treatment was called a "dimensional shift." On every second error in this group, the correct response criterion was shifted to the other of the two cues, or dimensions.

The Bower and Trabasso (1964) model assumes that subjects solve such problems by selecting a single cue value (such as red) after each error, without regard to whether the cue value has been tested previously. Under this assumption the advantage of having two redundant and relevant cues is that there are two chances to select a correct cue instead of just one. In the experiment just described, however, the model implies that the same advantage accrues to the subjects in the dimensional shift group, given the sampling-with-replacement assumption. After an error they may select the currently relevant one and solve or they may make a response that is not consistent with that cue, and still be given an opportunity to solve on the alternate cue. The probability of solution after an informed error should therefore have been the same for both groups. Trabasso and Bower found, however, that the dimensional shift task was the more difficult. They suggested a new model in which cues could not be resampled until some number, $\underline{k}$, of trials after it had been tested and rejected. Such a model, they noted, would account for the results reported by Levine (1962) and Holstein and Premack (1965) as well as those of their own study.

Levine (1966) tested the replacement axiom with a different experimental procedure. On the first trial the outcome information was provided to the subject. Four trials followed on which no outcome information was given. Three such blocks were given, followed by a final (sixteenth) trial on which the outcome was given. There were therefore only four outcome trials in the series. The stimuli on the outcome trials and those within test-trial blocks were "internally orthogonal." An important characteristic of such stimulus sequences is that any response sequence that correlates perfectly with one cue is uncorrelated with all other cues. Since test blocks were arranged in this way, response sequences could be analyzed to determine what hypothesis, if any, the subject was tracking.

Levine estimated the size of the hypothesis set from the probability of selecting one of the hypotheses consistent with the outcome-trial stimulus. Under the sampling-with-replacement assumption, the size of the hypothesis set should remain the same throughout the experiment. If subjects had been perfect information processors, the set should be reduced by half after each outcome trial. The obtained curve fit neither of these models perfectly, but was considerably closer to the curve for perfect processing. As in the Trabasso-Bower (1966) study, it was clear that the sampling-with-replacement axiom was inconsistent with the data.

Since all of the sampling schemes that imply the restart-after-errors principle are falsified by the results cited above, some kind of memory assumption is needed, and so the nature of what is remembered becomes important as well as the amount. Trabasso and Bower (1966) suggested that their earlier single-hypothesis model be modified by adding two assumptions dealing with two distinct kinds of memory. In the resulting model, subjects are assumed to remember rejected hypotheses for _k_ trials, where _k_ is a free parameter of the model. After _k_ trials a rejected hypothesis is assumed to be returned to the hypothesis pool.

The second kind of memory that was assumed dealt with stimulus information. On error trials it was assumed that the subject performed a consistency check, comparing stimulus information from the error trial with that from the preceding trial.

Use of all of the stimulus information provided on an error trial to limit hypothesis selection is called local consistency (Gregg and Simon, 1967; Trabasso and Bower, 1968). The Bower and Trabasso (1964) model has this property (Atkinson, Bower, and Crothers, 1965, p. 32), as do two more recent multiple-hypothesis models (Trabasso and Bower, 1968) and the model cited just above. The Trabasso and Bower (1966) model further assumes consistency over the trial preceding the error trial, but all of the models just cited have in common at least consistency with the error-trial information.

Kenoyer and Phillips (1968) tested the local consistency assumption in an experiment in which complementary pairs of stimuli were presented. Each cue (size, color, shape, and border) had two values. The alternative values of the cues were called complements. Red was thus the complement of blue, square was the complement of circle, large the complement of small, and presence of a border was the complement of absence of a border. Two stimuli were a complementary pair if the value of every cue in the first stimulus (C_1) was the complement of the value of the corresponding cue in the second stimulus in the pair (C_2). The first member of the pair was always presented before the subject had been given enough information to solve the problem, and the outcome on that trial was arbitrarily a W.

Assuming local consistency, the subject would select some cue and would make his category assignment agree with the correct category assignment on the error trial. If the subject classified C_1 as VEK, combining this response with the W outcome would result in a correct classification of NONVEK for that stimulus. Then regardless of which cue the subject selected after the error, the cue value present in C_1 would be assigned to NONVEK. When C_2 appears, the opposite value of that cue (and all other cues) is present, and the local consistency assumption implies that the subject must assign it to the VEK category. Thus the category assignment of C_2 must match that of C_1, according to the local consistency assumption. If no information is processed after a correct response, as the single-hypothesis models imply, this prediction on matches holds regardless of the number of trials intervening (the lag) between the trials on which C_1 and C_2 are presented, given that the responses on these trials are all classified as correct.

The multiple-hypothesis models developed by Trabasso and Bower (1968) assume processing after correct responses. Responses are assumed to be consistent with all hypotheses not yet eliminated from the sample, however, and this implies that one of the hypotheses consistent with the error-trial information is retained until another error occurs. Thus the C_1 to C_2 lag is unimportant to the match prediction within these multiple-hypothesis models as well as in the single-hypothesis models. Kenoyer and Phillips found that the probability of a match was not near 1, in general, as implied by the models. What is remembered immediately after an error cannot be determined with certainty from this result, but the complete stimulus-response-outcome information does not remain available over a series of correct trials. This result does not completely isolate the local consistency assumption, since some kind of forgetting process could be posited to account for the loss over trials.

Experiment 3 of the Restle and Emmerich (1966) study, cited previously, was more directly relevant to the local consistency assumption. Subjects were given one or six concurrent problems, and the same stimulus was presented on two consecutive trials, both very early and very late in the problem. On late trials, the probability of an error on the second presentation following an error on the first presentation was near chance (1/2). On early trials, for the one-problem group, three of the 61 subjects who made correct responses on the first presentation and 3 who made errors on the first presentation, made errors on the second presentation. This result simultaneously refutes the local consistency assumption and the assumption that the process restarts after errors without local consistency. The former assumption implies that the probability of a correct response on the second presentation following an error on the first is 1 and the latter implies that it is 1/2. Subjects in the six-problem condition made 8 errors following 62 correct responses and 19 errors, following 69 errors. Memory was less effective for this group than for the one-problem group, and less effective after errors than after correct responses. This result on early trials, like the corresponding data on the one-problem group, refutes both local consistency and restarting assumptions. It is consistent with Levine's (1966) theory, however. The overall results of this experiment may be explained by assuming that subjects processed multiple hypotheses and tried to keep track of rejected hypotheses on early trials, but changed strategies when they failed to solve and began processing single hypotheses and sampling with replacement.

Multiple hypotheses

In addition to providing evidence on the sampling-with-replacement question, Levine's (1966) experiment also yielded data relevant to the question of multiple versus single hypotheses. Since he was able to manipulate outcome sequences as an independent variable, Levine could compare sequences with different numbers of errors in terms of their effect on subsequent performance. He compared a one-error condition (RRW), a two-error condition (RWW and WRW), and a three-error condition (WWW). The dependent variable was probability of a correct hypothesis after trial 3, a correct hypothesis being defined as the one hypothesis that was consistent with the information provided to the subject on all three outcome trials. Levine found that the probability of a correct hypothesis was an increasing function of the number of correct (R) outcomes. It is evident from these data that subjects were processing information on R trials. If subjects processed only one hypothesis at a time, no information about that hypothesis would be provided on correct trials. Since each of the sequences Levine compared ended with a W, differences among probabilities for the three groups constitute further evidence that the problem-solving process does not simply restart after errors.

Richter (1965) presented subjects with a series of four-trial problems, in which the stimulus sequence was structured like those of the Levine (1963, 1966) experiments, and so logical solution of the problem was possible after three trials. The probability of a correct <u>response</u> on trial 4 was therefore comparable to the probability of a correct <u>hypothesis</u> after trial 3 in the Levine (1966) study. Richter used predetermined solution rules rather than fixed outcomes. He found that probability of a correct response on trial 4 was an increasing function of the number correct on trials 1 through 3.

Erickson, Zajkowski, and Ehmann (1966) and Erickson and Zajkowski (1967) found evidence for multiple-hypothesis processing in latency data from concept identification experiments. In both studies a post-criterion decrease was found. Pre-criterion latencies were analyzed separately for trials following errors and correct responses. Latencies following correct responses clearly decreased over pre-criterion trials. Results on latencies following errors were equivocal. For one analysis the median latency was computed for the first and last halves of pre-criterion trials following errors, and the means of these median latencies were compared. The mean for the last half was greater than that for the first half. A regression line on trials however, showed a slight negative slope. The post-criterion decrease in latency suggests processing of multiple hypotheses. If hypotheses are processed on correct trials as well as on error trials, solution is possible on correct trials and the post-criterion decrease can be explained by reduction of the hypothesis set after the last error. The pre-criterion decrease in latency indicated by the regression of latency on trials can also be explained in terms of multiple hypotheses. If the number of hypotheses being processed is reduced after an error, then the time required to process them should decrease.

The evidence for a multiple-hypothesis process is convincing, but the memory assumptions of the multiple-hypothesis models developed by Restle (1962) and Trabasso and Bower (1968) are inadequate on other grounds, as was stated above. Levine's (1966) multiple-hypothesis theory is similar to the Restle _n_-at-a-time model, but the memory assumptions are different. As in the Restle model, the hypotheses consistent with the classification responses are assumed to be retained. The treatment of the hypotheses discarded on that trial, i.e., those inconsistent with the classification response, differs for the two models. They are lost, according to the Restle model, to be recovered only by starting again with the whole hypothesis set. The assumption in the Levine theory is that these hypotheses can be retrieved, although with some difficulty. The difficulty of retrieval of this set of hypotheses provides an explanation for the decreased effectiveness of information processing on error trials as compared with correct trials.

Levine's Hypothesis Theory

The Levine theory includes none of the assumptions that are rejected by the above arguments. It assumes that subjects begin a problem with hypotheses rather than in a guessing state. It assumes that multiple hypotheses are processed, although only one hypothesis is assumed to be the basis of each response. Since it assumes that hypotheses, rather than specific stimulus information are remembered, it does not imply local consistency. The proposition that the solution process restarts after each error is neither assumed nor implied by the theory.

The assumed reduction of the hypothesis set after each trial on which information is provided implies, in the usual concept identification situation, an increase over trials in the probability of selecting the correct hypothesis as a basis for responding. Since this assumption is contrary to the restarting-after-errors property that has been supported by previous research, it requires further discussion. The increase in the probability of solution over trials defines an inhomogeneous Markov process (Cf. Atkinson, Bower, and Crothers, 1965). Stationarity of the probability of a correct response when the subject is in the pre-solution state, however, does not depend upon homogeneity of the probability of solution. If solution has not occurred, Levine's theory holds that some other hypothesis being entertained by the subject determines choice responses. If the cue values corresponding to hypotheses are varied independently, the hypothesis that determines the choice response brings about chance responding. Actually, as Restle (1962) noted, not all hypotheses are independent of the correct one. The occurrence of the complement of the correct cue value is completely redundant (perfectly correlated) with the nonoccurrence of the correct cue value. But this implies that the complement of the correct cue value is likely to be eliminated early. The remaining hypotheses have the required independence property. Given a reasonably large initial set of hypotheses, the probability of an error would not be greatly affected by this nonindependence and the hypothesis that always leads to a wrong response would tend to be eliminated early in the problem. The probability of an error prior to the last error should therefore decrease only slightly over trials as a result of eliminating the complement of the correct hypothesis. A slight decrease in the probability of an error is consistent with reported results, in fact (Trabasso, 1966, p. 45; Bower and Trabasso, 1964), although the decrease has not been found to be significant.

Levine did not explicitly state an assumption that all hypotheses are equally likely to be selected, but he estimated the size of the active hypothesis set as the reciprocal of the proportion of correct hypotheses. This estimation procedure suggests that the equal-likelihood assumption was intended, and it is therefore treated here as part of the theory. The mechanism for retrieving hypotheses was also left unspecified in Levine's outline of his theory. Some specific assumptions about this process are needed if the theory is to be tested.

Chumbley's Hypothesis Manipulation Model

Chumbley (1969) presented a Hypothesis Manipulation (HM) model based upon Levine's theory. In this model, the current set of hypotheses is partitioned into two subsets by the subject's choice response. The subset that is consistent with the choice response is retained and if the response is correct, the current hypothesis set for the next trial has been reduced. The subset that is not consistent with the choice response is discarded. If the response is called "wrong", the discarded hypotheses are the proper ones to retain as the new current set. The HM model assumes that the subject retrieves the discarded set (as a whole) with probability t. Otherwise the entire set is lost, and the subject begins again with the whole initial hypothesis set.

Chumbley performed an experiment in which the problems consisted of four training trials followed by seven test trials. Treatment groups solved either one problem or three concurrent problems and had either a 5-sec. or a 15-sec. intertrial interval. The parameter t was estimated separately for each of the four conditions. The HM model fitted the data from the test trials, but not the training-trials data. One puzzling result on training trials was a higher than chance occurrence of a right-hand button press. A second result was even more striking. The trials were experimenter-paced, and so it was possible to sit through training trials without responding, and without any loss of information. Chumbley found that some subjects did not respond at all on training trials but performed without error on test trials.

Chumbley suggested that this discrepancy between model and data was due to a procedural artifact. He claimed that test-trial performance was emphasized to the detriment of meaningful performance on training trials, and that the model was therefore not necessarily wrong, but should be tested in an experimental situation from which this artifact is absent. It seems appropriate, therefore, to test the HM model against data reported by Levine (1966).

Test of the Hypothesis Manipulation Model

Chumbley's parameter, t, is the probability that the set is retrieved and retained until the next set reduction operation. If the hypotheses are not retrieved, the assumption is that the subject must start over with the initial hypothesis set. When the stimulus sequence is internally orthogonal, as in the Levine (1966) study, the model states that half of the hypothesis set is discarded. The current set is reduced to half its former size after a correct trial in any case. After an error trial, this reduction occurs only if the discarded set is successfully retrieved, i.e., with probability $\underline{t}$. If the current set is not reduced, it is replaced by the full initial set. Thus, if there are two hypotheses in the current set on an error trial, the set is either reduced to one hypothesis or replaced by the initial set of (typically) eight hypotheses. With an initial set of eight hypotheses, then, every subject must have either four or eight hypotheses after trial 1.

If we define a Bernoulli random variable x_i such that x_n=1 when a tenable hypothesis is selected after an error on trial n, and x_n=0 otherwise, we have for the WWW condition:

$$E(\bar{x}_i) = Pr\ (x_i=1) = Pr\ (\text{tenable H is selected} \mid r \text{ Hs remain}).$$

$$\text{Then } E(\bar{x}_1) = Pr\ (\text{tenable H is selected} \mid 8 \text{ Hs remain}) \cdot (1-t) + Pr\ (\text{tenable H is selected} \mid 4 \text{ Hs remain}) \cdot t$$

Since four Hs are tenable after trial 1, the probability of selecting one of them is simply four divided by the total number of Hs remaining, and:

$$E(\bar{x}_1) = \frac{1-t}{2} + t = \frac{1+t}{2}$$

$$E(\bar{x}_2) = Pr\ (\text{tenable H is selected} \mid 8 \text{ Hs remain}) \cdot (1-t)$$
$$+ Pr\ (\text{tenable H is selected} \mid 4 \text{ Hs remain}) \cdot t(1-t)$$
$$+ Pr\ (\text{tenable H is selected} \mid 2 \text{ Hs remain}) \cdot t^2$$
$$= \frac{(1-t)}{4} + \frac{t(1-t)}{2} + t^2 = \frac{2t^2 + t + 1}{4}$$

$$E(\bar{x}_3) = \text{Pr (tenable H is selected} \mid \text{8 Hs remain)} \cdot (1-t)$$
$$+ \text{Pr (tenable H is selected} \mid \text{4 Hs remain)} \cdot t(1-t)$$
$$+ \text{Pr (tenable H is selected} \mid \text{2 Hs remain)} \cdot t^2(1-t)$$
$$+ \text{Pr (tenable H is selected} \mid \text{1 Hs remains)} \cdot t^3$$
$$= \frac{(1-t)}{8} + \frac{t(1-t)}{4} + \frac{t^2(1-t)}{2} + \frac{t^3}{} = \frac{4t^3 + 2t^2 + t + 1}{8}$$

Expressions may be derived similarly for sequences other than WWW. The Chumbley model assumes no loss of information on correct trials. Again referring to the Levine study, the model predicts four hypotheses remaining after an initial "right" reinforcement, two hypotheses remaining after the subject is told "right" on trials 1 and 2. Then for the RRW condition,

$$E(\bar{x}_1)=1,$$
$$E(\bar{x}_2)=1, \text{ and}$$
$$E(\bar{x}_3)=\frac{1-t}{8} + t = \frac{7t+1}{8}$$

for the RWW condition,

$$E(\bar{x}_1)=1$$
$$E(\bar{x}_2)=t + \frac{1-t}{4} = \frac{3t+1}{4}$$
$$E(\bar{x}_3)=t^2 + \frac{t-t^2}{4} + \frac{1-t}{8} = \frac{6t^2 + t + 1}{8}$$

and for the WRW condition,

$$E(\bar{x}_1) = t + t\,\frac{1-t}{2} = \frac{t+1}{2}$$
$$E(\bar{x}_2) = t^2 + \frac{t-t^2}{2} + \frac{1-t}{8} = \frac{4t^2 + 3t + 1}{8}$$

The remaining task is to obtain a distribution so that an appropriate test of fit may be applied. Since the x_i are Bernoulli random variables, the number of tenable hypotheses on any one problem is a sum of Bernoulli random variables over subjects. Assuming subject independence, the sum over subjects is a random variable, y_i, with a binomial distribution. The probability of tenable hypotheses on the _i_th problem, $\bar{x}_i$, estimates the parameter _p_ of the binomial distribution. Given Levine's sample of 80 subjects, the distribution of the mean is closely approximated by the normal. Now if _t_ is assumed to remain constant over problems, the distribution of the random variable y is identical on all problems within an outcome-defined condition. If interproblem independence is assumed, then the mean over problems is the mean of independent, identically distributed random variables. Two implications from the Central Limit Theorem are that the distribution of this mean approaches the normal as the number of problems over which the mean is taken increases, and that the variance of the sample mean is inversely proportional to the number of problems (Cf. Parzen, 1960). For the analysis at hand it is important to note simply that the variance of the mean is _less than_ that of any one of the variables averaged. The deviation of an observation of $\bar{y}$ from the population mean μ_y, is approximately normally distributed with mean 0 and variance less than the variance of the binomial variable, y. A test of fit to $\bar{y}$ based upon the binomial distribution of y is therefore a conservative test, in the sense that a deviation of a given size is more probable in the distribution of y, due to its larger variance.

The test is not necessarily conservative if interproblem independence does not hold. The variance of the mean of two random variables is given by:

$$\mathrm{Var}\ (\frac{w+z}{2}) = \frac{\mathrm{Var}(w+z)}{4} = \frac{\mathrm{Var}(w) + \mathrm{Var}\ (z) + 2 \cdot \mathrm{Cov}(w,z)}{4}$$

If w and z are independent, this reduces to:

$$\frac{\mathrm{Var}(w) + \mathrm{Var}(z) + 0}{4}$$

Assuming identical distributions, we have:

$$\frac{\mathrm{Var}(w+z)}{2} = \frac{2 \cdot \mathrm{Var}(w)}{4} = \frac{\mathrm{Var}(w)}{2} = \frac{\mathrm{Var}(z)}{2}$$

If the covariance is negative rather than zero, the variance of the mean is even smaller. If the covariance is positive, however, the variance of the mean is greater than indicated above, where the covariance is assumed to be zero. When the covariance is positive, the variance of the mean is:

$$\mathrm{Var}\ (\frac{W+z}{2}) = \frac{\sigma_W{}^2 + \sigma_z{}^2 \quad + 2 \cdot \mathrm{Cov}\ (w,z)}{4}$$

$$= \frac{\bar{\sigma}^2 + \mathrm{Cov}(w,z)}{2} = \frac{\bar{\sigma}^2 + \rho_{wz}\ \sigma_w\ \sigma_z}{2}$$

If $\sigma_w = \sigma_z$, $\mathrm{Var}\ (\frac{w+z}{2}) = \frac{\sigma^2 + \rho_{wz}\ \sigma^2}{2} = \frac{\sigma^2(1+ \rho_{wz})}{2} \leq \sigma^2 = \bar{\sigma}^2$

Otherwise, $\mathrm{Var}\ (\frac{w+z}{2}) = \frac{\bar{\sigma}^2 + \rho_{wz}\ \sigma_w\ \sigma_z}{2} \leq \frac{\sigma^2\ (1+ \rho_{wz})}{2} \leq \bar{\sigma}^2$

This last inequality holds because, for a fixed sum $\sigma^2_w + \sigma^2_z$, the product $\sigma_w^2 \cdot \sigma_z^2$, and therefore $\sigma_w \cdot \sigma_z$, is maximized when $\sigma_w = \sigma_z$. Then regardless of the equality of σ_w^2 and σ_z^2, we have:

$$\mathrm{Var}\ (\frac{w+z}{2}) \leq \sigma^2$$

In words, the variance of an average of two random variables is no greater than the average of their variances. This principle clearly can be extended to more than two random variables.

'Frequencies' were obtained from the reported proportions by multiplying by the number of subjects (80). The sampling distribution of these quantities, according to the above argument, have variances less than or equal to those of the corresponding binomial distributions of the scores for occasions, over which they are averaged. For a binomial distribution with N=80, the chi-square statistic is distributed approximately as chi-square. The expected frequencies generated from Chumbley's HM model were therefore compared to the data from Levine's study by means of the chi-square test.

The procedure was as follows: Trial values of the parameter (t) of the model were used in a Fortran program to generate expected proportions (i.e., probabilities) of tenable hypotheses. The observed proportions used were those reported by Levine (1966). Three Pearson chi-square statistics were computed from these observed and expected proportions. The parameter value selected was the one for which the sum of the three chi-squares statistics was a minimum. The procedure therefore differs from minimum chi-square techniques in that a different criterion (the sum of three chi-squares) was minimized. Each of the chi-square values was computed on a pair of frequencies. One of each pair was the frequency of a consistent hypothesis after a W on trial 1. The other was the frequency of a correct hypothesis after trial 3 for the one-error, two error, or three-error condition. Since different expected frequencies follow from WRW and RWW, these were averaged to yield the expected frequency for the two-error condition. For WWW the chi-square value was 8.05, for RRW it was 11.09, and for WRW-RWW, 8.91. The value of the parameter <u>t</u> selected in this way was .49. If two degrees of freedom are assumed for each chi-square, each is significant beyond the .025 level. The fit of the HM model is therefore unsatisfactory by this criterion.

The criterion just described is somewhat conservative since it does not reduce the degrees of freedom for the estimated parameter $\underline{t}$. A more stringent test of the model may be devised by using the sum of the three chi-square statistics as a test statistic, comparing it with values in a chi-square table. Since the observations used in calculating the three chi-squares are not independent, the sum cannot be expected to have the chi-square distribution. Such pseudo chi-squares have smaller variance, however, than the analogous chi-square distributions (Cf. Atkinson, Bower, and Crothers, 1965). Therefore the actual probability of Type 1 error is less than for the chi-square distribution, and the test is conservative. Combining the chi-squares yields a pseudo chi-square of 28.05 with six degrees of freedom, less one degree of freedom for the parameter $\underline{t}$, which is significant beyond the .001 level.

In the following chapter, models are presented in which different assumptions are made about retrieval and memory of hypotheses. These alternative assumptions may lead to a better fit to the Levine data. The models also include a modified response assumption suggested by Chumbley's experimental data. Chumbley's finding that some subjects did not respond on training trials but performed perfectly on test trials, and that subjects had a nonchance tendency to press the right-hand button suggests that pre-solution responding is not necessarily related to hypothesis processing. In a situation in which emphasis is placed on post-solution performance, it is reasonable to conjecture that subjects concern themselves with solving rather than with maximizing the chance of a correct response on early trials. If working out a response rule based on the hypothesis set interferes with processing of the hypothesis set, then disregarding the correspondence between hypotheses and responses early in the task could be an effective strategy.

Some support may be found for this notion. Goodnow and Pettigrew (1956) found that subjects in a prediction task reported solving the problem rather easily when they stopped trying to predict and simply observed. In that study subjects had to make some response in order to get feedback, and so the "just observe" strategy was not as readily identified as in the Chumbley study. Byers (1965) allowed subjects in a concept attainment experiment the option of offering hypotheses on each trial, and found that the tendency to offer hypotheses on early trials decreased significantly over problems. In this case, the process of selecting a hypothesis from the tenable set may have interfered with processing. In the model to be developed in the next chapter, it will be assumed for tasks stressing post-solution performance that subjects respond according to strategies not connected with the tenable hypothesis set until only one element remains, and then respond according to the single hypothesis. For comparison to the Chumbley model, however, the new model will be fitted to the Levine data, and hypothesis-relevant responding will be assumed.

STATEMENT OF THE PROBLEM

Findings cited in the preceding chapter lead to a fairly detailed picture of the concept identification process. Recent evidence (Restle and Emmerich, 1966; Levine, 1966; Trabasso and Bower, 1966) indicates that the concept identification process does not restart after errors. Something is remembered. Trabasso and Bower proposed a model in which both the eliminated hypotheses and cue values of the positive stimulus enter memory. Restle and Emmerich argued that memory for stimulus information was necessary to explain their results.

Memory for rejected hypotheses was suggested by Erickson and Zajkowski (1967) and Levine's results indicate that hypotheses are remembered after errors. Although what is remembered in Levine's experimental situation is almost certainly a set of hypotheses, the situation is sufficiently different from the standard concept identification experiment to leave room for doubt that Levine's findings extend to that situation. (Cf. Trabasso and Bower, 1968, p. 50.) A test of some implications of Levine's theory in an ordinary concept identification task seems to be needed.

Chumbley (1969) developed a model based on Levine's theory and applied it to a situation in which subjects were given four training trials followed by seven test trials. The model fit the test trials, but Chumbley reported that it did not fit the training trials. The test described in the preceding chapter shows that prediction of the proportion of tenable hypotheses in the Levine study was also inadequately accurate.

A model consistent with the findings discussed in the preceding chapter is still needed. A major purpose of this study is to develop such a model, and to test it in an experimental situation that conforms to the usual concept identification arrangement.

Although something, probably a set of hypotheses, is remembered, it is equally clear that something is lost, or forgotten. What is not clear about the forgetting is when it occurs. It is reasonable to hypothesize that processing of a large or otherwise difficult set of hypotheses results in both loss from the hypothesis set and forgetting of previously stored information (retroactive interference). Restle and Emmerich's (1966) data on repeated presentation of a stimulus showed that there was some immediate loss of information, since performance was not perfect on the second presentation. The data on complementary stimuli (Kenoyer and Phillips, 1968) suggests that even more loss occurs over trials. One way of investigating this loss of information over trials is to present complementary pairs of stimuli, as in the Kenoyer and Phillips study, and manipulate the number of trials intervening between the presentation of the first and second member of a complementary pair. In the present study the lag effect was arranged factorially with the initial outcome sequences, in order to facilitate this kind of analysis. Versions of the model both with and without the retroactive interference assumption were developed and compared.

The use of fixed outcomes on initial trials in this study provides particularly powerful tests of the extant hypothesis models. When the "process model" (Cf. Gregg and Simon, 1967, p. 250) is examined rather than the stochastic model that is derived from it, several of the models discussed in the preceding chapter (Restle, 1962; Bower and Trabasso, 1963; Trabasso and Bower, 1966, 1968) yield deterministic predictions. These predictions require analysis of error trial stimuli so that consistency between the information provided on that trial and later performance can be determined. If the position of the error trial in the trial sequence can be predetermined, as in the fixed-outcome procedure, this consistency checking is facilitated considerably.

THE INDEPENDENT HYPOTHESIS ELIMINATION MODELS

General Development

The strategy of the present study is to isolate component assumptions of extant models and to test the assumptions individually when such tests can be devised. As Sternberg (1963) noted, a test of the whole model is a test of the logical conjunction of all of its assumptions. A test of a single assumption therefore serves as a test of the whole model, since falsity of any one element of a logical conjunction implies falsity of the conjunctive assertion. Whenever an assumption can be falsified in a reasonably simple experiment, therefore, it seems profitable to test it in isolation.

Besides serving to falsify models, tests of individual assumptions are useful in the construction of new models. Rejection of a given assumption may suggest an alternative treatment of a mechanism within a model. A framework of sorts has been established for the model to be developed in this chapter, simply by the nature of the models already discussed.

Several assumptions have been rejected in studies discussed in the preceding chapter. The sampling-with-replacement axiom has been falsified in a number of the studies cited (Levine, 1962, 1966; Holstein and Premack, 1965; Richter, 1965; Trabasso and Bower, 1966; Restle and Emmerich, 1966; Erickson and Zajkowski, 1967). An alternative assumption is sampling without replacement. Richter (1965) and Levine (1966) both found that subjects failed to display the perfect performance implied by this assumption. Restle and Emmerich (1966) and Kenoyer and

Phillips (1968) have presented evidence against the local consistency assumption. The assumption that subjects improve performance only after error trial has been refuted by Levine's (1966) results. In the same study Levine also demonstrated that subjects are capable of processing information about hypotheses that are not currently being used as a basis for responding.

An adequate model must not include any of the rejected assumptions. In the case of those assumptions that were refuted by Levine's data, it seems advisable to acquire further evidence in a standard experimental situation, but it is probably best to consider alternative assumptions when constructing a new model. Lack of fit of Chumbley's (1969) model to Levine's data suggests that alternatives to his process assumptions should be considered. Finally, Chumbley's presolution (training trial) results suggest a modification of the response assumption.

Common to all the models discussed here thus far is the concept of a set of hypotheses available to the subject, from which he selects elements to be tested against the feedback or information provided on each trial. Even in view of empirical evidence eliminating several assumptions included in various models, Levine's theory remains intact. The model to be proposed here is consistent with Levine's general hypothesis processing framework although it differs from Chumbley's more completely specified process assumptions. A reasonable alternative to Chumbley's assumption of all-or-none retrieval of the whole hypothesis set is all-or-none retrieval of each individual hypothesis. An example of this kind of model is Phillips, Shiffrin, and Atkinson's (1965) register model of short-term memory. In the hypothesis model to be developed here, however, the memory mechanism must be combined with other mechanisms, and so a register model of the memory process without simplifying assumptions leads to prohibitive complexity in the overall model. The assumption that hypotheses are retrieved independently, while probably not true, seems adequate for the purposes of the model being developed here.

A decision must be made as to what hypotheses are assumed to be remembered. If only the hypotheses that have not been eliminated are remembered, then loss of the correct hypothesis from this memory store would render the problem unsolvable. This difficulty can be handled by assuming perfect memory, but this assumption does not fit available data (e.g., Levine, 1966; Richter, 1965). Another solution is to assume, as Chumbley (1969) did, that the subject starts with the entire hypothesis set if memory fails. Given that the hypothesis set can be reconstructed from the stimuli, this is quite reasonable. It could even be assumed if it required the subject to store the initial hypothesis set in memory. The Chumbley model, however, has been shown to yield unsatisfactory fit to Levine's data, and so an alternative explanation should be considered.

An alternative assumption is that what is remembered is the set of logically eliminated hypotheses. The complement of this set yields the currently entertained set, and so the information needed for responding is always available. Equivalently, the subject could scan the stimulus, matching its elements with eliminated hypotheses, and so avoid dealing with the entertained set. Under this assumption any forgotten hypotheses simply become part of the set of hypotheses that are currently entertained by the subject and have to be eliminated again. In this view, a set of hypotheses is not forgotten. Rather, the subject only forgets which hypotheses have been eliminated.

A flow diagram of the Independent Hypothesis Elimination (IHE) models appears in Figure 1. As the figure indicates, the subject is assumed to begin the task by establishing two sets, or lists. The set U_o is the set of hypotheses held by the subject to be untenable at the beginning of the task. U_o may be described as containing all hypotheses that are disallowed by the experimental instructions, but the model deals only with those hypotheses that are described to the subject as legitimate. In the context of this set (H) of hypotheses, U_o is assumed to be empty. Since information provided to the subject makes logical elimination of hypotheses possible, U_n is not generally empty for $n > 0$. The residual hypothesis set, R_n, is the complement of U_n with respect to H.

Hypotheses sufficient for solution of the kind of problem of interest here must specify a partition of the stimulus set in which the subsets are assigned to categories established by the experimenter. (Cf. Haygood and Bourne, 1965.) A hypothesis could, for example, associate red figures with VEK and green figures with NONVEK. An equivalent partition would be obtained by associating red figures with VEK and "everything else" to NONVEK. In a two-category problem, specification of the second category, is redundant. The nonredundant representation is assumed in the present model. Adoption of this assumption requires assumption of an additional step in the process, in which one of the two

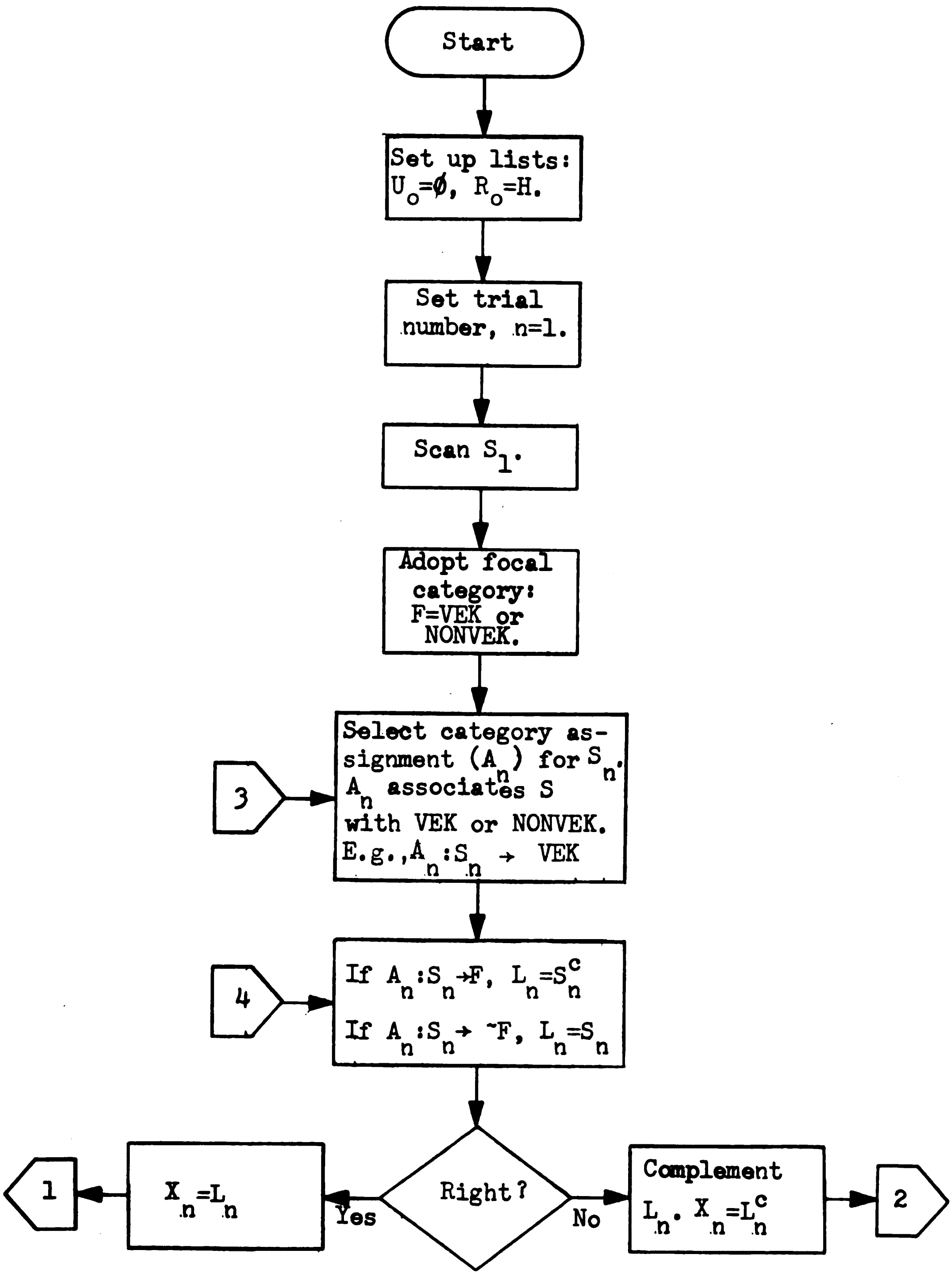


Figure 1. Flow diagram of IHE models.

(Continued on page 42)

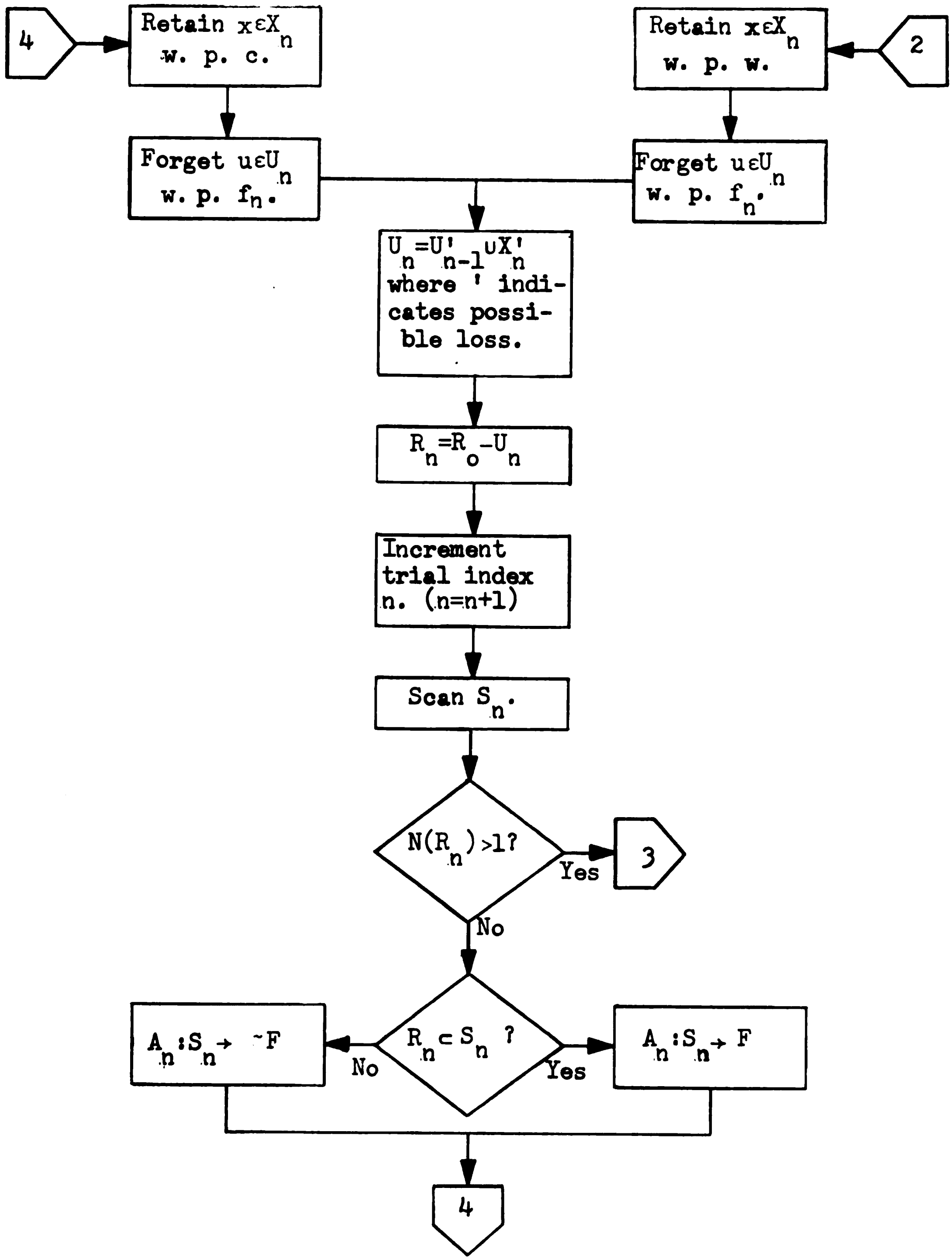


Figure 1. Flow diagram of IHE models.

(Continued from page 41)

categories is adopted as a focal category, i.e., the subject chooses to consider VEKs or NONVEKs to be positive instances. Stimuli are assumed to be assigned to this focal category, then, throughout the task.

The next step in the assumed process is to select a category assignment, A_n, which associates the stimulus on trial n with VEK or NONVEK. If A_n assigns S to the focal category (F), then the subject sets up a list (L_n) of the cue values in the complement of the stimulus (i.e., those not present in the stimulus), which can be eliminated if he is correct. If the subject does not assign the stimulus to the focal category, then the cue values present in the stimulus are placed in the list L_n.

If the response is called "right", the subject has only to retain L_n and add its members to U_n, the untenable set. Because little processing is assumed to occur at this stage, the probability of loss is relatively small. If the subject is told "wrong", however, he must recover the complement of L_n with respect to H. The recovery process is assumed to increase the likelihood of an error. Elements are then forgotten from U_n with probability f_n. After U_n has been obtained in this way, the residual hypothesis set R_n, can be recovered by eliminating the elements of U_n from the full hypothesis set H.

On later trials, after enough information has been presented to the subject for logical solution of the problem, the nature of the assumed process depends upon whether solution has occurred. If only one hypothesis remains, the problem is solved, and the response is determined by whether the cue value that the hypothesis associates with the focal category is present in the stimulus. If so, the stimulus is assigned to the focal category; otherwise it is assigned to the other category. If more than one hypothesis remains in the residual set, the subject selects a category assignment by a strategy that is not specified in the flow chart. In an experiment for which pre-solution performance has been stressed, such as Levine's (1966) experiment, the assumed strategy is to respond according to a randomly selected hypothesis from the residual set. In an experiment such as Trabasso and Bower's (1966, 1968), using redundant relevant cues, it is assumed that subjects notice the redundancy of the cues corresponding to hypotheses in the residual set rather quickly when all other hypotheses have been eliminated, and respond consistently with those hypotheses. In the ordinary concept identification task, however, the response selection process is assumed to be unrelated to hypotheses in the residual set (including the correct one), and so responses are randomly correct or incorrect. This property of the model would account for Chumbley's (1969) finding that training-trial data were not predictable by his HM model, since choice responses on the training trials would not be related to the subjects' hypotheses.

Hypothesis States

It is more convenient to represent IHE Models in terms of hypothesis states than in terms of subject states. If we consider the probability, v_{in}, that hypothesis H_i is in R_n, the state probability vector for hypotheses on trial n is:

$V_n = \langle v_{1n}, v_{2n}, \ldots, v_{in}, \ldots, v_{mn} \rangle$, where there are _m_ hypotheses in R_o. Since all hypotheses are, by assumption, in R_o with probability 1,

$$V_o = \langle 1,1,1,1,1,1,1,1\rangle .$$

Now if a transformation, T_{in}, can be specified such that

$v_{in} = v_{i,n-1}\, T_{in}$, then v_{in} can be obtained by successive applications of transformations to hypothesis states, and so V_n can be obtained for $n = 1,2, \ldots.$ Given V_n, the probability distribution may be obtained for the number of hypotheses in R_n. The probability that there is exactly one hypothesis in R_n is

$$\sum_{i=1}^{m} \Pr\,[H_i \in R_n \,\&H_j \notin R_n, j\neq 1] = \sum_{i=1}^{m} \Pr\,[H_i \in R_n] \prod_{j\neq i} \Pr\,[H_j \notin R_n]$$

$$= \sum_{i=1}^{m} v_i \prod_{j\neq 1} (1-v_j)$$

since hypotheses are eliminated independently. In general, if we define an m-element vector X_n such that

$$x_{i,n} = \begin{cases} 1 \text{ if } H_i \in R_n \\ 0 \text{ if } H_i \notin R_n \end{cases}, \quad i=1,\ldots,m$$

then $\Pr\,[N(R_n) = k] = \sum_{X_n \in K} \prod_{i=1}^{m} v^x(1-v)^{1-x},$

where subscripts for v and x have been excluded for clarity. Thus v should be read as v_{in} and x should be read as x_{in}. K is the set of all vectors X such that

$$\sum_{i=1}^{m} x_{in} = k.$$

When $v=x=0$, v^x is taken to be 1, and for $v=x=1$, $(1-v)^{1-x}$ is taken to be 1. Thus the probability distribution can be derived from the state probabilities for individual hypotheses.

Response Assumptions

For situations such as the Levine (1966) experiment, in which the subject is encouraged to optimize pre-solution responding, it is assumed that the hypothesis upon which his responses are based is selected from R_n, and that all elements of R_n are equally likely to be selected. In this process there is no way to eliminate a hypothesis unless information on the current trial allows its logical elimination, and so every hypothesis in U_n (i.e., every hypothesis not in R_n) is in the set (D_n) of hypotheses that have been logically eliminable on or before the *n*th trial. It follows that the complement of D_n is a subset of R_n, and therefore that

$$N(D_n^c) < N(R_n)$$

The probability that the working hypothesis (H*), which is selected from R_n, is also a member of D_n^c, is

$$\Pr [H^* \varepsilon D_n^c] = \sum_{k=1}^{8} \Pr [H^* \varepsilon D_n^c \mid r_n = k] \cdot \Pr [r_n = k]$$

$$= \sum_{i=1}^{8} \frac{N(D_c^n)}{k} \cdot \Pr [r_n = k]$$

Eliminability Indicators

In the development that follows, it is convenient to define an eliminability indicator, e_{in}, for the *i*th hypothesis on trial *n*. The indicator e_{in} takes on the value 1 if H_i is eliminable on trial *n*, 0 otherwise. For the case of eight hypotheses, there is an ordered set of eight such indicators for each trial, which may be represented as an eight-element vector, E_n. Symmetry in the hypothesis set makes it possible to

place the elements in any arbitrary order, given that the same ordering is maintained over all trials of a problem. On each trial, half of the hypotheses are eliminable and half are not. Thus we can represent the vector for trial 1 as:

$$E_1 = \langle 1,1,1,1,0,0,0,0 \rangle$$

Half of the elements have the same value on trial 2 as on trial 1, and the other half have the opposite value, assuming "orthogonality" of stimuli (Cf. Levine, 1963). We can therefore represent the vector for trial 2 as:

$$E_2 = \langle 1,1,0,0,1,1,0,0 \rangle$$

A vector for trial 3 that satisfies the orthogonality requirement for the two vectors above is:

$$E_3 = \langle 1,0,1,0,1,0,1,0 \rangle$$

The principles outlined above apply to all of the Independent Hypothesis Elimination (IHE) models. The individual IHE models differ with respect to the nature of the transformation, T_{in}, that operates on $v_{i,n-1}$ to yield v_{in}.

IHE Model 1

In IHE Model 1, it is assumed that hypotheses are not lost (i.e., forgotten) from U_n. The probability that an eliminable hypothesis in R_n is also in R_{n+1} is the probability that the elimination process fails for that hypothesis, i.e., 1-*w* if trial *n* is an error trial or 1-*c* if it is a correct trial. Thus

$$v_{in} = v_{i,n-1} \cdot \Pr [\text{ not eliminated } | \text{ eliminable }]$$

$$\cdot \Pr [\text{ not eliminable }] + v_{i,n} \cdot \Pr [\text{ not eliminable }]$$

$$= v_{i,n-1} (1-p_n e_{in}), \text{ where } p_n = w \text{ or } p_n = c.$$

Given a pair of values for _w_ and _c_, the transformation rule is specified for IHE Model 1 for the first three trials, and hypothesis state probabilities can be generated in a Fortran program. Thus the same procedure described above for the Chumbley model can be used to evaluate IHE Model 1 against Levine's data.

Test values for the parameters of the model (_w_ and _c_) were used to generate expected proportions of tenable hypotheses for four situations: following a W outcome on trial 1, and following trial 3 for the RRW, WWW, and RWW-WRW conditions. A Pearson chi-square statistic was computed for each of three pairs of proportions consisting of the trial 1 proportion and one of the three trial 3 proportions. The observed proportions used in the computations were those from Levine's study. The parameter values selected were those for which the sum of the three chi-square statistics was a minimum.

The procedure differed somewhat, because two parameters were being varied. First a relatively coarse grid was used, in which _w_ and _c_ varied in steps of .10. In regions where fit was best, a finer grid was applied, until steps of .01 were used in the best-fit regions. While it must be recognized that extrema of functions (in this case, the chi-square value) may be missed by such procedures, inspection of the values generated did not suggest failure of monotonicity as parameters were varied from an optimum value.

IHE Model 2

A hypothesis in R_{n-1} is assumed to enter U_n with a probability determined by e_{in} and p_n, just as in IHE Model 1. If we represent the probability that the hypothesis is in R_n immediately after the elimination process as v_{in}^* we have:

$$v_{in}^* = v_{i,n-1} (1-p_n e_{in}).$$

Hypotheses in U_n, however, are assumed to be forgotten (and hence enter R_n) with probability *f*. Applying the forgetting operator to v_{in}^*,

$$\begin{aligned} v_{in} &= v_{in}^* + f(1-v_{in}^*) \\ &= f+v_{in}^* (1-f) \\ &= f+v_{i,n-1} (1-p_n e_{in}) (1-f) \end{aligned}$$

The probability *f* is a free parameter of the model, and has the same value on every trial. The transformation characterizing IHE Model 2 is therefore completely specified.

The same method of parameter estimation and test of fit described for IHE Model 1 was also applied to IHE Model 2 in order to fit the model to Levine's data.

IHE Model 3

This model differs conceptually from IHE Model 2 only in that the probability of retaining a hypothesis in U_n is not constant. If the memory store for rejected hypotheses is separate from the memory store for hypotheses currently being logically manipulated, the assumption of a constant forgetting parameter seems reasonable. In IHE Model 3, however, these memory stores are assumed to be affected similarly by processing requirements. More precisely, both retention of hypotheses in U_n and elimination of those currently being processed are assumed to occur with the same probability p_n on a given trial. Then,

$$v_{in}^* = v_{i,n-1} (1-p_n e_{in})$$

Now, since hypotheses in U_n are assumed to be forgotten with probability $1-p_n$,

$$v_{in} = v_{in}^* + (1-p_n) (1-v_{in}^*)$$

$$= (1-p_n) + v_{in}^* p_n$$

$$= (1-p_n) + v_{i,n-1} (1-p_n e_{in}) p_n$$

The transformation rule characterizing IHE Model 3 is completed. The procedure for evaluating IHE Model 3 was the same as that for IHE Model 1.

IHE Model 4

If the order of the two processes, hypothesis elimination and forgetting of eliminated hypotheses, is reversed, a new model results. This modified model describes a process in which the forgetting operation occurs when the subject is analyzing the stimulus and processing information that leads to hypothesis elimination, rather than after information processing has occurred. Such a model could, for example, describe a process in which storage of new information tends to result in displacing old information. IHE Model 4 is described here in terms of the operators already described for IHE Model 3, applied in reverse order. Thus we have

$$v_{in}^* = v_{i,n-1} + (1-p_n) (1-v_{i,n-1}) = 1-p_n (1-v_{i,n-1})$$

$$v_{in} = v_{in}^* (1-p_n e_{in})$$

$$= [1-p_n (1-v_{i,n-1})] (1-p_n e_{in})$$

The parameters w and c were estimated, and IHE Model 4 was tested, by the same procedure used for IHE Model 3.

IHE Model 5

Just as IHE Model 4 was obtained from IHE Model 3 by reversing the order in which the forgetting and elimination operators are applied, the corresponding operators for IHE Model 2 may be reversed to yield IHE Model 5. Thus we obtain

$$v_{in}^* = v_{i,n-1} + (1-v_{i,n-1})\, f = 1+f-fv_{i,n-1}$$

$$v_{in} = v_{in}^* \,(1-p_n\, e_{in}) = (1+f-fv_{i,n-1})\,(1-p_n\, e_{in})$$

The model is identical in all other respects to IHE Model 2, and the same procedure was used for parameter estimation and fit that were used for that model.

Comparison of Models

Levine (1966) reported the proportions that were used in the present study for preliminary evaluation of the models described above. The first of these is the proportion of hypotheses following an error on trial one that are consistent with the information provided by that trial. For each of three conditions, one error (RRW), two errors (WRW or RWW), and three errors (WWW), Levine reported the proportion of hypotheses following the third trial (for conditions in which the outcome was "wrong") that were consistent with the information provided by the first three trials. For each of the models discussed above, parameters were varied to generate probabilities corresponding to these proportions and the parameter values that yielded the best fit to the

observed proportions (by the criterion described previously) were taken as estimates of the parameters. These estimates appear at the right-hand side of Table 1.

At the top of Table 1 are the observed proportions reported by Levine. The corresponding predicted values are given for each model. The chi-square value for each condition appears in Table 1 just below the expected third-trial proportion on which it was computed. The sum of the three chi-squares appears on the same line with them, to the right.

Each chi-square has two degrees of freedom if none are deducted for parameter estimation. Thus 6.0 is the critical value at the .05 level of significance. Thus each of the chi-square values for Chumbley's HM model leads to rejection of the model. IHE Model 1 and IHE Model 3 fit even more poorly using either the sum of the chi-squares or each chi-square as a criterion. One of the chi-square values of IHE Model 4 is significant beyond the .01 level and the sum is significant beyond the .005 level. The fit to Levine's data is better for IHE Model 2, but is not really good, since two of the three chi-squares are significant beyond the .10 level and the sum is significant beyond the .05 level with 5 degrees of freedom. For IHE Model 5 the fit is much better. None of the chi-squares is significant at the .25 level. The pseudo chi-square for the sum has three degrees of freedom after correction for estimating $\underline{w}$, $\underline{c}$, and $\underline{f}$. It is significant beyond the .25 level, but not at the .10 level. If only two degrees of freedom are deducted for the nondegenerate parameters $\underline{w}$ and $\underline{f}$, four degrees of freedom remain. By this reckoning, the sum is not significant even at the .25 level.

Table 1

Summary of Fit of Models to Levine's Data

Proportion of Hypotheses that are Logically Tenable

		After Trial One	After Trial Three-Conditions One Error	Two Errors	Three Errors	Sum of Chi-squares	Parameters
	Observed Proportions	.873	.44	.32	.25		
Chumbley's HM Model	Predicted Proportions	.745	.55	.40	.31		
	Chi-square		11.09	8.91	8.05	28.04	t=.49
IHE Model 1	Predicted Proportions	.764	.36	.43	.50		w=.65
	Chi-square		10.41	12.70	34.98	58.09	c=.45
IHE Model 2	Predicted Proportions	.823	.34	.34	.34		w=1.0
	Chi-square		4.97	1.95	5.32	12.23	c=1.0 f=.25
IHE Model 3	Predicted Proportions	.787	.37	.41	.46		w=.83
	Chi-square		6.82	8.30	24.11	39.23	c=.73
IHE Model 4	Predicted Proportions	.811	.39	.39	.38		w=.73
	Chi-square		3.54	4.40	10.17	18.12	c=.75
IHE Model 5	Predicted Proportions	.868	.35	.34	.32		w=.82
	Chi-square		2.47	.14	2.37	4.99	c=1.0 f=.50

Certain features of the models became apparent as they were fitted to the Levine data. All of the models predicted proportions for trial one that are too small. This is quite pronounced for Chumbley's HM Model, IHE Model 1, and IHE Model 3. For the last two models mentioned, the optimal value of the parameter w for fitting the data point for trial one (after one error), was obviously higher than the optimal value for fitting the data points for trial three. The parameter c, however, was unaffected by the trial one proportion since trial one was an error trial. One result was that the estimate of w was higher than that of c, and the expected proportions of consistent hypotheses after one, two, and three errors were in increasing order rather than in the decreasing order of observed proportions. This kind of prediction by the model is qualitatively unacceptable. The evidence presented both by Levine (1966) and by Richter (1965) indicates that information provided by "right" trials is more effectively used than information provided by "wrong" trials.

In IHE Model 2, which has an additional parameter for forgetting, both w and c lose their potency as parameters. The best estimate for each is 1.00. In effect, this means that the effect of adding the forgetting parameter is to override the other two parameters.

The qualitative defect found in IHE Model 1 and 3 did not appear in IHE Model 4. The parameters w and c, and therefore also the expected proportions for the three outcome conditions, are in the proper ordinal relationship. Quantitative fit is not impressive, however. There is virtually no difference among the expected proportions on trial three, and the expected proportion for trial one is low.

The fit of IHE Model 5 is the best of all the models tested. The expected proportion for the first trial fits well, and those for the third trial in the three experimental conditions are properly ordered. It is apparent, however, that the model does not differentiate strongly enough among the three conditions. The expected proportions are more similar than the observed proportions.

Application of the models to Levine's data served as a screening process by which some of the models could be excluded from further testing against data collected in the present study. The least promising of the models discussed above are Chumbley's HM model, IHE Model 1, and IHE Model 3. Besides generally poor fit, the two IHE models displayed serious qualitative defects. The HM model did not seem to warrant further testing, and there is reason to doubt that its author intended that the model be applied to experiments such as those of the present study, in which solution is stressed rather than pre-solution performance.

Of the remaining three models, IHE Models 2 and 5 yield the best fit to Levine's data. These models are quite similar, differing only in the order in which the forgetting process (operator) and the hypothesis elimination process (operator) are applied to the hypothesis state probabilities. The remaining model, IHE Model 4, fits Levine's data less adequately than the two just discussed, but was retained for further testing. The procedure of Levine's study, in which blank trials were administered, may have led to a greater degree of forgetting of feedback information than occurs when feedback is given on every trial. Such a state of affairs is suggested by the finding that the forgetting parameter

(f) in IHE Models 2 and 5 overrides the parameters w and c. Therefore IHE Model 4, the best of the models that did not include the parameter f, was tested against the data of the present study.

METHOD

Design

Two separate experiments were conducted. In Experiment 1, pairs of complementary stimuli were separated by one trial (lag 1) or five trials (lag 5). The other independent variable, the sequence of outcomes on the first three trials, was combined with the lag variable in an incomplete factorial design. There were two types of problems. In the predominant type, all outcomes were predetermined. For these problems, all possible outcome sequences were used on the first three trials, and responses were called correct on the remaining six trials. These are called fixed-outcome problems. In the other type of problem, the first three outcomes were fixed, but the responses on later trials were considered right or wrong depending upon whether they were consistent with the stimulus-response outcome information on the first three trials. These are called contingent-outcome problems. The first two problems were of this type, as well as the first two problems in the last half of the eighteen-problem set (problems 10 and 11). Each subject performed on all problems, but two groups were given the problems in two different orders. The outcome sequences are shown in Table 2 for all problems for group 1, where an underscore with no letter indicates that the outcome for that trial was contingent upon agreement with the hypothesis determined by trials 1 through 3. Asterisks mark the trials on which the complementary stimuli, C_1 and C_2, were presented. For group 2 the problems were arranged in a different order. Problems

Table 2

Outcome Sequences for Experiment 1, Group 1

Task	Outcome								
1.	W	R	R	_	_	_	_	_	_
2.	W	W	R	_	_	_	_	_	_
3.	R	R	W*	R	R*	R	R	R	R
4.	W	W	W*	R	R	R	R	R	R*
5.	R	R	W*	R	R	R	R	R	R*
6.	R	W*	R	R	R	R	R	R*	R
7.	W	R	W*	R	R*	R	R	R	R
8.	R	W*	R	R*	R	R	R	R	R
9.	W*	R	R	R	R	R	R*	R	R
10.	W	R	W	_	_	_	_	_	_
11.	W	W	R	_	_	_	_	_	_
12.	R	W	W*	R	R*	R	R	R	R
13.	W	R	W*	R	R	R	R	R	R*
14.	R	W	W*	R	R	R	R	R	R*
15.	W	W*	R	R	R	R	R	R*	R
16.	W	W	W*	R	R*	R	R	R	R
17.	W	W*	R	R*	R	R	R	R	R
18.	R*	R	R	R	R	R	R*	R	R

1, 2, 10, and 11 were presented in the same order and the problem blocks 3-9 and 12-18 were interchanged.

The same initial-outcome variable was used in Experiment 2. Three warmup problems were administered. Sixteen contingent-outcome problems were administered in which the lag was always 5 and each of eight initial-outcome conditions was administered twice. Apparatus considerations made it convenient to administer 18 rather than 16 problems besides the warmup problems. Therefore the first problem in the first half and the first problem in the last half of the experimental problem set were extra fixed-outcome problems, inserted so that the number of problems conformed to the apparatus constraint. The treatment orders were varied by interchanging the problem blocks in the first and second halves of the problem set as in Experiment 1. The warmup and extra problems were administered in the same order for both groups. The outcome structures for group 1 appear in Table 3. In Experiment 2 subjects were asked to state the correct hypothesis at the end of each problem if they knew it, but were not asked to guess.

Subjects

College students fulfilling an introductory psychology course requirement served as subjects. In Experiment 1 all Group 1 subjects were run before Group 2 subjects rather than in random order because it was necessary to reorder all stimuli before changing groups. Since all subjects had the same stimulus sequence in Experiment 2, subjects were randomly assigned with the constraint that the sizes of the two groups

Table 3

Outcome Sequences for Experiment 2, Group 1

Task			Outcome						
1.	R	R	W	_	_	_	_	_	_
2.	W	R	R	_	_	_	_	_	_
3.	W	W	R	_	_	_	_	_	_
4.	R	W	R	R	R	R	R	R	R
5.	W	W	W	_	_	_	_	_	_
6.	R	R	W	_	_	_	_	_	_
7.	R	W	W	_	_	_	_	_	_
8.	W	R	W	_	_	_	_	_	_
9.	W	W	R	_	_	_	_	_	_
10.	R	W	R	_	_	_	_	_	_
11.	R	R	R	_	_	_	_	_	_
12.	W	R	R	_	_	_	_	_	_
13.	R	R	W	R	R	R	R	R	R
14.	W	R	W	_	_	_	_	_	_
15.	R	R	R	_	_	_	_	_	_
16.	R	W	W	_	_	_	_	_	_
17.	W	W	W	_	_	_	_	_	_
18.	R	W	R	_	_	_	_	_	_
19.	W	W	R	_	_	_	_	_	_
20.	W	R	R	_	_	_	_	_	_
21.	R	R	W	_	_	_	_	_	_

were kept as nearly equal as possible throughout. In each experiment a few subjects were discarded because of apparatus failures and procedural errors. One subject was discarded because of avowed red-green color blindness, although there was no evidence that he had any difficulty with color discrimination in the experiment. Data from 61 subjects were analyzed in each experiment. There were 31 subjects in Group 1 and 30 in Group 2 for each.

Apparatus

For both experiments stimuli were presented on a rear-projection screen of flashed opal glass. The screen was installed in an open-backed cabinet of wood and hardboard, and was visible through a clear plastic window 4 inches high by 12 inches wide, in the front of the cabinet. The window was in three sections, and each section was hinged at the top to form a transparent movable panel. The bottom of each section rested against a Microswitch, which served to register responses. Figure 2 shows this cabinet. The labels "VEK" and "NON-VEK" were placed on the leftmost and rightmost panels, respectively. The center panel, on which the stimulus appeared, was locked so as to be immovable. The categorizing response on each trial was indicated by pressing the panel with the appropriate label. This device was described previously in reports of similar research (Kenoyer, 1968; Kenoyer and Phillips, 1968).

The subject inputs (switch closures) could be rendered ineffective by the experimenter by means of a pushbutton control held in his hand; another button on the same control device rendered the subject's inputs effective. When these inputs were ineffective, a red light just above

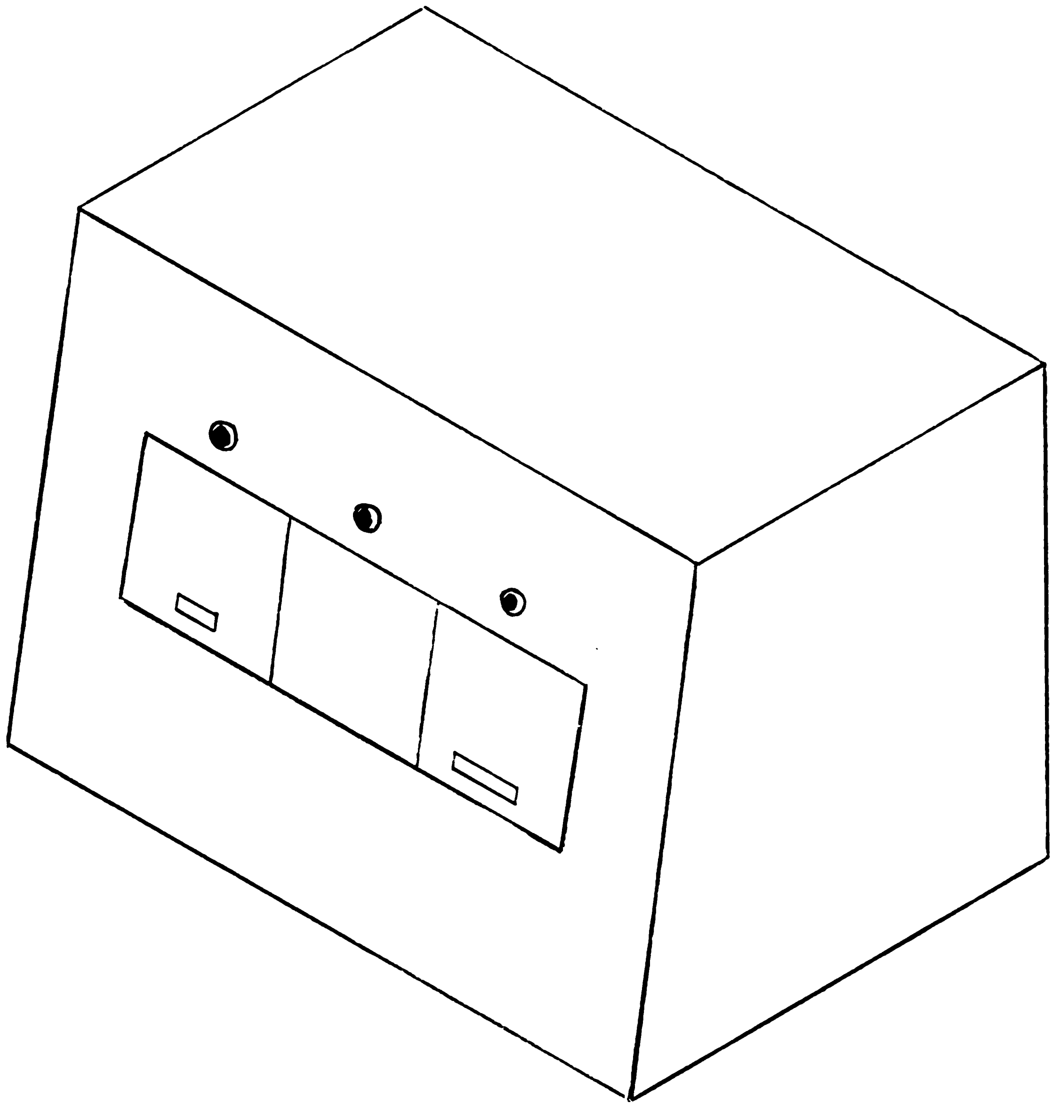


Figure 2. Stimulus display and response device.

the stimulus window was turned on. When the inputs were effective a response by the subject advanced the Carousel projector by which the stimuli were displayed, showing the stimulus for the next trial.

For Experiment 2 all stimuli could not be loaded simultaneously into a single Carousel tray, and so two Carousels were used. The stimuli for the three warmup problems were loaded in one Carousel projector and the remaining stimuli were loaded in a second Carousel, in order to avoid interrupting the procedure to change trays.

Procedure

The instructions shown in Appendix A were read to the subjects. A demonstration of the subject response panels was given, with the inputs disabled, and the functions of the red signal light and response panels were explained. When subjects had questions, the instructions were paraphrased. As the instructions state, the subject was supplied with a card (Figure 3) listing the stimulus dimensions and the values on each dimension. Another card, pictured in Figure 4, was shown to subjects when the nature of the concepts was being described.

In the first experiment the subject progressed through the 18 tasks with only momentary breaks between consecutive tasks. During this interval the red light indicating the end of a problem was on. Subjects typically began the new problem immediately when the light was extinguished; if not, the experimenter informed the subject that it was time to start a new problem. There was considerable variation in time to complete the set of tasks, but nearly all subjects required more than 15 minutes and less than 30 minutes.

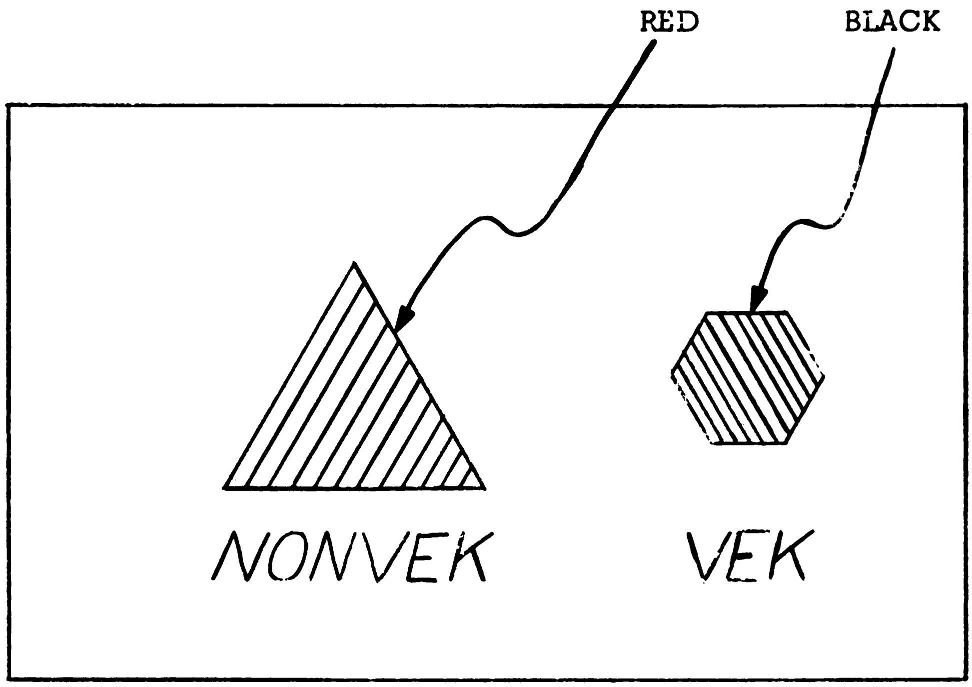


Figure 3. Card shown to subjects to illustrate the nature of concepts.

ATTRIBUTES		
	Size:	large small
	Shape:	circle square
	Color:	red green
	Border:	bordered unbordered

Figure 4. List of attributes and values.

In the second experiment there were 21 tasks in all, and total performance time was slightly longer. The procedure also differed in that there was a pause after the familiarization trials to change projector connections, and subjects were asked to state the correct hypothesis at the end of each task.

The experimenter sat to one side of the subject in a chair with a writing surface. A record booklet was placed on the writing surface. The subject sat in a chair facing the presentation device, which sat on a table. The booklets for experiments 1 and 2 are shown in Appendices B and C, respectively. The experimenter provided feedback for the first three responses in accordance with the predetermined sequence of outcomes for the first three trials in every case. For fixed-outcome problems, all remaining responses were called "right." For contingent-outcome problems, the experimenter tracked the subject's response sequence on the decision tree shown in the protocol booklet. This procedure determined the correct hypothesis for a problem after three trials, and subsequent outcomes were made contingent upon agreement with the hypothesis determined in this way.

Stimulus Materials

The stimuli were figures projected on the rear-projection screen, varying on four binary attributes: Size, shape, color, and border. Figures were either red or blue, squares or circles, and either had a white border or no border. The large figures were four times the area of the smaller, and squares were approximately equal to circles in area. All figures appeared on a dark background. For experiment 1 two different randomized stimulus orders were used for the two groups. The

two orders appear in Tables 4a and 4b. The stimuli for the first half of the problem set are the same as those for the last half. This repetition was caused by progressing through the entire set of slides in the Carousel slide tray twice. Only one order was used in Experiment 2. The stimuli for problem blocks 4-12 and 13-21 were identical for the same reason just given for Experiment 1. The stimulus order appears in Table 5.

Table 4a

Stimulus Sequences for Experiment 1, Group 1

Task	Stimulus								
1.	LGCN	LGQB	SGCB	SGQB	LRQB	SRCN	LRCN	SRCB	LGCB
2.	LGQB	LGCN	LRQN	SGCN	LRCN	SRQB	SRQN	SGQN	LGQN
3.	SGQB	SRCB	SRQN	LRQN	LGCB	SRQB	LGQB	SGCN	LRCB
4.	SRCN	SRQB	LRCB	LRQN	SRQN	SGQB	LGQN	SGCN	SGQN
5.	LGCB	LRQB	SRCB	SGQN	LGQB	SRQB	LRCN	SGCB	LGQN
6.	LRCB	SGCB	SRQB	SRCN	SGQB	LRQB	LGCB	LRQN	LGQN
7.	SRCB	SGQB	LRQB	LGCN	SGCN	LRCB	SGCB	SRQN	LGQB
8.	LRCN	SGCN	LGQN	LRQB	SRQB	LRQN	SGQN	LGCB	SRCN
9.	LGCN	SGCB	SGQN	LRQN	SRQN	LRCN	SRQB	SRCN	SGCN
10.	LGCN	LGQB	SGCB	SGQB	LRQB	SRCN	LRCN	SRCB	LGCB
11.	LGQB	LGCN	LRQN	SGCN	LRCN	SRQB	SRQN	SGQN	LGQN
12.	SGQB	SRCB	SRQN	LRQN	LGCB	SRQB	LGQB	SGCN	LRCB
13.	SRCN	SRQB	LRCB	LRQN	SRQN	SGQB	LGQN	SGCN	SGQN
14.	LGCB	LRQB	SRCB	SGQN	LGQB	SRQB	LRQN	SGCB	LGQN
15.	LRCB	SGCB	SRQB	SRCN	SGQB	LRQB	LGCB	LRQN	LGQN
16.	SRCB	SGQB	LRQB	LGCN	SGCN	LRCB	SGCB	SRQN	LGQB
17.	LRCN	SGCN	LRQN	LRQB	SRQB	LRQN	SGQN	LGCB	SRCN
18.	LGCN	SGCB	SGQN	LGQN	SRQN	LRCN	SRQB	SRCN	SGCN

Code: L; large, R: red, C: circle, B: border, S: small,

G: green, Q: square, N: border.

Table 4b

Stimulus Sequences for Experiment 1, Group 2

Task	Stimulus								
1.	LGCN	LGQB	SGCB	SGQB	LRQB	SRCN	LRCN	SRCB	LGCB
2.	LGQB	LGCN	LRQN	SGCN	LRCN	SRQB	SRQN	SGQN	LGQN
3.	LRCB	SGCB	SRQB	SRCN	SGQB	LRQB	LGCB	LRQN	LGQN
4.	LRCN	SGCN	LGQN	LRQB	SRQB	LRQN	SGQN	LGCB	SRCN
5.	SRCB	SGQB	LRQB	LGCN	SGCN	LBCB	SGCB	SRQN	LGQB
6.	SGQB	SRCB	SRQN	LRQN	LGCB	SRQB	LGQB	SGCN	LRCB
7.	LGCN	SGCB	SGQN	LGQN	SRQN	LRCN	SRQB	SRCN	SGCN
8.	LGCB	LRQB	SRCB	SGQN	LGQB	SRQB	LRCN	SGCB	LGQN
9.	SRCN	SRQB	LRCB	LRQN	SRQN	SGQB	LGQN	SGCN	SGQN
10.	LGCN	LGQB	SGCB	SGQB	LRQB	SRCN	LRCN	SRCB	LGCB
11.	LGQB	LGCN	LRQN	SGCN	LRCN	SRQB	SRQN	SGQN	LGQN
12.	SGQB	SRCB	SRQN	LRQN	LGCB	SRQB	LGQB	SGCN	LRCB
13.	SRCN	SRQB	LRCB	LRQN	SRQN	SGQB	LGQN	SGCN	SGQN
14.	LGCB	LRQB	SRCB	SGQN	LGQB	SRQB	LRCN	SGCB	LGQN
15.	LRCB	SGCB	SRQB	SRCN	SGQB	LRQB	LGCB	LRQN	LGQN
16.	SRCB	SGQB	LRQB	LGCN	SGCN	LRCB	SGCB	SGQN	LRQB
17.	LRCN	SGCN	LGQN	LRQB	SRQB	LRQN	SGQN	LGCB	SRCN
18.	LGCN	SGCB	SGQN	LGQN	SRQN	LRCN	SRQB	SRCN	SGCN

Code: L: large, R: red, C: circle, B: border, S: small,
G: green, Q: square, N: no border.

Table 5

Stimulus Sequences for Experiment 2

Task	Stimulus								
1.	LRCN	SRCB	SGCN	SGCB	SGQN	LGCN	LGCB	SRCN	SGCN
2.	SGQB	LGQN	SRQN	SGCN	SRCN	LRCB	SRQB	LRQN	LRCN
3.	LRCN	SRCB	SRQN	LGCN	LGQB	LRCB	SGCN	LRQN	LGCB
4.	LGQB	LRQN	LGCN	SGCN	LRCN	SGQN	LGQN	SRQN	SRQB
5.	SGCB	SGQN	SRQB	SRCN	LGQN	SRQN	LRCN	SGCN	SRQB
6.	LRCB	SRQB	SGCB	SRCN	SGQB	LRQB	LGCB	LGQN	LRQN
7.	LRCN	LGQN	SGCN	SRQB	LRQN	SGQN	SRCN	LGCB	LRQB
8.	SRCB	SGQB	LRQB	LGCN	LRCB	SGCB	LGQB	SRQN	SGCN
9.	SGQB	SRCB	SRQN	LRQN	SRQB	LGQB	SGCN	LRCB	LGCB
10.	LGCN	SGCB	LGQB	SRCB	SGQB	LGCB	LRCN	LRQB	SRCN
11.	SRCN	SRQB	LRCB	LRQN	SRQN	SGQB	LRQN	SGCN	SGQN
12.	LGCB	LRQB	SRCB	SGQN	LGQB	SRQB	LRCN	SGCB	LGQN
13.	LGQB	LRQN	LGCN	SGCN	LRCN	SGQN	LGQN	SRQN	SRQB
14.	SGCB	SGQN	SRQB	SRCN	LGQN	SRQN	LRCN	SGCN	SRQB
15.	LRCB	SRQB	SGCB	SRCN	SGQB	LRQB	LGCB	LGQN	LRQN
16.	LRCN	LGQN	SGCB	SRCN	SGQB	LRQB	LGCB	LGQN	LRQN
17.	SRCB	SGQB	LRQB	LGCN	LRCB	SRCB	LGQB	SRQN	SGCN
18.	SGQB	SRCB	SRQN	LRQN	SRQB	LGQB	SGCN	LRCB	LGCB
19.	LGCN	SGCB	LGQB	SRCB	SGQB	LGCB	LRCN	LRQB	SRCN
20.	SRCN	SRQB	LRCB	LRQN	SRQN	SGQB	LGQN	SGCN	SGQN
21.	LGCB	LRQB	SRCB	SGQN	LGQB	SRQB	LRCN	SGCB	LGQN

Code: L: large, R: red, C: circle, B: border, S: small,

G: green, Q: square, N: no border.

RESULTS

Test of Models

Detailed predictions may be derived from some current models for the experimental conditions of the present study. Several such predictions are evaluated below.

The first of these follows from the assumption (Bower and Trabasso, 1964) that subjects begin a concept identification task in a guessing state and remain in that state until an error occurs, at which time they select a hypothesis. In the RRR condition it follows that a subject cannot have solved the problem at the end of three trials, since there have been no errors. The probability that a subject in this condition makes a correct response on any trial after the third is then 1/2, provided that no error has occurred. The probability of making no errors on the remaining six trials is $(1/2)^6=1/64$. The observed proportions of errorless solutions for the RRR condition were 0.836 for Experiment 1 and 0.869 for Experiment 2. These observations were based on 61 subjects in each experiment, each performing on one RRR problem in Experiment 1 and on two in Experiment 2 and so the proportions clearly are reliably different from 1/64.

For the WRR condition predictions from two models (Bower and Trabasso, 1964; Trabasso and Bower, 1966) are equivalent and quite clear. Since subjects responses were called "wrong" on trial 1, these models assume that selection of a new cue occurred after that trial. The models assume "local consistency," and so the subject's selection of a cue following an error trial must, according to this assumption, be consistent with the information provided by that trial. Since the correct hypothesis in this condition is selected by the experimenter so that it agrees with the subject's choices on trials 2 and 3, the models predict errorless solutions with probability 1. The observed proportions of errorless solution were 0.738 for Experiment 1 and 0.694 for Experiment 2. Both proportions are reliably different from 1.

The Restle (1962) model does not include a consistency check on the error trial, and so predicts only that all responses will be consistent with trials 2 and 3 in such fixed outcome problems. The proportion of WRR problems in Experiment 1 for which this two-trial consistency held was 0.82.

The Bower and Trabasso (1964) model assumes consistency checks after errors, in which the cue to be selected is checked against the error-trial information only, and so for the WWR condition this model predicts that all responses will be consistent with trials 2 and 3, but not necessarily with trial 1. The same prediction holds for RWR, since the second-trial consistency check looks back to the correct choice on trial 2 and the experimental procedure ensures agreement with trial 3 for this condition as well as for WWR. The proportion of WWR problems for which consistency with trials 2 and 3 was found was .75 and the corresponding proportion for RWR problems was .79. Ninety-nine per cent confidence intervals for the probabilities associated with these proportions were computed by means of the normal approximation to the binomial distribution. Since both intervals lie below .88, it is apparent that the probabilities are not near 1.

A pair of models developed more recently yield the same predictions for the two conditions discussed just above. Trabasso and Bower (1968) presented two models which differ in their predictions about behavior such as learning a second redundant relevant cue, but cannot be discriminated on the basis of manipulations of the outcomes on the first three trials, as in the present experiment. These models assume consistency checking against the error trial as does the Bower and Trabasso (1964) model and, although they assume multiple-hypothesis processing, their prediction for this case is similar to that of the 1964 model. After the second trial the subject is assumed to select a new "focus sample" without regard to trial 1 information. According to these models, the sample is then narrowed down on correct trials by discarding those hypotheses inconsistent with the chosen stimulus. Consequently there are at least one, at most two hypotheses left in the sample after trial 3. If there is one hypothesis, the subject's responses are consistent with it, and perfect consistency with trial 2 results. If there are two hypotheses, the subject's response is consistent with both of them on each trial until a trial occurs on which they are placed in opposition. When they are opposed, the subject narrows the focus by discarding one of them and so retains the remaining hypothesis throughout the remainder of the problem. In either case, then, every response is consistent with the trial 2 response and with trial 3 as well. This is the same prediction made by the 1964 model for the WWR condition and the RWR condition. The observed proportions were .75 and .79, as stated in the preceding paragraph.

The model (Trabasso and Bower, 1966) that relinquished the sampling-with-replacement axiom also added consistency checking against the trial preceding the error trial. For the WWR and RWR conditions, according to this model, a consistency check occurs, comparing trials 1 and 2, eliminating any cue that is inconsistent with those outcomes, and selecting a new cue value that agrees with trial 1 and trial 2 outcomes. Consistency with trials 1 and 2 is thus assured by the subject's behavior and consistency with trial 3 is generated by the experimental procedure. This model therefore predicts errorless performance after trial 3 with probability 1 for both RWR and WWR. The observed proportions of errorless performance of these conditions were 0.694 and 0.410, respectively.

The outcome combinations still to be considered are those with a "wrong" on trial 3 (XXW). The Bower-Trabasso (1964) model predicts for this condition that all responses will be consistent with trial 3 information. By the same argument given above, with respect to trial 2 consistency in the XWR conditions, the two more recent multiple-hypothesis models (Trabasso and Bower, 1968) yield the same prediction for this condition. The proportion of trial-three-consistent protocols observed for XXW conditions was 0.795.

Since it assumed consistency checking against the trial before the error trial, the Trabasso-Bower (1966) model predicts perfect consistency with trial 2 as well as trial 3 in the XXW case. The observed relative frequency of such consistency on XXW problems was 0.504.

Lag Between Complementary Stimuli

Another test of local consistency is that described by Kenoyer and Phillips (1968), in which consistency is indicated by the subject's matching responses on complementary stimuli. A description of this procedure and its rationale was given in a previous chapter. The finding by Kenoyer and Phillips that matches occurred with probabilities different from 1 was confirmed in the present study. The present design also considers two values of lag (the number of trials intervening between complementary stimuli). In Experiment 1 the relative frequencies of matches were 0.702 for lag 1 and 0.586 for lag 5. The decrease over lag is significant, and indicates some loss of information over trials. Although this loss could be interpreted as a forgetting process, further examination of the data suggests another possibility.

Whenever errorless solution occurs, the choice responses corresponding to the complementary stimuli necessarily match. Since the criterion for correct responding is established partly by the trial on which the first member of the complementary pair (C_1) is presented, a correct response to the second member is necessarily the same response that is called "wrong" for the first member. Therefore, any subject who solves the problem before the presentation of the second member of the complementary pair (C_2) scores a match on that problem. Methods have not been devised for identifying all subjects who solve before the trial on which C_2 is presented, but some improvement can be effected by eliminating those subjects who solve with no errors after the third trial.

It is useful, therefore, to examine the conditional proportion of matches given that errorless solution does not occur. For Experiment 1 the conditional proportion was 0.498 for the lag 1 condition and 0.328 for lag 5. If no information from the error trial and subsequent trials were utilized, the corresponding probability would be .5. The lag 1 proportion is not significantly different from this chance level, but the lag 5 proportion is below chance. The number of observations (i.e., the number of problems with at least one error) from which these proportions were computed was 479.

The finding that the lag 5 proportion was below chance suggests that the decrement is not simple forgetting. If it is regarded as information loss, it must be attributed to misinformation. A plausible source of misinformation in Experiment 1 is the series of "right" reinforcements between the two complementary stimuli. If subjects do process information on those trials, then any response that is not consistent with the hypothesis established on the first three trials leads to category information that is inconsistent with the established hypothesis.

Experiment 2 did not provide this potential source of misinformation, since feedback after the first three trials was contingent on the response, and feedback on the first three trials, though arbitrary, was necessarily consistent with the correct hypothesis. If the specified kind of misinformation did occur in Experiment 1, the conditional probability of a match should be greater in Experiment 2. The local consistency assumption, on the other hand, predicts a lower conditional probability of a match in Experiment 2, since each error trial is assumed to "restart" the subject. The conditional relative frequency observed in Experiment 2 was 0.682, which was reliably greater than that for either lag in Experiment 1. This proportion was calculated for a sample of 330 instances. The sample size was smaller in Experiment 2 because that experiment was not designed primarily to gather data on matches, and consequently the first of the complementary pair of stimuli did not always coincide with an error trial. The variation was not due to differences in the number of subjects who made errors; the number of subjects making at least one error averaged over conditions was approximately 36 in Experiment 1 and approximately 37 in Experiment 2.

General Results

Although the major emphasis in this study was on the evaluation of models and of certain theoretical assumptions, several results should be reported because of their relevance to other questions that may be raised about the study. Such results are included in this section of the Results chapter.

In Experiment 1 subjects were not informed of errors on trials after the third in most of the problems (problems 3-9, 12-13), but were told "right" regardless of their responses on these trials. Regardless of the effectiveness of subjects' initial strategies, feedback indicated perfect performance, and so there was no apparent need to improve. In this situation it seems reasonable to expect little or no improvement in actual performance. This expectation was checked by means of two dependent variables, a binary indicator variable indicating either that one or more errors occurred (1) or that solution occurred without error (0), and the number of errors occurring after the third trial. By "error" is meant a response that is not consistent with the hypothesis established on the first three trials. Subjects were not informed of these errors in Experiment 1. The comparison was between the mean for the first half of the problems and the mean for the last half (problems 3-9, 12-18). The relative frequency of at least one error was 0.550 on the first half and 0.517 on the second half. The mean numbers of errors were 1.852 and 1.813 for the first and last halves of the problems, respectively. The difference between neither of these pairs of numbers is significant. It may be noted that while the number of errors decreased over problems, the relative frequency of at least one error increased slightly. The comparisons were based on data from 61 subjects.

The situation was different for Experiment 2. On all but the two filler problems, consistent feedback was provided on all trials. Under these conditions it is reasonable to expect some improvement over problems. The same dependent variables described just above were used, as well as a third variable, an indicator variable which took the value 1 if the subject verbalized the hypothesis correctly after the problem, or 0 otherwise. The probability of at least one error was .502 for the first half (problems 5-12) and .395 for the last half (problems 14-21). The mean number of errors was 1.256 for the first half and 1.029 for the last half. The probability of correct verbalization was .730 for the first half and .793 for the later ones. None of these differences is significant although all are in the proper direction to indicate improvement. Problems 4 and 13 (the filler problems) were excluded from the analysis. Warmup problems (1-3) were also excluded.

Another indicator of change in performance is match frequency. If a subject becomes more efficient in encoding the stimulus, he may be expected to retain more information about the stimulus over trials. If so, the redundancy in the second member of a complementary pair of stimuli (C_2) would result in an increasing tendency to respond correctly when C_2 is presented, and match frequency would increase. In Experiment 1 the relative frequency of a match was .567 for the first half and .520 for the second half. The difference is not significant.

Matching responses on complementary stimuli are not independent of errorless solution. If solution occurs at any time before C_2 (the second member of the complementary pair) is presented, a correct response, and therefore a match, occurs on that trial. It is possible that the effect of lag on match frequency may be due in part to the effect of lag on the proportion of errorless solutions. The effect of lag on the proportion of errorless solutions was therefore assessed. Proportions of problems with at least one error before solution appear in Table 6, in which rows are experimental conditions defined by the outcomes on the first three trials of the problem, and columns are lag conditions. The marginal proportions for the two lag conditions were .612 for lag 1 and .618 for lag 5. The difference is not significant, and seems too small to mediate effects of any consequence.

The marginals for outcome conditions vary more strongly. The variability among these conditions was significant (χ^2=34.18, df=5). The observations on which the chi-square was computed were on the same subjects, and the independence assumption underlying the use of chi-square is therefore questionable. However, the result of the test serves as an indication of rather large variability among the proportions.

Table 6

Proportions of Problems of Which At Least

One Error Occurred, By Experimental Conditions

	Lag 1	Lag 5	Row Mean
WWW	.721	.852	.787
WWR	.459	.410	.435
WRW	.721	.721	.721
RWW	.770	.754	.762
WRR		.311*	
RWR	.459	.426	.443
RRW	.541	.541	.541
RRR		.164*	
Column Mean	.612	.617	.615

*The WRR and RRR conditions were excluded from the analysis, since both lags were not used for these conditions.

The number of VEK presentations in the first three trials is a dependent variable, since the outcomes on those trials are predetermined and the classification on each trial is jointly determined by the response and the predetermined outcome. The number of VEK presentations was counted for each subject and outcome condition, and intercorrelations were computed among these numbers. The intercorrelations appear in Table 7. It is apparent that the problems with the same first-trial outcome intercorrelate positively and that the correlations between these and the problems with the opposite first-trial outcome are negative, although the correlations are not large. Under the fixed-outcome condition that characterizes these first three trials, the stimulus is what the subject calls it (VEK or NONVEK) on "right" trials and the opposite of what he calls it on "wrong" trials. The correlation pattern suggests, then, that individual subjects tend to choose VEK or NONVEK consistently on these first three trials.

The following procedure was used to evaluate this conjecture. Correlations were computed on a binary variable indicating VEK (1) or NONVEK (0) for the first trial. The intercorrelations among problems are shown in Table 8 for group 1. With few exceptions (9 out of 162), the correlations are positive, and many of them are greater than .352, which is the smallest correlation that is significantly different from zero at the .05 level for 31 subjects. The correlations for group 2 appear in Table 9. Here there is only one negative correlation and again several of the correlations are significantly greater than zero. For 30 subjects, $\underline{r}$ is significant at the .05 level when $\underline{r} > .358$. These correlations indicate some individual consistency in the selection of a

Table 7

Correlations Among Numbers of VEK Presentations

in Experimental Conditions

(All Correlations Multiplied by 100)

WWW	100	44	22	10	-25	-54	-22	-07
WRW	44	100	37	13	-29	-51	-15	-18
WRW	22	37	100	06	12	-26	-02	02
WRR	10	13	06	100	-06	-07	-36	05
RRR	-25	-29	12	-06	100	18	20	21
RRW	-54	-51	-26	-07	18	100	15	19
RWR	-22	-15	-02	-36	20	15	100	08
RWW	-07	-18	02	05	21	19	08	100

Table 8

Correlations Among First-Trial Responses Over Eighteen Problems

Group 1
(All Correlations Multiplied by 100)

Problem Number

Problem Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	100	22	-4	19	22	9	6	26	-19	14	15	-4	11	35	4	19	0	-17
2	22	100	-3	36	4	22	30	16	18	-3	29	22	10	35	-10	17	24	35
3	-4	-3	100	9	18	61	17	29	31	24	29	10	49	10	29	43	24	22
4	19	36	9	100	13	-4	8	10	13	4	28	9	13	36	-9	8	4	23
5	22	4	18	13	100	18	46	31	30	38	54	18	41	18	36	33	52	18
6	9	22	61	-4	18	100	17	42	18	24	55	22	36	35	42	56	50	48
7	6	30	17	8	46	17	100	0	19	13	25	4	15	30	35	34	40	17
8	26	16	29	10	31	42	0	100	17	36	48	56	46	42	37	54	36	56
9	-19	18	31	13	30	18	19	17	100	24	40	31	55	18	9	33	38	31
10	14	-3	24	4	38	24	13	36	24	100	6	37	34	24	56	13	32	11
11	15	29	29	28	54	55	25	48	40	6	100	29	56	42	23	38	32	42
12	-4	22	10	9	18	22	4	56	31	37	29	100	10	35	42	30	37	48
13	11	10	49	13	41	36	15	46	55	34	56	10	100	23	29	41	34	10
14	35	35	10	36	18	35	30	42	18	24	42	35	23	100	15	56	11	35
15	4	-10	29	-9	36	42	35	37	9	56	23	42	29	16	100	22	29	29
16	19	17	43	8	33	56	34	54	33	13	38	30	41	56	22	100	53	43
17	0	24	24	4	52	50	40	36	38	32	32	37	34	11	29	53	100	37
18	-17	35	22	23	18	48	17	56	31	11	42	48	10	35	29	43	37	100

Table 9

Correlations Among First-Trial Responses Over Eighteen Problems

Group 2
(All Correlations Multiplied by 100)

Problem Number

Problem Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	100	24	47	24	33	29	31	25	43	43	19	28	38	64	42	5	53	38
2	24	100	13	5	53	38	28	24	38	30	40	34	28	31	34	26	40	4
3	47	13	100	26	33	46	61	19	5	22	33	40	47	54	40	20	37	53
4	24	5	26	100	13	11	28	52	38	58	40	47	28	31	20	26	24	32
5	33	53	33	13	100	46	32	33	33	49	20	53	21	40	40	33	37	26
6	29	38	46	11	46	100	39	29	3	39	46	54	49	51	54	5	44	25
7	31	28	61	28	32	29	100	15	24	36	17	36	70	35	22	47	15	57
8	25	24	19	52	33	29	15	100	29	58	47	57	24	34	42	47	53	24
9	43	38	5	38	33	3	24	29	100	39	5	41	8	23	14	46	37	38
10	43	30	22	58	49	39	36	58	39	100	35	41	46	48	41	35	24	30
11	19	40	33	40	20	46	17	47	5	35	100	27	34	26	67	6	37	-14
12	28	34	40	47	53	54	36	57	41	41	27	100	40	62	33	27	50	47
13	38	28	47	28	21	49	70	24	8	46	34	40	100	57	40	34	30	41
14	64	31	54	31	40	51	35	34	23	48	26	62	57	100	48	12	48	59
15	42	34	40	20	40	54	22	42	14	41	67	33	40	48	100	27	50	20
16	5	26	20	26	33	5	47	47	46	35	6	27	34	12	27	100	20	53
17	53	40	37	24	37	44	15	53	27	24	37	50	30	48	50	20	100	24
18	38	5	53	32	26	25	57	24	38	30	-14	47	41	59	20	53	24	100

given categorization response on these pre-solution trials. The means (which are, of course, probabilities) do not reveal this nonrandom behavior. For group 1, the proportions of VEK responses were .468, .516, and .556 for the three trials, respectively. For group 2, the corresponding proportions were .554, .568, and .493. From these proportions it is reasonable to infer that there is no preference in the population of subjects for either response. The correlations indicate, however, that there are consistent preferences at the individual level which are not apparent in group means. The above results support the contention that subjects begin such problems nonrandomly rather than in a guessing state. Some of the variability of early choice responses is therefore accounted for by response preference.

Another potential source of behavior regularity that was investigated is the correlation of responses with the presentation of cue values. Trabasso and Bower (1964) have dealt with the tendency of groups of subjects to select a given cue by including cue weighting parameters in their models. The present method, however, deals with tendencies of individual subjects to assign stimuli with a given property to a given category. If a subject tends to assign large stimuli to VEK, for example, then his VEK responses are, in a loose sense of the term, "correlated" with the appearance of large stimuli. If VEK and NONVEK are coded as 1 and 0, respectively, and size is coded so that 1 indicates large and 0 indicates small, the stimulus and response are quantified and the term "correlation" can be applied in the more rigorous sense of the Pearson product-moment correlation. A positive correlation between classification and size then indicates a tendency to emit VEK responses when large stimuli are presented, and a negative correlation indicates the opposite classification preference. For each subject a correlation can be obtained between his string of responses and the string of values for each cue.

The string of numbers identifying trial numbers within problems and the string of problem numbers can also be correlated with the response variables just described. A positive correlation between a subject's responses and trial numbers indicates a greater tendency to emit VEK responses (coded 1) on later trials than on earlier trials, while a negative correlation indicates a decreasing preference for the VEK response. Either a positive or a negative correlation may then be taken to indicate a change in response preference over trials. Similarly, nonzero correlations between the response variable and problem number indicate a shift in response tendency over problems.

The correlations described above were computed for a limited set of trials. The set of trials that were of interest are those over which the subject cannot reasonably be expected to change hypotheses and for which sufficient information has not been provided for solution to the problem. Therefore no trials after the first three were included and, of the first three, only those trials that were not preceded by a "wrong" outcome were used in this analysis. A nonzero correlation between any cue and the classification response therefore serves as a measure of the contingency relation between an individual's responses and the presence of a particular cue <u>value</u>. If subjects began problems consistently with the same cue (e.g., size) but alternately classified small stimuli as VEK and large stimuli as VEK consistent selection of a cue would not necessarily yield nonzero correlations, but the stronger consistency, i.e., a consistent contingency relation between classification response and cue value over observations, appears as a nonzero correlation.

For this analysis it is necessary to consider subjects, cue values, trial number, and problem number as variables. Observations of values taken on by these variables are taken over different occasions (trials). The portion of the correlation matrix that shows inter-correlations among subjects is not relevant to the analysis, since the object is not to identify similar response strings. The part that is of interest is the set of intercorrelations between response strings and the other variables, and the intercorrelations among the non-subject variables.

Twenty-four trials were used in the analysis of each of the two groups. The critical value for the correlations between qualitative variables (phi coefficients) with this sample size is .40. Most of the correlations with cues do not reach significance, but there are a few exceptions. For example, the classificational responses of Subject 16 in group 1 correlated significantly with both size and border (Table 10).

An especially impressive regularity is indicated by the entries for Subject 11 in Table 11. PACKAGE (Cf. Hunter and Cohen, 1969), the set of correlational programs used for this analysis, enters "900" in the correlation table for variables with zero variance. Subject 11 made the same response on all trials used in the analysis. Inspection of the data card showing the classification responses for the first three trials for all 18 problems revealed that all but two of the 54 responses were VEK.

Table 10

Correlations of Subjects' Responses With Cue

Values, Trial, and Problem Number

Group 1
(All Correlations Multiplied by 100)

Subject	Large	Red	Circle	Border	Problem	Trial
1	9	9	32	39	33	-23
2	-16	16	-5	30	-29	-3
3	- 8	41	-18	-4	3	21
4	-21	-7	-27	-2	16	11
5	-21	-21	-11	-16	11	21
6	-14	-29	-25	-19	2	15
7	-47	-0	-13	2	-14	85
8	-22	22	2	20	-18	-12
9	14	0	-16	-7	13	-5
10	28	9	12	-18	14	4
11	-7	-7	2	-25	34	31
12	-0	-14	-25	-34	-8	26
13	-32	-0	5	-14	-64	38
14	-7	21	11	-13	7	-11
15	7	-7	-2	-5	-1	-1
16	45	-0	2	47	25	-37
17	-0	-0	13	34	38	-22
18	8	8	-7	-39	16	9
19	7	-22	2	20	-20	10
20	35	-12	29	19	-10	-61
21	7	-21	-11	-31	-28	0
22	36	21	-5	2	-15	-32
23	-0	-15	2	-29	8	-37
24	-25	-8	0	-4	-23	21
25	-21	-7	-5	-13	9	21
26	28	-0	44	34	-6	-63
27	-28	28	-12	18	-28	64
28	-41	8	-18	13	8	58
29	7	-7	2	20	-9	-1
30	-14	14	0	-7	-53	-5
31	-22	-7	2	5	-55	31
Large	100	-14	0	-7	5	-47
Red	-14	100	16	7	-5	16
Circle	0	16	100	-18	11	-27
Border	-7	7	-18	100	-8	1
Problem	5	-5	11	-8	100	-4
Trial	-47	16	-27	1	-4	100

Table 11

Correlations of Subjects' Responses With Cue

Values, Trial, and Problem Number

Group 2
(All Correlations Multiplied by 100)

Subject	Large	Red	Circle	Border	Problem	Trial
1	-7	-14	16	-8	39	26
2	-21	-14	-16	38	-8	5
3	16	7	2	-23	26	20
4	-6	-0	23	-18	23	-6
5	14	12	4	4	34	-44
6	-6	-29	-25	13	-15	15
7	14	-0	14	-20	-8	-10
8	-32	-12	29	-24	-18	-27
9	39	-15	14	3	7	4
10	-9	-12	4	1	0	-44
11	900	900	900	900	900	900
12	-8	-0	-7	2	-0	17
13	-23	-0	-23	33	-21	6
14	6	-14	-7	33	31	-15
15	-4	35	-6	36	10	54
16	20	-0	-31	-3	-16	-15
17	6	9	-12	-8	-25	-18
18	-11	0	-40	92	-8	32
19	-14	-7	-11	-43	-12	10
20	-14	7	-20	33	2	-0
21	-29	7	-11	13	3	-10
22	-3	20	-42	28	-12	77
23	14	-50	-27	13	0	-10
24	13	9	32	-12	13	-50
25	2	-28	44	-19	27	-63
26	39	17	44	-36	5	-16
27	2	-28	44	-19	27	-63
28	-10	30	19	13	-14	-4
29	9	12	-4	-1	-2	61
30	-29	21	20	13	8	0
Large	100	-7	5	-18	4	-21
Red	-7	100	16	8	14	26
Circle	5	16	100	-44	16	-27
Border	-18	8	-44	100	4	36
Problem	4	14	16	4	100	-4
Trial	-21	26	-27	36	-4	100

There are large correlations with both trial number and problem number. Unfortunately, the assumptions necessary to determine a significance level for these correlations cannot be justified. The magnitudes of some of the correlations are such as to suggest, however, that the subjects with which they are associated shifted their response preferences over problems or over trials.

Evaluation of the IHE Models

The probability of each possible sequence of errors and correct responses can be generated from the IHE models, and so it is possible, in principle, to test the fit of the models against observed frequencies of the error protocols. The procedure is not feasible for the present study, however, since for 6 trials there are $2^6=64$ possible sequences. There are three fixed outcomes at the beginning of each problem and six response-contingent outcomes constitute the remainder of the problem. Each of these conditions can therefore yield $2^6=64$ different outcome sequences for the trials after the third. There are only two observations on each condition for each subject, for a total of 122 observations on each condition. The number of observations for each condition is therefore less than twice as large as the number of categories. This ratio is not sufficiently large for minimum chi-square methods of estimating parameters. The usual way of avoiding this problem is to consolidate categories.

One way of consolidating in the present study is to place all protocols for a given condition having the same trial of last error in the same category. This procedure yields a separate probability distribution for trial of last error for each condition. There are only seven categories (one for errorless runs on the last six trials). Given that the expected frequencies for a given set of parameter values are all sufficiently large, a chi-square measure of goodness of fit is reasonable.

A learning curve can also be obtained for each condition. For each protocol generated by the model, the probability of that protocol is added to each point on the error probability curve (learning curve) where the protocol shows an error. For example, a probability must be generated for the sequence 111000111 (where 1 indicates an error, 0 a correct response), as well as for all other sequences of the same length. This protocol can occur only in the WWW condition, as indicated by the errors on the first three trials. Then for the WWW condition, the probability that this sequence occurs is added to error probabilities for trials 7, 8, and 9. When all possible sequences have been tallied in this way, the resulting probabilities constitute the learning curves for the eight experimental conditions.

The procedure used for this test was similar to that described for the preliminary test discussed previously, in which Levine's (1966) data were used. Test values of the parameters were entered as data into a Fortran program, and theoretical (predicted) learning curves and trial of last error (TLE) curves were generated. A chi-square statistic was computed for each of the two curves for each experimental condition, and the sum of these chi-squares was taken as the indicator of goodness of fit.

Each of the eight experimental conditions then had seven TLE data points and six learning curve points, for a total of 104 data points.

Computing the theoretical curves and the chi-square statistics for so many points is obviously more time-consuming than the preliminary analysis. Results of that analysis were therefore used to simplify the present one. Since the best estimate of the parameter <u>c</u> was 1.00 for both IHE Model 2 and IHE Model 5 in the preliminary test, this parameter was not varied in the present test. With <u>c</u>=1.00 for these two models, each had only two parameters, <u>w</u> and <u>f</u>. As in the preliminary test, IHE Model 4 had the two parameters <u>w</u> and <u>c</u>.

The parameter values yielding the best fit, the chi-square values for the TLE curve and the learning error curve for each condition, and the sum of the chi-squares appear in Table 12 for IHE Models 2, 4, and 5. The minimum sum found for IHE Model 2 was 1100., that for IHE Model 5 was 643., and that for IHE Model 4 was 266. There are thirteen frequencies (twelve degrees of freedom) for each condition, and eight conditions, yielding 96 degrees of freedom for the overall chi-square before correction for estimated parameters. The sums for all three IHE models are therefore tested against chi-square with ninety-four degrees of freedom. All three models then, deviate from the observed data of the present study sufficiently to be rejected beyond the .001 level.

Of the three models, IHE Model 4 is clearly superior. For this model, the observed and expected proportions for trial of last error appear in Table 13, and those for the mean error curve are presented in Table 14. The predicted proportions shown are all generated from the parameter values shown in Table 12, those that yielded the minimum sum of chi-squares. Although the model does not fit adequately by this

Table 12

Chi-squares for Fit of Learning Curves and TLE Curves to Each Experimental Condition for IHE Models 2, 4, and 5

	RRR	RRW	RWR	RWW	WRR	WRW	WWR	WWW	Sums	Para-meters
IHE Model 2										
Trial of Last Error	120.	37.	65.	18.	62.	17.	44.	18.	281.	W=.99
										F=.03
Learning Curve	43.	8.	14.	213.	15.	249.	8.	170.	719.	
									1000.	
IHE Model 4										
Trial of Last Error	2.	48.	20.	11.	6.	11.	25.	14.	139.	W=.73
Learning Curve	0.	32.	21.	7.	5.	11.	33.	17.	127.	
									266.	
IHE Model 5										
Trial of Last Error	35.	40.	58.	32.	57.	28.	104.	39.	393.	W=.55
										F=.03
Learning Curve	9.	15.	16.	53.	15.	56.	52.	43.	250.	
									643.	

Table 13

Expected and Observed Proportions for Trial of Last Error in Each Experimental Condition

IHE Model 4

Trial Number

		3	4	5	6	7	8	9
RRR	Expected	1.	.0	.0	.0	.0	.0	.0
	Observed	.877	.032	.000	.024	.000	.016	.049
RRW	Expected	.353	.113	.100	.092	.091	.104	.147
	Observed	.631	.065	.032	.032	.016	.114	.106
RWR	Expected	.557	.087	.071	.062	.060	.068	.095
	Observed	.721	.073	.016	.040	.008	.032	.106
RWW	Expected	.254	.115	.112	.107	.109	.125	.178
	Observed	.286	.073	.098	.106	.040	.147	.245
WRR	Expected	.757	.051	.040	.034	.032	.036	.050
	Observed	.737	.016	.057	.040	.024	.057	.065
WRW	Expected	.300	.116	.106	.100	.100	.115	.163
	Observed	.270	.098	.073	.057	.098	.180	.221
WWR	Expected	.444	.105	.088	.078	.076	.087	.122
	Observed	.655	.057	.081	.040	.015	.049	.098
WWW	Expected	.226	.114	.114	.112	.114	.132	.188
	Observed	.262	.057	.057	.090	.114	.131	.286

Table 14

Expected and Observed Proportions of Mean Errors in Each Experimental Condition

IHE Model 4

Trial Number

		4	5	6	7	8	9
RRR	Expected	.0	.0	.0	.0	.0	.0
	Observed	.073	.024	.065	.049	.057	.049
RRW	Expected	.387	.313	.259	.215	.177	.147
	Observed	.172	.163	.147	.163	.188	.106
RWR	Expected	.253	.204	.168	.139	.115	.095
	Observed	.114	.131	.114	.057	.090	.106
RWW	Expected	.451	.376	.312	.259	.215	.178
	Observed	.434	.401	.278	.270	.286	.245
WRR	Expected	.135	.109	.089	.074	.061	.050
	Observed	.122	.114	.090	.098	.098	.065
WRW	Expected	.422	.345	.286	.237	.196	.163
	Observed	.483	.352	.352	.303	.270	.221
WWR	Expected	.330	.263	.215	.178	.147	.122
	Observed	.106	.180	.139	.073	.122	.147
WWW	Expected	.468	.395	.329	.273	.226	.188
	Observed	.327	.352	.327	.377	.254	.286

criterion, it accounts for 91 per cent of the variance among the 56 proportions in the TLE curve and 97 per cent of the variance among the 48 proportions in the mean learning curve.

DISCUSSION

Current Models

One of the purposes of this research was to evaluate certain assumptions of current models of concept identification. Although some of these assumptions had been tested in other contexts, the procedure of the experiments reported here led to particularly strong predictions from the assumptions tested, and so was expected to provide a sensitive test of them.

The simplest of these assumptions was that subjects begin a problem with no hypothesis and select hypotheses only after error trials. This assumption was included in Bower and Trabasso (1964) model for mathematical convenience rather than for substantive reasons, but its evaluation by the relatively direct method of the present study seemed appropriate. The finding, in the present study, that subjects in the RRR condition performed without error with high probability extends the findings of Levine (1966) and Richter (1965) by providing evidence for hypotheses at the outset of the standard concept identification situation.

Several deterministic predictions following from the local consistency assumption were tested and found to be in conflict with the data. This assumption, like the one discussed just above, may be analyzed into two process assumptions. The first of these is that the hypothesis selected after an error is consistent with the information on that trial. The second is that hypotheses are not abandoned on

correct trials. Previous research (Restle and Emmerich, 1966) has shown that repeated presentation of a stimulus on successive trials does not lead to perfect performance on the second presentation after an error on the first. The deviation from perfect local consistency was not large in that study, however. The greater discrepancies found in the present study are probably due to processes occurring over a longer series of trials, since the predictions tested here have to do with consistency over the remainder of a problem after an error trial. Predictions that were examined included (a) errorless performance in the WRR condition, (b) consistency with trials 2 and 3 in the RWR and WWR conditions, (c) consistency with trial 3 in the WWW, WRW, RWW, and RRW conditions, and (d) matching responses on complementary stimuli after either one or five intervening trials.

Effect of Lag on Matches

An earlier study (Kenoyer and Phillips, 1968) showed that the proportion of matching responses to complementary stimuli was not near 1 in general, and the present study added support to that finding with a larger sample of subjects and an exhaustive set of combinations of outcomes on the first three trials. The present study also varied the number of trials intervening between presentations of the two complementary stimuli (lag) independently of other variables, and so permitted a comparison of lag 1 and lag 5. The longer lag led to a lower proportion of matches, suggesting information loss over trials. Since all responses were called correct on the lag trials, this difference is inconsistent with the notion that correct trials have no effect on subjects. The importance of correct trials, indicated by performance differences in studies reported by Levine (1966) and Richter (1965), thus generalizes to a different concept identification paradigm.

Matches would occur in the present study whenever solution of the problem preceded the second complementary stimulus. Match proportions were therefore computed for those problems on which at least one error occurred. The mean of these conditional proportions was found to be significantly below .5, the chance level. In Experiment 1 all responses on lag trials were called correct irrespective of their consistency with the established concept, and so it was conjectured that this below-chance proportion of matching responses was due to misinformation on lag trials. Such misinformation can occur only if the subject is tracking more than a single hypothesis that determines his response. This explanation of the low match proportions implies that the proportions in Experiment 2 would not be below chance, since misinformative feedback was not given. This prediction was confirmed. There is some support, therefore, for this interpretation.

Alternative explanations for these data exist, of course. It is possible, for example, that subjects have preferred hypotheses that guide their early choice responses. A subject might, for example, prefer "red VEK." Then presentation of a LRCB stimulus would be followed by a VEK response with high probability. If the subject then forgot that the favored "red VEK" hypothesis had been eliminated, later presentation of the complementary stimulus SGQN would be followed by a NONVEK response with high probability, and a failure to match would occur frequently. Previous evidence for information processing on correct trials, however, lends support for the former explanation, while the generally low correlations found in the present study between early responses and stimulus dimensions do not lend support to the "preferred hypothesis" explanation.

Effect of Outcome Sequence on Difficulty

Data on proportions of problems with one or more errors indicated that the difficulty of a problem depends significantly upon the outcome sequence on early trials. The most difficult condition was WWW and the least difficult was RRR (Table 2). Previous research (Levine, 1966; Richter, 1965) has indicated that some of these experimental conditions were more facilitative than others under rather special circumstances. The results of the present study show that the ordering of these tasks on difficulty generalizes to the usual kind of concept identification experiment, in which post-solution performance is emphasized. Besides extending these findings on task difficulty to a new situation, the present study has also elaborated the set of conditions investigated. Richter did not manipulate outcomes as an independent variable and Levine reported only the RRW and WWW conditions and combined data from the RWW and WRW conditions. The present study deals with all eight possible R-W sequences over the first three trials.

IHE Models

The IHE models developed in the present investigation were subjected to rather rigorous criteria for acceptance. First, each model was constructed so as to be consistent with recent evidence for

(a) multiple hypothesis processing, (b) differential information processing on correct and error trials, (c) failure of strict local consistency, and (d) failure of the assumption that errors serve to eliminate the effects of previous trials, "restarting" subjects.

Second, the IHE models, along with Chumbley's (1969) HM model, were tested against data from Levine's (1966) experiment. Some of the models, IHE Models 1 and 3, displayed qualitative characteristics that were in conflict with available data and were pursued no further, although only one of the models (IHE Model 5) fit the Levine data adequately, IHE Models 2, 4, and 5 were all consistent with qualitative criteria, and were all tested against the data of Experiment 2 of the present study.

This last test of the models yielded several interesting results. The first is the finding that IHE Model 4 gave the best fit, rather than IHE Model 5, which fit Levine's data best. Although any interpretation of this kind of finding should be made with caution, such a result is consistent with certain differences between the two experimental situations. In the Levine experiment, four blank trials intervened between successive feedback trials, and the hypothesis assumed to govern the subject's response was inferred from the series of blank trials. It is therefore plausible that forgetting of eliminated hypotheses over the series of blank trials could be due primarily to mental activity during the blank trials and hence be virtually unaffected by outcomes. IHE Model 5 assumed an elimination operator that depends upon the nature of the outcome, but its forgetting operator is the same for every outcome trial, regardless of the nature of the outcome.

Thus this model seems more appropriate for such an experiment than IHE Model 4, in which the forgetting operator is determined by the outcome.

In the present research no blank trials are administered. Forgetting therefore occurs only during a feedback trial. It is reasonable, in this case, to expect any differential cognitive strain due to trial outcomes to affect the forgetting of eliminated hypotheses as well as the elimination of hypotheses. IHE Model 4, in which the probability of forgetting an eliminated hypothesis is determined by the trial outcome, fits these data better than IHE Model 5. This finding supports the contention that the forgetting processes are different for the two situations.

Another point of interest is the fit of IHE Model 4 to each condition. Although the model can be rejected on the basis of a chi-square fit to the data, the theoretical curve for each condition seems to resemble the data for that condition more than the data for other conditions. It may therefore be fruitful to consider other models that are similar to it, perhaps taking additional sources of variation into account by including additional parameters.

It is also interesting to note that the best estimate of $\underline{c}$ in IHE Model 4 was 1.00. One implication of this result is that the model attained the degree of fit described earlier without assuming any loss on correct trials, either of information from that trial or of previously eliminated hypotheses. In terms of simplicity of the model, this result means that only the parameter $\underline{w}$ remains. The development of new models by adding parameters is therefore more feasible than for a

model with two parameters, since the time required for parameter estimation increases exponentially with the number of parameters.

Two directions for further model development were suggested by results in this study. The first was suggested by the evidence that subjects vary substantially in terms of strategies. In view of this variation, it may be more fruitful to attempt to fit large behavior samples for individual subjects rather than to extend a model to a large population of subjects, all of whom must be described by the same parameter values. At the simplest level, this approach consists of application of the same model to all subjects, but with a new set of parameter estimates for each subject. At a second level, qualitatively different models may be necessary for different subjects. Bruner, Goodnow, and Austin (1956) found it helpful to classify subjects in two or more strategy categories. Their "successive scanner" category corresponds closely to the kind of subject described by Restle's (1962) and Bower and Trabasso's (1964) models, while their "focusser" corresponds to subjects described by the IHE models. If such categorial differences are used to determine which model is to be applied to each subject, it may be possible to improve fit considerably.

A second direction follows from the notion of a register model (e.g., Phillips, Shiffrin, and Atkinson, 1967), which formed the conceptual basis for the processes assumed in the IHE models. In IHE Models 2 and 5, the register for eliminated hypotheses and that for hypotheses being processed on the current trial were assumed to operate independently, and each was represented by a separate parameter. In IHE Models 3 and 5, the two functions were seen as shared in the same register, so that increased cognitive strain on error trials affected both alike, and both functions were represented by a single parameter. Of course, in both cases the probability operators at best only approximated what would be developed from a well specified register model. Actually imbedding a register memory process in the IHE models was seen as too complex at this stage of the research.

A second approximation can perhaps be obtained, however, by noting that two memory functions may share a common register, in the sense that they can displace each other, without having exactly the same probability parameter. In other words, one function may take priority over the other although both are subjected to the same stresses. The second approximation that will be attempted in subsequent research will simply include a parameter for adjusting the relationship between the two probability functions. The first and simplest of these will be a proportionality parameter relating the probability of remembering an eliminated hypothesis to the probability of hypothesis elimination. A register model, while difficult to formulate in this context, may be expected to be the end product in this line of development.

All of the models derived in this study include the assumption that the subject stores and imperfectly retains the set of rejected hypotheses. It was noted that memory for tenable hypotheses alone would lead to serious consequences if the subject forgot the correct hypothesis, since there would be no way of recovering it short of beginning the problem again with the whole hypothesis set. It is plausible, however, that the strategy of remembering both a list of eliminated hypotheses and a list of hypotheses not yet eliminated is employed. The problem of distinguishing between single-list and two-list models was beyond the scope of this study, but will become necessary if register models of hypothesis processing prove viable.

BIBLIOGRAPHY

BIBLIOGRAPHY

Atkinson, R.C., Bower, G.H., and Crothers, E.J. An introduction to mathematical learning theory. New York: Wiley, 1965.

Bower, G.H., and Trabasso, T. Concept identification. In Atkinson, R.C. (Ed.) Studies in mathematical psychology. Stanford: Stanford University Press, 1964.

Bruner, J.S., Goodnow, J.J., and Austin, G.A. A study of thinking. New York: Wiley, 1956.

Byers, J.L. Hypothesis behavior in concept attainment. Journal of Educational Psychology, 1965, 56, 337-342.

Chumbley, J. Hypothesis memory in concept learning. Journal of Mathematical Psychology, 1969, 6, 528-540.

Erickson, J.R., and Zajkowski, M.M. Learning several concept-identification problems concurrently: A test of the sampling-with-replacement assumption. 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: Analysis of latency data. Journal of Experimental Psychology, 1966, 72, 690-697.

Goodnow, J.J., and Pettigrew, J.F. Some sources of difficulty in solving simple problems. Journal of Experimental Psychology, 1956, 51, 385-392.

Gregg, L.W., and Simon, H.A. Process models and stochastic theories of simple concept formation. Journal of Mathematical Psychology, 1967, 4, 246-276.

Harlow, H.F. Learning set and error factor theory. In Koch, S. Psychology: A study of a science. New York: McGraw-Hill, 1959.

Haygood, R.C., and Bourne, L.E., Jr. Attribute and rule-learning aspects of conceptual behavior. Psychological Review, 1965, 72, 175-195.

Holstein, S.B., and Premack, D. On the different effects of random reinforcement and pre-solution reversal on human concept identification. Journal of Experimental Psychology, 1965, 70, 335-337.

Hunter, J.E., and Cohen, S.H. PACKAGE: A system of computer routines for the analysis of correlational data. Educational and Psychological Measurement, 1969, 29, 697-700.

Kenoyer, C.E., and Phillips, J.L. Some direct tests of concept identification models. Psychonomic Science, 1968, 13, 237-238.

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 behavior 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.

Martin, E. Concept utilization. In Luce, R.D., Bush, R.R., and Galanter, E. Handbook of mathematical psychology. Vol. III. New York: Wiley, 1965.

Phillips, J.L., Shiffrin, R.M., and Atkinson, R.C. Effects of list length on short-term memory. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 303-311.

Restle, F. The selection of strategies in cue learning. Psychological Review, 1962, 69, 329-343.

Restle, F., and Emmerich, D. Memory in concept attainment: Effects of giving several problems concurrently. Journal of Experimental Psychology, 1966, 71, 794-799.

Richter, M.L. Memory, choice, and stimulus sequence in human discrimination learning. Unpublished Doctoral Dissertation, Indiana University, 1965.

Spence, K.W. Continuous versus non-continuous interpretations of discrimination learning. Psychological Review, 1940, 47, 271-288.

Sternberg, S. Stochastic learning theory. In Luce, R.D., Bush, R.R. and Galanter, E. Handbook of mathematical psychology. New York: Wiley, 1963.

Trabasso, T.R. The effect of stimulus emphasis on strategy selection in the acquisition and transfer of concepts. Unpublished Doctoral Dissertation, Michigan State University, 1961.

Trabasso, T.R., and Bower, G. Pre-solution 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.R., and Bower, G.H. Attention in learning: Theory and research. New York: Wiley, 1968.

APPENDICES

APPENDIX A

APPENDIX A

INSTRUCTIONS READ TO SUBJECTS IN BOTH EXPERIMENTS

In this problem, we're interested in finding out how college students learn to classify patterns. For each set of patterns I will have in mind a classification rule, and your task will be to figure out what it is. There will be several of these tasks, each very short.

This is how we'll proceed. A pattern will be projected on the screen here in front of you, like that one (pointing). You will classify each picture as either VEK or NONVEK, and will indicate your choice by pressing the panel with the label corresponding to your decision. Either here (demonstrating) or here (demonstrating). These labels have no meaning; but are just convenient names for the two classifications. After you classify each picture, I'll say "RIGHT" or "WRONG." As we continue, you should be able to figure out a rule that will enable you to classify all the pictures correctly. The pictures have been randomly ordered, and so the order in which they appear has no bearing on your task.

From picture to picture the pattern can change in any of four ways so that there are four attributes to consider. The four attributes are: color: either red or green; shape: either a square or a circle; size: either large or small; and that's a large one; and border: either the figure has a white border or it has no border, like that one (pointing).

The solution to the problem will depend upon only one of these four attributes. By this, I mean that only one attribute is is crucial in your decision of how to classify the pictures.

Let me illustrate to you what I mean by using one attribute to classify a picture. This sample will not contain the pictures in your problem, but the principle is the same. That is, the classification depends upon only one attribute of the picture. (Holding card with figures before S.) If the classification rule I had in mind placed all hexagons in the VEK category and triangles in the NONVEK category, then I would say "RIGHT" if you indicated a hexagon to be a VEK or a triangle to be a NONVEK, or I would say "WRONG" otherwise, regardless of other characteristics of the figure.

Here is a card listing some information you should remember. Refer to it as often as you like throughout the experiment.

Do you have any questions?

There is one more procedural point I'd like to cover. You'll notice that this redlight (pointing) is on; this indicates that the box is turned off and so pressing the panels had no effect (demonstrating). When the box is turned on, the projector advances each time you press a panel. At the end of each of your tasks, I'll simply turn the box off, and you'll know the task is over when the red light comes on.

APPENDIX B

APPENDIX B

PROTOCOL BOOKLET FOR EXPERIMENT 1

DATE ____________________

TIME ____________________

DISSERTATION
EXPERIMENT 1
GROUP 1

(TASK 1)
WRR

- VEK
  - VEK
    - VEK(BORDERED) ___ ___ ___ ___ ___ ___
    - NONVEK(SQUARE) ___ ___ ___ ___ ___ ___
  - NONVEK
    - VEK(SMALL) ___ ___ ___ ___ ___ ___
    - NONVEK(RED) ___ ___ ___ ___ ___ ___

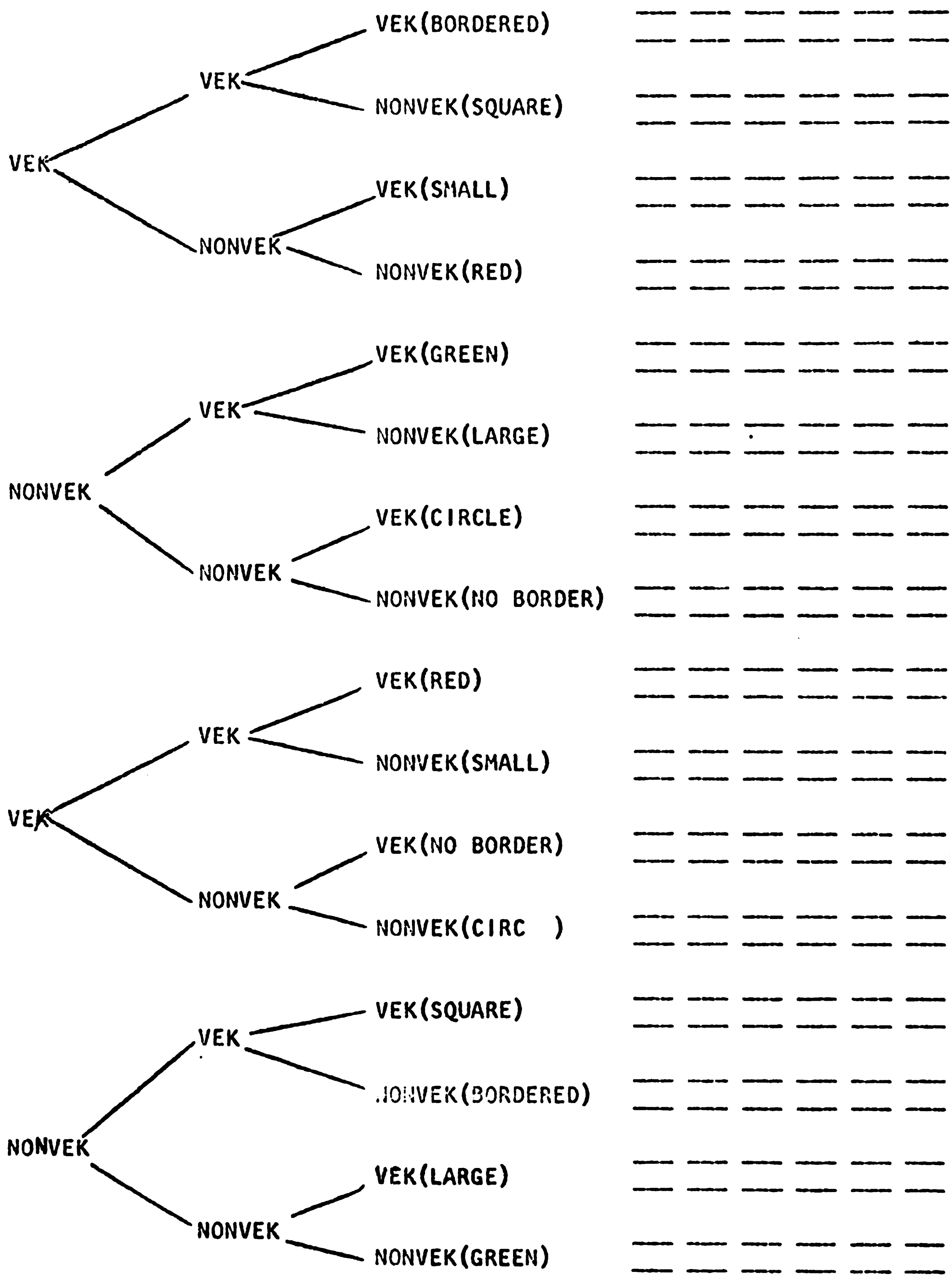
- NONVEK
  - VEK
    - VEK(GREEN) ___ ___ ___ ___ ___ ___
    - NONVEK(LARGE) ___ ___ ___ ___ ___ ___
  - NONVEK
    - VEK(CIRCLE) ___ ___ ___ ___ ___ ___
    - NONVEK(NO BORDER) ___ ___ ___ ___ ___ ___

(TASK 2)
WWR

- VEK
  - VEK
    - VEK(RED) ___ ___ ___ ___ ___ ___
    - NONVEK(SMALL) ___ ___ ___ ___ ___ ___
  - NONVEK
    - VEK(NO BORDER) ___ ___ ___ ___ ___ ___
    - NONVEK(CIRCLE) ___ ___ ___ ___ ___ ___

- NONVEK
  - VEK
    - VEK(SQUARE) ___ ___ ___ ___ ___ ___
    - NONVEK(BORDERED) ___ ___ ___ ___ ___ ___
  - NONVEK
    - VEK(LARGE) ___ ___ ___ ___ ___ ___
    - NONVEK(GREEN) ___ ___ ___ ___ ___ ___

TASK 3:	R	R	W*	R	R*	R	R	R	R
TASK 4:	W	W	W*	R	R	R	R	R	R*
TASK 5:	R	R	W*	R	R	R	R	R	R*
TASK 6:	R	W*	R	R	R	R	R	R*	R
TASK 7:	W	R	W*	R	R*	R	R	R	R
TASK 8:	R	W*	R	R*	R	R	R	R	R
TASK 9:	W*	R	R	R	R	R	R*	R	R


VEK
VEK
NONVEK
VEK(BORDERED)
NONVEK(SQUARE)
VEK(SMALL)
NONVEK(RED)
NONVEK
VEK
NONVEK
VEK(GREEN)
NONVEK(LARGE)
VEK(CIRCLE)
NONVEK(NO BORDER)
VEK
VEK
NONVEK
VEK(RED)
NONVEK(SMALL)
VEK(NO BORDER)
NONVEK(CIRC)
NONVEK
VEK
NONVEK
VEK(SQUARE)
NONVEK(BORDERED)
VEK(LARGE)
NONVEK(GREEN)

TASK 12:	R	W	W*	R	R*	R	R	R	R
TASK 13:	W	R	W*	R	R	R	R	R	R*
TASK 14:	R	W	W*	R	R	R	R	R	R*
TASK 15:	W	W*	R	R	R	R	R	R*	R
TASK 16:	W	W	W*	R	R*	R	R	R	R
TASK 17:	W	W*	R	R*	R	R	R	R	R
TASK 18:	R*	R	R	R	R	R	R*	R	R

REMEMBER TO PUT RECORD GAP ON TAPE

APPENDIX C

APPENDIX C

PROTOCOL BOOKLET FOR EXPERIMENT 2

DATE ____________

TIME ____________

WARMUP 1:
R R W

VEK
- VEK
  - VEK
  - NONVEK
- NONVEK
  - VEK
  - NONVEK

NONVEK
- VEK
  - VEK
  - NONVEK
- NONVEK
  - VEK
  - NONVEK

(RED)
(CIRCLE)
(LARGE)
(NO BORDER)
(BORDERED)
(SMALL)
(SQUARE)
(GREEN)

HYPOTHESIS ____________

WARMUP 2:
W R R

VEK
- VEK
  - VEK
  - NONVEK
- NONVEK
  - VEK
  - NONVEK

NONVEK
- VEK
  - VEK
  - NONVEK
- NONVEK
  - VEK
  - NONVEK

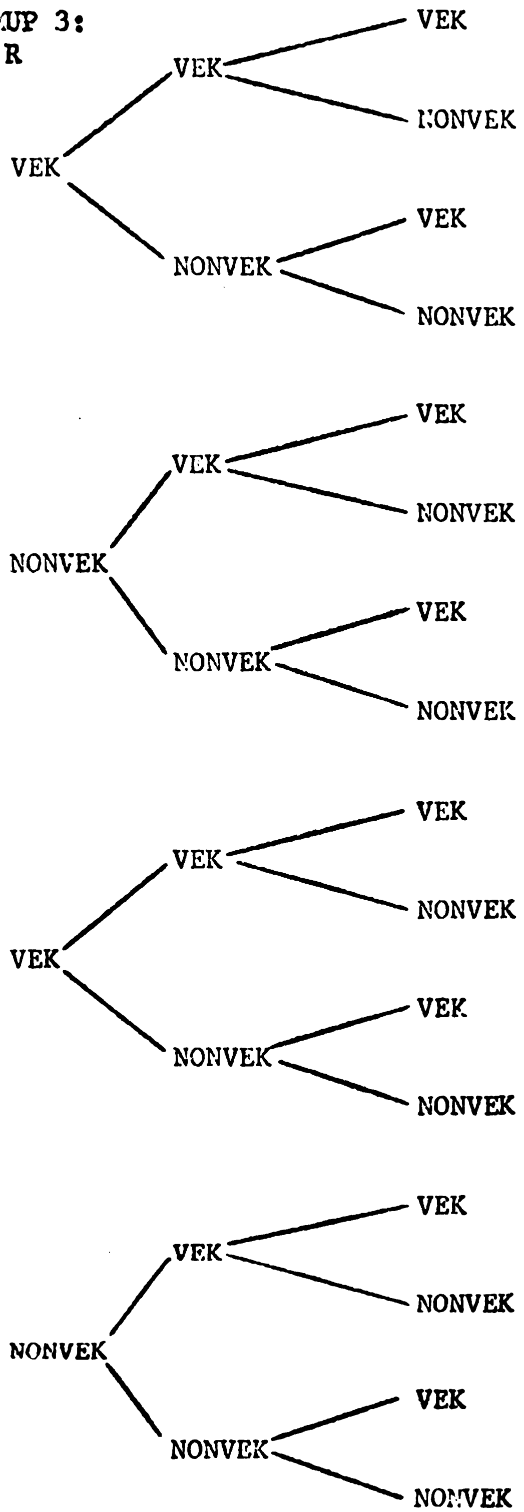
(NO BORDER)
(LARGE)
(RED)
(CIRCLE)
(SQUARE)
(GREEN)
(SMALL)
(BORDERED)

HYPOTHESIS ____________

DATE ____________________

TIME ____________________

WARMUP 3:
W W R



(SQUARE) ___ ___ ___ ___ ___ ___

(GREEN) ___ ___ ___ ___ ___ ___

(SMALL) ___ ___ ___ ___ ___ ___

(BORDERED) ___ ___ ___ ___ ___ ___ HYPOTHESIS

(NO BORDER) ___ ___ ___ ___ ___ ___ ____________

(LARGE) ___ ___ ___ ___ ___ ___

(RED) ___ ___ ___ ___ ___ ___

(CIRCLE) ___ ___ ___ ___ ___ ___

___ ___ ___ ___ ___ ___

___ ___ ___ ___ ___ ___

___ ___ ___ ___ ___ ___

___ ___ ___ ___ ___ ___ HYPOTHESIS

___ ___ ___ ___ ___ ___ ____________

___ ___ ___ ___ ___ ___

___ ___ ___ ___ ___ ___

___ ___ ___ ___ ___ ___

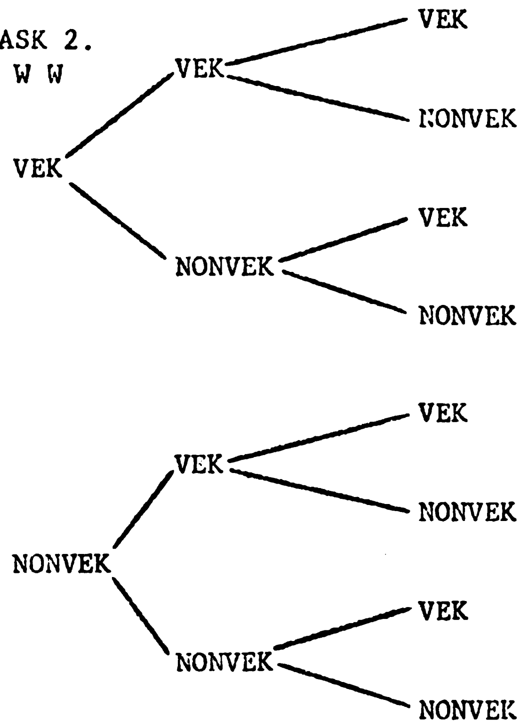
TASK 1:	___	___	___	___	___	___	___	___	___	HYPOTHESIS
	R	W	R	R	R	R	R	R	R	__________

GROUP 1

DATE ____________________

TIME ____________________

TASK 2.
W W W



(RED)

(LARGE)

(SQUARE)

(NO BORDER)

(BORDERED)

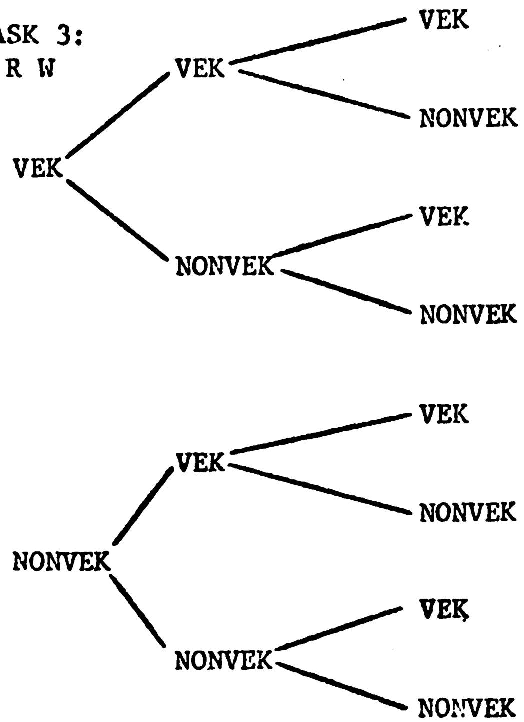
(CIRCLE)

(SMALL)

(GREEN)

HYPOTHESIS ____________

TASK 3:
R R W



(RED)

(BORDERED)

(LARGE)

(CIRCLE)

(SQUARE)

(SMALL)

(NO BORDER)

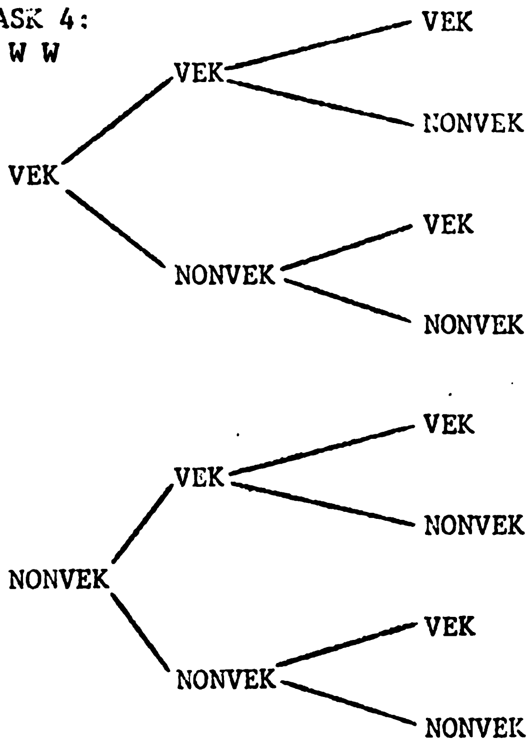
(GREEN)

HYPOTHESIS ____________

DATE ____________________

TIME ____________________

TASK 4:
R W W



(RED) ___ ___ ___ ___ ___ ___

(CIRCLE) ___ ___ ___ ___ ___ ___

(LARGE) ___ ___ ___ ___ ___ ___

(NO BORDER) ___ ___ ___ ___ ___ ___ HYPOTHESIS

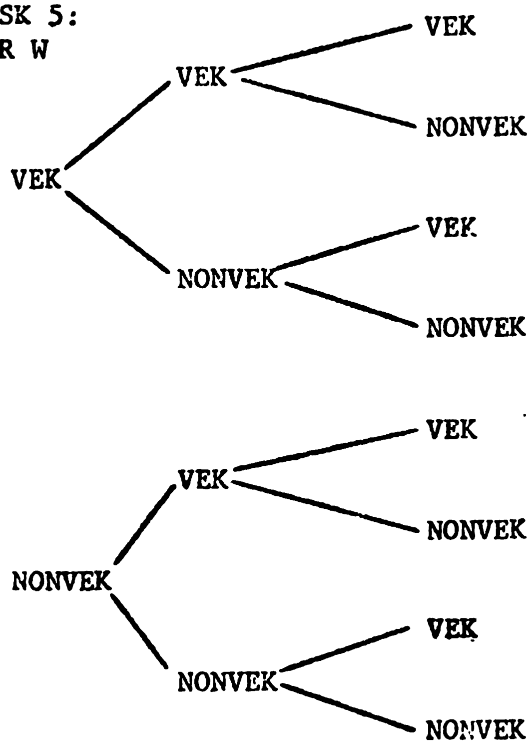
(BORDERED) ___ ___ ___ ___ ___ ___ __________

(SMALL) ___ ___ ___ ___ ___ ___

(SQUARE) ___ ___ ___ ___ ___ ___

(GREEN) ___ ___ ___ ___ ___ ___

TASK 5:
W R W



(GREEN) ___ ___ ___ ___ ___ ___

(SQUARE) ___ ___ ___ ___ ___ ___

(NO BORDER) ___ ___ ___ ___ ___ ___

(LARGE) ___ ___ ___ ___ ___ ___ HYPOTHESIS

(SMALL) ___ ___ ___ ___ ___ ___ __________

(BORDERED) ___ ___ ___ ___ ___ ___

(CIRCLE) ___ ___ ___ ___ ___ ___

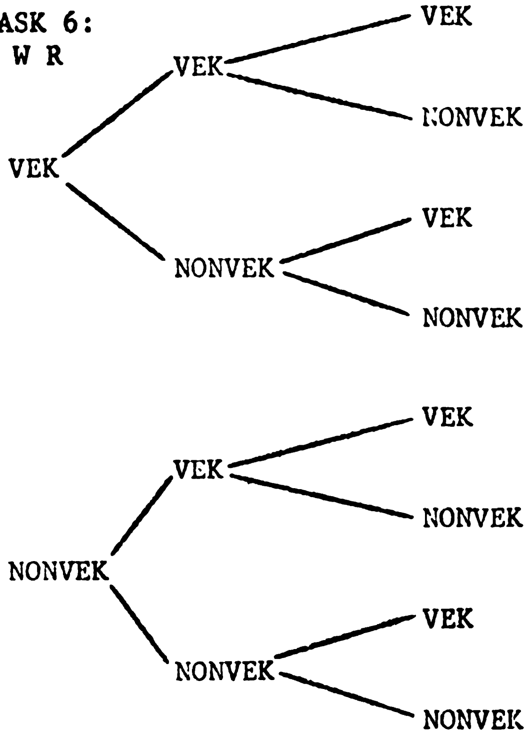
(RED) ___ ___ ___ ___ ___ ___

GROUP 1

DATE ____________

TIME ____________

TASK 6:
W W R



(NO BORDER)

(LARGE)

(RED)

(CIRCLE) HYPOTHESIS

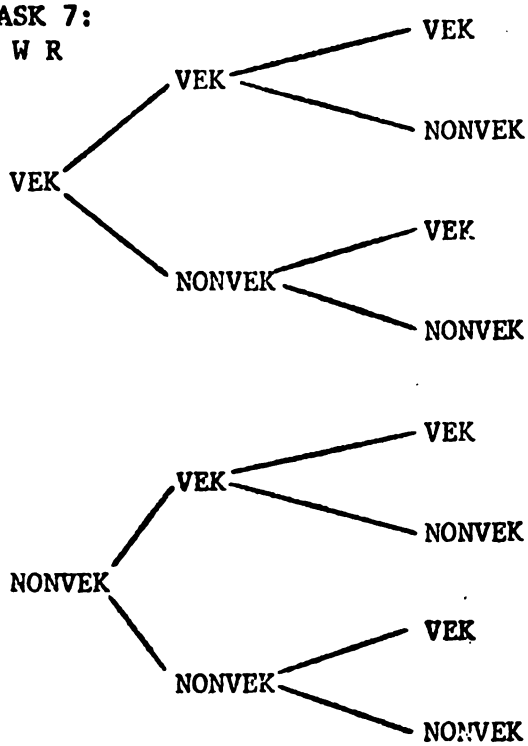
(SQUARE) ____________

(GREEN)

(SMALL)

(BORDERED)

TASK 7:
R W R



(LARGE)

(NO BORDER)

(GREEN)

(CIRCLE) HYPOTHESIS

(SQUARE) ____________

(RED)

(BORDERED)

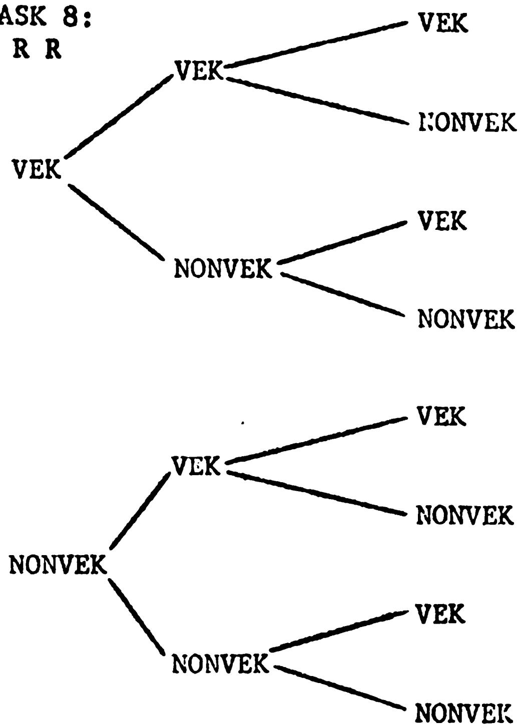
(SMALL)

GROUP 1

DATE ____________

TIME ____________

TASK 8:
R R R



(RED) ___ ___ ___ ___ ___ ___

(SMALL) ___ ___ ___ ___ ___ ___

(CIRCLE) ___ ___ ___ ___ ___ ___

(BORDER) ___ ___ ___ ___ ___ ___ HYPOTHESIS ____________

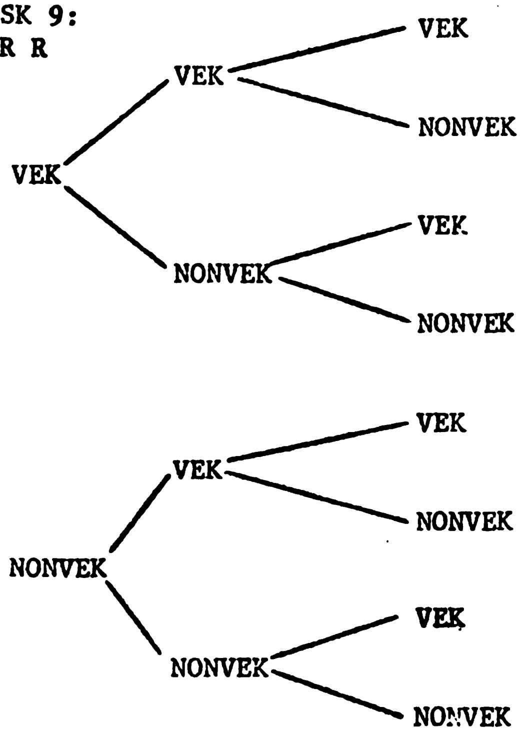
(BORDER) ___ ___ ___ ___ ___ ___

(SQUARE) ___ ___ ___ ___ ___ ___

(LARGE) ___ ___ ___ ___ ___ ___

(GREEN) ___ ___ ___ ___ ___ ___

TASK 9:
W R R



(RED) ___ ___ ___ ___ ___ ___

(SQUARE) ___ ___ ___ ___ ___ ___

(SMALL) ___ ___ ___ ___ ___ ___

(NO BORDER) ___ ___ ___ ___ ___ ___ HYPOTHESIS ____________

(BORDERED) ___ ___ ___ ___ ___ ___

(LARGE) ___ ___ ___ ___ ___ ___

(CIRCLE) ___ ___ ___ ___ ___ ___

(GREEN) ___ ___ ___ ___ ___ ___

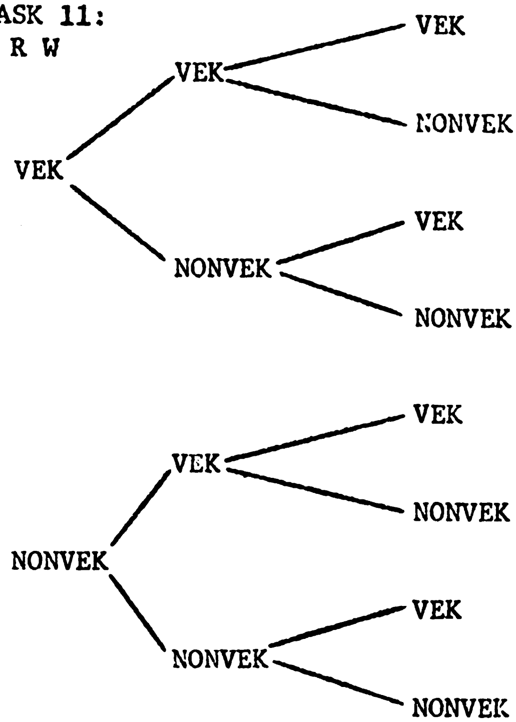
TASK 10:	___	___	___	___	___	___	___	___	___	HYPOTHESIS
	R	R	W	R	R	R	R	R	R	________

GROUP 1

DATE ____________________

TIME ____________________

TASK 11:
W R W



(SQUARE) ___ ___ ___ ___ ___ ___

(NO BORDER) ___ ___ ___ ___ ___ ___

(RED) ___ ___ ___ ___ ___ ___

(LARGE) ___ ___ ___ ___ ___ ___ HYPOTHESIS

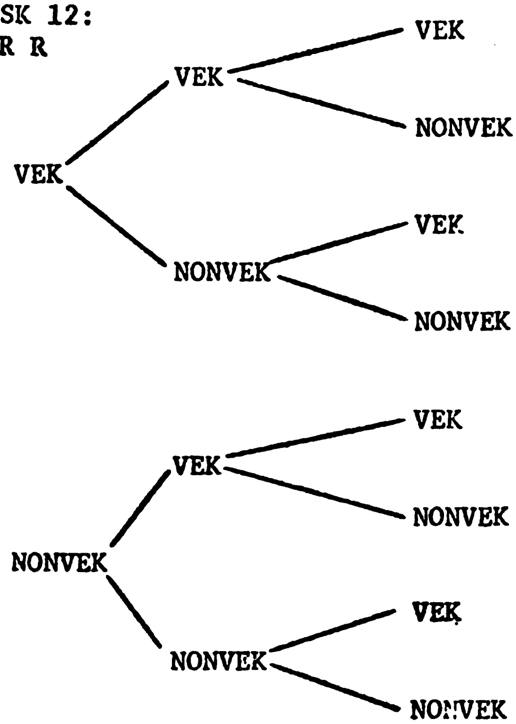
(SMALL) ___ ___ ___ ___ ___ ___ ____________

(GREEN) ___ ___ ___ ___ ___ ___

(BORDERED) ___ ___ ___ ___ ___ ___

(CIRCLE) ___ ___ ___ ___ ___ ___

TASK 12:
R R R



(BORDERED) ___ ___ ___ ___ ___ ___

(RED) ___ ___ ___ ___ ___ ___

(CIRCLE) ___ ___ ___ ___ ___ ___

(LARGE) ___ ___ ___ ___ ___ ___ HYPOTHESIS

(SMALL) ___ ___ ___ ___ ___ ___ ____________

(SQUARE) ___ ___ ___ ___ ___ ___

(GREEN) ___ ___ ___ ___ ___ ___

(NO BORDER) ___ ___ ___ ___ ___ ___

GROUP 1

DATE ____________

TIME ____________

TASK 13:
R W W

VEK	VEK	VEK	(RED)
		NONVEK	(CIRCLE)
	NONVEK	VEK	(LARGE)
		NONVEK	(NO BORDER)
NONVEK	VEK	VEK	(BORDERED)
		NONVEK	(SMALL)
	NONVEK	VEK	(SQUARE)
		NONVEK	(GREEN)

HYPOTHESIS ____________

TASK 14:
W W W

VEK	VEK	VEK	(NO BORDER)
		NONVEK	(LARGE)
	NONVEK	VEK	(GREEN)
		NONVEK	(SQUARE)
NONVEK	VEK	VEK	(CIRCLE)
		NONVEK	(RED)
	NONVEK	VEK	(SMALL)
		NONVEK	(BORDERED)

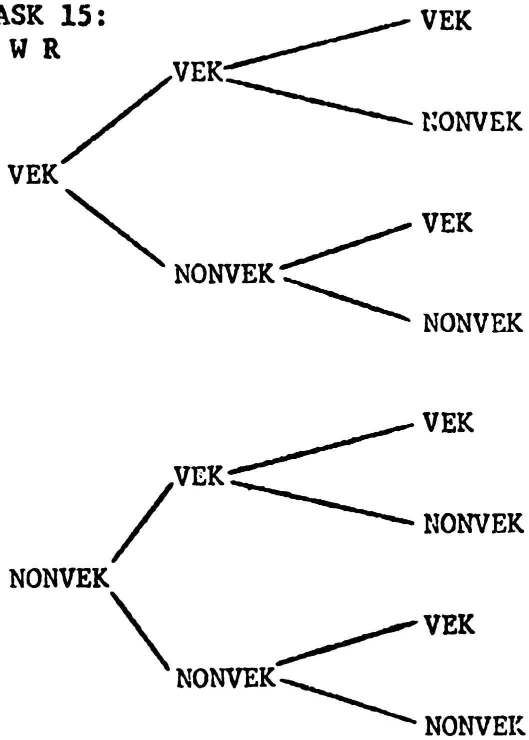
HYPOTHESIS ____________

GROUP 1

DATE ______________

TIME ______________

TASK 15:
R W R



(SQUARE) ___ ___ ___ ___ ___ ___

(GREEN) ___ ___ ___ ___ ___ ___

(SMALL) ___ ___ ___ ___ ___ ___

(BORDERED) ___ ___ ___ ___ ___ ___ HYPOTHESIS

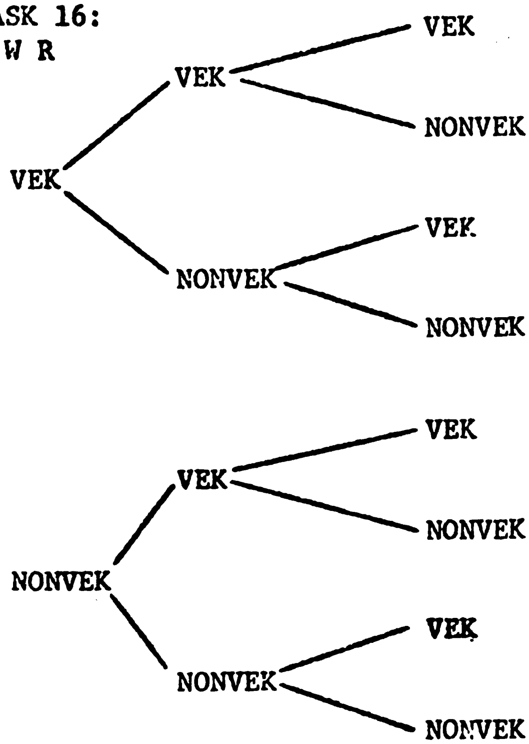
(NO BORDER) ___ ___ ___ ___ ___ ___ ______________

(LARGE) ___ ___ ___ ___ ___ ___

(RED) ___ ___ ___ ___ ___ ___

(CIRCLE) ___ ___ ___ ___ ___ ___

TASK 16:
W W R



(SQUARE) ___ ___ ___ ___ ___ ___

(RED) ___ ___ ___ ___ ___ ___

(BORDERED) ___ ___ ___ ___ ___ ___

(SMALL) ___ ___ ___ ___ ___ ___ HYPOTHESIS

(LARGE) ___ ___ ___ ___ ___ ___ ______________

(NO BORDER) ___ ___ ___ ___ ___ ___

(GREEN) ___ ___ ___ ___ ___ ___

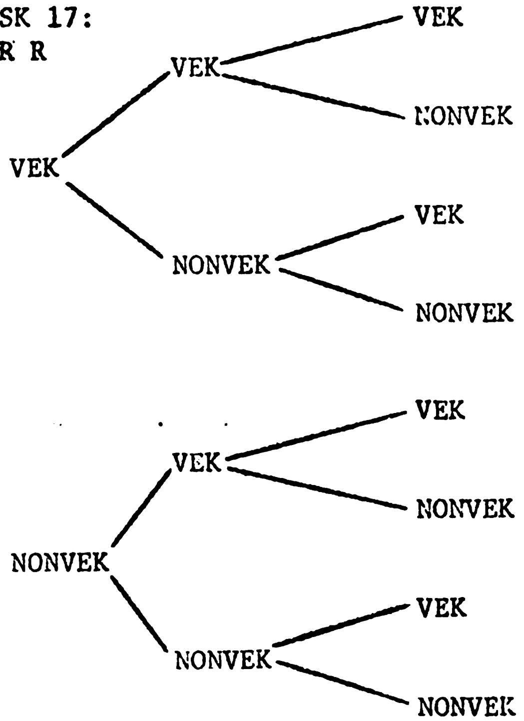
(CIRCLE) ___ ___ ___ ___ ___ ___

GROUP 1

DATE ____________

TIME ____________

TASK 17:
W R R



(BORDER) ___ ___ ___ ___ ___ ___

(SQUARE) ___ ___ ___ ___ ___ ___

(LARGE) ___ ___ ___ ___ ___ ___

(GREEN) ___ ___ ___ ___ ___ ___ HYPOTHESIS ________

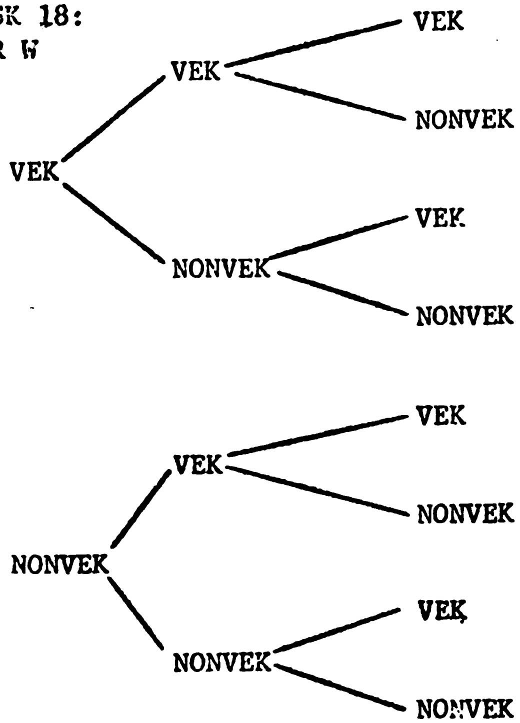
(RED) ___ ___ ___ ___ ___ ___

(SMALL) ___ ___ ___ ___ ___ ___

(CIRCLE) ___ ___ ___ ___ ___ ___

(BORDER) ___ ___ ___ ___ ___ ___

TASK 18:
R R W



(LARGE) ___ ___ ___ ___ ___ ___

(BORDERED) ___ ___ ___ ___ ___ ___

(GREEN) ___ ___ ___ ___ ___ ___

(CIRCLE) ___ ___ ___ ___ ___ ___ HYPOTHESIS ________

(SQUARE) ___ ___ ___ ___ ___ ___

(RED) ___ ___ ___ ___ ___ ___

(NO BORDER) ___ ___ ___ ___ ___ ___

(SMALL) ___ ___ ___ ___ ___ ___