



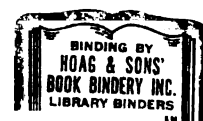
144  
672  
THS

HYPOTHESIS SAMPLING AND  
INFORMATION PROCESSING IN  
CONCEPT IDENTIFICATION

Thesis for the Degree of M. A.  
MICHIGAN STATE UNIVERSITY  
DAVID JOHN DePALMA  
1971



3 1293 10382 9085



## ABSTRACT

### HYPOTHESIS SAMPLING AND INFORMATION PROCESSING IN CONCEPT IDENTIFICATION

By

David John DePalma

Many researchers have replicated the "outcome effect" in experiments in concept identification (Richter, 1965; Levine, 1966; Kornreich, 1968; DePalma, 1969; and others). Some of the methodological problems, however, have received little attention. DePalma tried to avoid some of these problems by using a modified version of Richter's design. The approach proved successful, but it too had some deficiencies. The present study extends this earlier investigation, and examines the effects of memory aid and frequency of experimental question on subjects' performance. The relationship between problem type, sequence of outcomes and performance is also investigated.

Subjects were asked the question, "How did you make that choice?" either after each trial or only once per four-trial problem. Half of each of these groups were allowed to use a "memory" aid, paper and pencil, while half were not. All subjects were given sixteen four-trial, four-dimensional problems, with each dimension (color, letter,

size or position) correct (relevant) an equal number of times. The variables of question frequency and memory aid were controls in this study, since it was expected that neither one would have differential effects on performance.

It was predicted that: 1) Problem type and sequence of outcomes would be influential factors on performance; 2) Changes of hypothesis would occur after rights as well as after wrongs; 3) A new hypothesis would not always be consistent with the information the subject received on an error trial; 4) The subject would consider as hypotheses stimuli which had failed to pass a consistency check; 5) Hypothesis-sampling would occur with replacement; 6) The outcome effect would be replicated; 7) No one particular strategy (Win-Stay, Lose-Shift) would be used more frequently than any other, a) subjects would respond on the basis of one hypothesis while processing one or more hypotheses, b) a more important factor (than strategy) in processing would be how subjects used the information they received, especially "wrong" information.

Ninety-six undergraduates were individually tested with the experimenter giving outcomes, asking the experimental question, and recording the subjects' responses.

Analysis of question frequency, memory aid and problem type showed a significant effect of problem type. No other effects were significant. Size problems resulted in the lowest level of performance. This result replicates

DePalma's earlier work, but requires further research for an explanation. In another analysis of variance, the main effects of sequence of outcome and problem type were significant. Of these, sequence proved to be more influential. The other predictions were confirmed.

From these data, it was observed that some subjects processed and recorded "wrong" information with "right" information. Other subjects only utilized a portion of the available input. The latter method of processing caused many difficulties, and usually did not lead to solution-attainment by these subjects. It was concluded that subjects who can use correct ("right") and incorrect ("wrong") information effectively, will solve the problems despite the experimental conditions employed. However, subjects who have trouble processing information (especially "wrong") may be affected positively or negatively by the same methodology. The hypothesis-sampling and processing of such subjects should be more closely investigated in future studies.

APPROVED: Ellen A. Strommen

DATE: 3 May 1971

HYPOTHESIS SAMPLING AND INFORMATION  
PROCESSING IN CONCEPT IDENTIFICATION

By

David John DePalma

A THESIS

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

MASTER OF ARTS

Department of Psychology

1971

To my parents, who have made all this possible.

## ACKNOWLEDGMENTS

I would like to express my gratitude to the members of my committee: Dr. Ellen Strommen, Dr. William Stellwagen, Dr. Gordon Wood and Dr. Donald Johnson for all their assistance in planning this research. Dr. Strommen particularly deserves my appreciation for the time, energy and encouragement she offered me during the entire project.

I am especially indebted to Dr. Martin Richter, Lehigh University, who introduced me to this research area, and without whose generous assistance this project would not have been realized.

The students who served as subjects should be thanked for their time and efforts. I would also like to thank Charlotte Wright for her help with the data, and her advice and enthusiasm during the study.



## TABLE OF CONTENTS

	Page
LIST OF TABLES . . . . .	v
INTRODUCTION . . . . .	1
METHOD . . . . .	12
RESULTS . . . . .	20
DISCUSSION . . . . .	31
REFERENCES . . . . .	37
APPENDIX	
A EXPERIMENTAL DESIGN . . . . .	40
B QUESTION MATRIX . . . . .	41

## LIST OF TABLES

Table		Page
1.	MEAN NUMBER OF CORRECT PROBLEMS AND VARIANCE FOR EACH PROBLEM TYPE BY GROUP CELL . . . . .	20
2.	SUMMARY OF ANALYSIS OF VARIANCE OF QUESTION X MEMORY AID X PROBLEM TYPE . . . . .	21
3.	PERCENTAGES OF CORRECT-CHOICE RESPONSE ON TRIAL FOUR FOR PROBLEM TYPE AND SEQUENCE OF OUTCOME . .	23
4.	SUMMARY OF ANALYSIS OF VARIANCE OF PROBLEM TYPE X SEQUENCE OF OUTCOME . . . . .	25
5.	PERCENTAGES OF CORRECT-CHOICE RESPONSE ON TRIAL FOUR FOR QUESTION AND MEMORY GROUPS BY SEQUENCE OF OUTCOME . . . . .	26
6.	FREQUENCY OF SUBJECTS BY NUMBER OF CORRECT PROBLEMS . . . . .	27
7.	NUMBER OF SUBJECTS GIVING CORRECT-CHOICE RESPONSE ON TRIAL FOUR FOR EACH PROBLEM AND PROBLEM TYPE . . . . .	29
8.	PERCENTAGES OF CORRECTNESS AND USE OF HYPOTHESIS-SAMPLING STRATEGIES FOR PROBLEM TYPES . . . . .	30

## INTRODUCTION

Many theories of hypothesis sampling and information processing in concept identification have been proposed (Levine, 1966; Bower and Trabasso, 1964; Rogers and Haygood, 1968; and others). However, many of the assumptions of these theories appear questionable in light of recent research.

One of the problems with these theories is that they usually describe only good solvers, and the situation where subjects use information correctly. The shortcomings of these descriptions are twofold. First, few (if any) generalizations can be made in the application of these assumptions to the processing of poor solvers. And secondly, DePalma (1969) has shown that even the assumptions regarding good solvers are not completely accurate.

Some of the other problems with these theories involve the methodologies used to support them. These procedures have in some way been inappropriate, too highly structured or unsatisfactory to examine the broad range of questions involving the theories.

The present study was designed to test the assumptions concerning hypothesis sampling, avoid some of the methodological problems of past research, and investigate

factors involved in subjects' processing of the information they received from the outcomes.

The conceptualization in human discrimination learning of subjects as information processors and analyzers is relatively recent in psychology. Researchers in concept identification characteristically utilize computer terminology for such description.

Levine (1966) proposed a theory (similar to Bruner, Goodnow and Austin's (1956) "focusing" strategy) in which he assumed that subjects remember (encode) all the logically correct cues after an outcome ("right" or "wrong"), store these "hypotheses," and then test these hypotheses on subsequent trials. This allows the subject to eliminate hypotheses until the one correct solution remains.

To test this theory, Levine proposed a method whereby the set of possible hypotheses was determined by the experimental situation and the one hypothesis which the subject was "holding" after each trial could be inferred. The outcomes, "right" or "wrong," were controlled so that the effects of outcome on retention or rejection of a hypothesis held could be analyzed. To obtain the necessary information, Levine devised the "blank" trial method, in which four blank (no outcome) trials are presented between the outcome trials. If the subject responded on the basis of a single hypothesis, that hypothesis manifested itself in a distinguishable sequence over the four blank trials.

From his studies, Levine found that subjects respond on the basis of a hypothesis until a "wrong" outcome is received, at which time they shift to another hypothesis. These data yielded evidence that the subjects hold several hypotheses at one time and eliminate several simultaneously. This led Levine to the formulation of his "focusing" strategy, and in a recent study (1969) he proposes:

- a) the subject samples a subset of hypotheses,
- b) then the subject takes one, a working hypothesis, from this subset as the basis for his response,
- c) the subject uses the outcome "right" or "wrong," to evaluate those hypotheses in the subset.

The emphasis here is on the subject's monitoring of more than one hypothesis at a time even though he uses only one hypothesis as the basis for his response.

Another hypothesis-testing model of concept identification is the Bower and Trabasso (1964) theory. The basic assumptions of this theory postulate:

- 1) a change of hypothesis occurs only after a "wrong" outcome, which infirms the hypothesis on which that response was based.
- 2) a new hypothesis is always consistent with information (about stimulus and response assignment) given on an error trial.
- 3) stimulus dimensions failing to pass consistency check are not considered possible hypotheses during the selection process.

Although these two theories have been the most noteworthy in the literature, a recent probabilistic model by Rogers and Haygood (1968) attempts to explain hypothesis-testing in concept identification as a process in which:

- 1) the subject discovers a working hypothesis by changing his hypothesis frequently until he is "right" more than 50% of the time.
- 2) after discovering a hypothesis which works better than chance, the subject adds amendments, until the solution hypothesis (minus irrelevant hypotheses) has been obtained.
- 3) the subject no longer changes his hypothesis, and continues to respond on the basis of this hypothesis (see also Falmagne, 1970).

There have been many studies criticizing the assumptions of these theories. Bower and Trabasso's contention that after an error trial, the subject resamples with replacement from the "hypothesis-pool," has been questioned. Restle (1962) provided the original proposition for such sampling, then Bower and Trabasso (1964, 1966) presented supporting evidence. Most recently, Merryman, Kaufmann, Brown and Dames (1968) concluded that the sampling-with-replacement theory could not be rejected. However, not all of the available evidence supports such a conclusion. Levine (1966), Erickson (1968) and Nahinsky and Slaymaker (1969) have obtained strong support for the contrary

proposal that after an error, sampling cannot occur with replacement, but instead occurs without replacement. One aim of the present study was to obtain further evidence relevant to this issue.

Another area of some dispute has been the effects of "right" and/or "wrong" on the information-processing of the subject. Rogers and Haygood (1968) found that for a block of errorless trials, the subject is just as likely to change his hypothesis as he is to keep it; and with at least one error, the subject is as likely to keep his hypothesis as he is to change it. The authors point out that the subjects could have changed hypotheses because of implication; that is, because of the experimental procedure. Unfortunately, Rogers and Haygood seem to dismiss this possibility as easily as they proposed it. They also found that subjects who take longer to respond make more errors, and change hypotheses more often than low-latency subjects. Merryman, Kaufmann, Brown and Dames (1968) found that after six non-contingent trials of either "right" or "wrong," the "wrong"s had no effects on performance, while the "right"s produced a retarding effect on subsequent learning. From their data, they also decided to reject the idea that the subject keeps his hypothesis after a correct trial. Similarly, Nahinsky and Slaymaker (1969) and Dodd and Bourne (1969) found evidence that subjects change hypotheses not only after an error trial, but also after a

correct trial. However, not all the evidence substantiates these conclusions. Bourne, Dodd, Guy and Justesen (1968) observed earlier that although learning occurs on all trials, changes occur only after an error. Levine (1966) and Bower and Trabasso (1966) concur on this point, as mentioned earlier. And more recently, Trabasso and Staudenmayer (1968) have obtained data which indicate that random reinforcement effects, that is, non-contingent feedback, are problem or dimension specific, especially if the subject is familiar with (knows) the stimulus dimensions. In the present study these random reinforcement effects were avoided by using contingent outcomes. It was hoped that this procedure would provide more relevant (and more accurate) information regarding subjects' processing than non-contingent feedback.

Some other problems which have received relatively little attention are concerned with the methodology of the experiments themselves. In Levine's "blank" trial method, only one hypothesis is tapped on a given trial, although it has been shown that subjects hold several hypotheses simultaneously. Levine himself realized this, but he continues to use this method. This conflict between experimental procedure and the "observed" mode of information-processing is an important shortcoming of the methodology, not only with regard to the sampling of hypotheses, but also with respect to the effects of the outcomes.



Kornreich (1968) used two procedures to circumvent these trouble areas. In the first, he used a modified "blank" trial procedure and preprogrammed the outcomes. In the second phase of the study, the outcomes were dependent on the subject's responses. During the experiment, the subjects were faced with eight buttons on which all the possible hypotheses were written. Subjects were asked to indicate which hypotheses still could be correct after each of the outcomes. This procedure supposedly taps all the hypotheses held by the subject. Another group of subjects was run under Levine's "blank" trial method. No significant differences were found between procedures in effect on correct processing or selecting. However, there may not have been any differences because of the highly structured cue aid (the eight buttons). The present study examined this problem more closely by providing some subjects with a completely unstructured cue aid, paper and pencil. These subjects were able to use the information they received on each trial in a manner more consistent with their own mode of processing (without the influence of the eight buttons).

Another methodological problem occurs in experiments such as that of Merryman, Kaufmann, Brown and Dames (1968) in which a group of non-contingent "right"s or "wrong"s is presented to the subject. It does not seem reasonable that the same mode of processing operates in this setting and in

the situation where the subject receives contingent "right"s or "wrong"s (or a mixture of contingent outcomes). The effort to make the outcomes non-contingent fails in the former procedure, because the subject begins to use more information than he would under a contingent paradigm. That is, the subject notices that no matter what he says the outcome is the same. So he experiments with many possibilities, changing hypotheses frequently--probably more frequently than he would under the contingent situation (at least for "right"s). Thus, the changing of hypotheses and the retarding effects of the group of errorless trials are artifacts of the procedure, not true indications of "what is happening." Certainly, such random reinforcement results cannot provide prototypes for successful hypothesis-sampling theories in the mixed outcome condition.

In an effort to avoid these problems, and investigate some other aspects of the concept identification task, DePalma (1969) used the four-trial, four-dimensional discrimination problems of Richter (1965) with one important modification. A question designed to tap the hypotheses held by the subject (but without the cue aids of Kornreich) was asked by the experimenter once during each problem. The question was purposely very vague--"How did you make that choice?"--and was asked only once per problem to keep interference with the subject's processing at a minimum.

Otherwise, the problems might have become question-answering tasks. The question was asked after the outcome, because prior questioning might have shaped or interfered with the subject's processing. It was hoped that asking the question once per problem would not have a detrimental effect on the subject's processing ability.

The data indicated that the experimental question interfered no more than Kornreich's procedure had, if it interfered at all. The effect of wrongs on performance as observed by Richter (1965), Levine (1966) and Kornreich (1968) was replicated. That is, the probability of correct-choice response on trial four decreased as the number of errors on the first three trials (from 0 to 3) increased. This has been labelled the "outcome effect." But further analysis of the data suggested that the effect was much more complex. The sequence of rights and wrongs and the problem type (color, letter, size and position) appeared to be related somehow to the probability of correct-choice response on trial four. This differed from the outcome effect because sequences which had the same number of wrongs on the first three trials had different probabilities of problem solution. For two of the problem types, o+o (wrong, right, wrong) was more detrimental than ooo (wrong, wrong, wrong)!

These results indicate the importance of sequential effects of the outcomes. The traditional outcome effect

explanation fails to account for this. What seems so simple at first glance appears so only because of averaging of results--when each sequence is studied separately, the complexity is revealed.

The responses given by the subjects contingent upon the outcome on the previous trial were also analyzed. It is interesting to note that being incorrect on the previous trial resulted in a fairly constant level (probability) of being correct on trial four of approximately 67%. Being correct, however, on the previous trial led to much higher percentages, averaging around 84%. It seems from these data that one of the effects of error on the previous trial is to make some subjects use more elements (hypotheses) at a time when they should be narrowing down the choices, not expanding them.

Of course, being correct on the first three trials increased the probability of the subject's being correct on trial four. The subjects who responded correctly on a given trial outperformed subjects who responded incorrectly on corresponding trials with regard to final solution attainment for every trial. Thus, this experiment supported Levine's contention that "wrong"s affect problem-solving differently. However, the subjects did not code (attempt to remember) the stimulus cues as Levine says, before the outcome, but after it. Kornreich stated that after the outcome the subject simply encodes the correct stimulus

cues. This corresponds to the Bower and Trabasso theory mentioned earlier. Logically, subjects should encode only the correct cue information, but DePalma observed that subjects will, in fact, encode "wrong" cue information instead of the correct stimulus cues. This ultimately led to incorrect response choice on trial four. The subject "knew" how to solve the problem, since he had solved some correctly, but sometimes he used the "wrong" information. The interference was not in the "focusing" strategy employed, but in the coding--either incorrect information, or the non-utilization of all the available information. It also seems that subjects encode all the hypotheses or stimulus cues, but decide on the basis of only one. This result has been confirmed recently by Levine (1969). Contrary to Levine's theory (and others), it was observed that sampling occurred with replacement, since hypotheses were frequently repeated during a problem.

Thus, the problem type and sequence of outcomes play an important part in the subject's processing and performance. However, it is possible that the subject's performance might be affected by the availability of "memory" aid, or by the frequency of the probes (experimental questions). The present study extends DePalma's (1969) study by examining the effects of memory aid and frequency of probes on performance. The relationship between problem type, sequence of outcomes and performance is investigated more closely.

## METHOD

Studies such as Kornreich's (with their cue aids) were criticized for structuring the subject's responses so that the observed data are not true "tapping"s of the subject's hypotheses. In the present study we hope to remedy this by using a less structured methodology, allowing subjects to use paper and pencil while working the problems. This unstructured cue aid will permit the subject to use all or a portion of the information he receives from the outcomes according to his own method of processing. If memory for previous responses is an important component in problem solution, subjects using pencil and paper should perform differently from subjects who have no cue aid. One group of subjects will be allowed to use paper and pencil to help them, while the other group will not.

The main purpose of the present study was to examine problem type and sequence of outcome in greater detail. From DePalma's (1969) data and the considerations reviewed above, we predict:

Hypothesis one: Problem type and sequence of outcome will be important factors for correct-choice response on trial four.

Hypothesis two: Changes of hypothesis will occur after "right"s as well as after "wrong"s.

Hypothesis three: A new hypothesis will not always be consistent with the information the subject receives on an error trial.

Hypothesis four: Stimuli which fail to pass consistency check (as determined by the experimenter) will be considered possible hypotheses during the selection process. This hypothesis and the previous one could be combined by stating that subjects will not always use hypotheses which are consistent with (logically follow) information they receive. Hypotheses two through four are in disagreement with the Bower and Trabasso theory mentioned earlier. We will agree, however, that:

Hypothesis five: Hypothesis-sampling will occur with replacement.

Hypothesis six: As the number of "wrong"s on the first three trials increases from 0 to 3, the probability of correct-choice response on trial four will decrease. This will not be a simple relationship, however, if DePalma's sequential effects are replicated.

Hypothesis seven: No one particular strategy of processing (Win-Stay, Lose-Shift, for example) will be used more frequently than any other, a) subjects will respond on the basis of one hypothesis while processing one or more,

b) a much more important concern will be how subjects use the information they receive, especially "wrong" information.

In DePalma's study it was observed that asking the experimental question once per problem was not detrimental to the subject's performance. And, it was assumed that asking the question after each trial would change the nature of the task. However, this assumption was not tested. In the present study both conditions are used, to see whether frequency of question influences performance on the task.

### Subjects

Ninety-six undergraduates (76 females, 20 males) enrolled in an introductory psychology course at Michigan State University served as subjects.

### Stimulus Cards and Problems

The discrimination problems consisted of sets of cards on which were drawn two stimuli about 1-1/2 inches apart. The stimuli varied on four dimensions--color, letter, size and position. The colors and letters differed for each problem. Large letters were 1-1/2 inches, small letters one inch in height.

A problem was composed of four such cards, and the outcome, "right" or "wrong," was given after the subject's



response to the card. The four cards formed a set with several properties. Each value of each dimension was combined exactly twice with the values of all the other dimensions. The set provided that, after the first outcome, four of the eight cues remained as logically possible solutions; after the second outcome, two remained; and after the third outcome, the solution was logically determined. This was true whether the outcomes were "right" or "wrong." Also, the subject had a 50% chance of choosing the correct stimulus on the first three cards.

Using the three cards (trials) #2, 3, and 4, it was possible to construct (for each problem) three combinations of the cards so that each card type was present on each trial over the three combinations. That is, these combinations were possible: 1,2,3,4; 1,3,4,2; and 1,4,2,3. This balanced for sequential effects across subjects and enabled the experimenter to make inferences that were not problem specific. The three combinations were labelled A, B, C.

### Design and Procedure

The design was a simple 2x2x4 question x memory aid x problem type factorial design with repeated measures on the last factor. The two question conditions were: 1) question after each trial, and 2) question once per problem. This condition existed to test the effects

(facilitative or detrimental) of the experimental question. The two memory aid conditions were: 1) the paper and pencil group, and 2) no paper and pencil. These groups tested the effects of an unstructured cue aid on the subject's processing. Subjects in each of these groups were given color, letter, size and position problems (see Appendix A). These conditions provided that there be two analyses of variance for the data. The first analyzed the effects of question, memory aid and problem type on performance (hypothesis given on trial four). And the second analysis examined the effects of sequence of outcome (on the first three trials) and problem type on performance.

The question--"How did you make that choice?"--was asked in the question-once-per-problem condition according to a schedule determined by a 16x16 matrix of trials vs problem type (see Appendix B). This matrix provided for the experimental question to be asked after each trial across all subjects (N.B. not for each subject). Of course, no matrix was needed in the question-after-each-trial condition. The deck types A, B, or C were assigned to the subject as he entered the experimental situation in the order--(A,B,C,A,B,C...). Thus, deck types and question trial were controlled across subjects.

Each subject was instructed and tested individually. The instructions were nearly identical to those used by Richter (1965). The difference was that the subject was

given four practice problems--one of each problem type--to the criterion of correct solution. After practice, all subjects were given one more problem (which did not count in the experiment) and then the sixteen test problems. It should be noted here that no two problem types followed one another more than twice over the sixteen problems. The subjects turned the cards over at their own speed. The subjects' responses to the experimental question were recorded, as were the outcomes given by the experimenter.

### Instructions

The instructions as given to the subject were as follows:

"This is an experiment in problem-solving. We want to see how quickly you can solve some very simple problems."

"I will show you a card like this (show cX). Each card will have two different letters side by side, each of a different color and different size. Each problem will consist of a series of cards with different combinations of the two letters, two colors, two sizes and two positions, (left and right)--like this...(show first three or four cards of cX)."

"For each card I want you to point with this pen to the one you think is correct, either this one or that one (demonstrate). Hold the pen in that position until I tell you whether you are right or wrong. Then you may turn to

the next card and again point to the one you think is correct. After you have turned over a card you may not turn back to it. The idea, of course, is for you to try to get as many right as possible. (The paper-and-pencil subjects were instructed that they could use paper and pencil provided by the experimenter in the task)."

"In all these problems the solution is of the simplest kind; either the same letter, the same color, the same size or the same position will be correct throughout a single problem. In order to give you a better idea of the procedure and the kind of problems you will have, let's begin with the first practice problem. Are there any questions before we begin?"

After practice:

"(As you can see yourself, you were getting them all right toward the end).\* On all the cards of this problem, one element (color, letter, size or position) was always correct. All the other problems you will have will also have solutions as simple as these. For some problems the large letter (or the small) will always be correct; for some the one on the right (or left) will always be

---

\*The bracketed material was used only if the subject had actually gotten at least the last three trials correct. For those who did not, only the second half of the sentence was read, followed by a demonstration by the experimenter of the correct responses on the last four trials of the problem.

correct. Sometimes it will be one of the colors, and sometimes it will be one of the letters itself that will always be correct."

"However, the problems will be much shorter than the practice problems; there will only be four cards in each problem, while there were twelve in the practice problems. Thus, although the solutions are simple, you must solve the problems very fast in order to get as many right as possible. Also remember that once you have turned over a card you may not refer back to it again. Are there any questions?"

## RESULTS

Since each subject was given four of each problem type, he could have had from 0 to 4 correct-choice responses on trial four for each type. The mean number of correct problems and the variance for each problem type and condition appear in Table 1.

TABLE 1  
MEAN NUMBER OF CORRECT PROBLEMS AND VARIANCE  
FOR EACH PROBLEM TYPE BY GROUP CELL

				Problem Type			
				Color	Letter	Size	Position
Q	NP	Mean	=	3.17	3.21	2.79	3.08
		Variance	=	.58	.17	.65	.69
	PP	Mean	=	3.17	2.96	2.88	3.04
		Variance	=	1.36	.99	.90	.74
QAE	NP	Mean	=	3.21	3.04	2.75	3.29
		Variance	=	1.04	.56	.84	.56
	PP	Mean	=	3.42	3.17	2.92	3.29
		Variance	=	.43	1.28	1.12	.39

Q = question once per problem  
QAE = question after each trial  
NP = no paper and pencil  
PP = paper and pencil

The solution data were analyzed by a three-way question x memory aid x problem type analysis of variance with repeated measures over problem type (see Table 2). This analysis showed a significant main effect of problem

TABLE 2  
SUMMARY OF ANALYSIS OF VARIANCE OF  
QUESTION X MEMORY AID X PROBLEM TYPE

Source of Variance	Sums of Squares	DF	F	P
<u>Between</u>				
A Question no.	0.94	1	.86	NS
B Memory aid	0.31	1	.28	NS
AB	0.75	1	.69	NS
<u>S</u> within groups (error)	100.56	92		
<u>Within</u>				
C Problem type	11.34	3	5.73	<.01
AC	0.84	3	.42	NS
BC	1.09	3	.55	NS
ABC	0.41	3	.21	NS
C x <u>Ss</u> within (error)	182.57	276		
Total	298.81	383		

type ( $F = 5.73$ ,  $df = 3/276$ ,  $p < .01$ ). None of the other main effects or interaction was significant. The question frequency and presence of a "memory" aid had no effect on subjects' performance. Evidently asking the question after each trial does not affect performance on the task, and factors other than memory are critical for correct solution. However, the paper and pencil group provided valuable information regarding subjects' processing. This will be discussed later.

Individual comparisons (Winer, pp. 65-69) showed that size problems differed significantly ( $F = 6.64$ ,  $df = 1/276$ ,  $p < .01$ --for the smallest difference) from the other problem types. That is, there were significantly fewer correct-choice responses on trial four (solutions) for the size problems than for any other type. This partially confirmed the hypothesis that problem type would be an influential factor of performance, and replicated DePalma's earlier results. Size problems again led to the lowest level of performance, while letter, position and color problems (in order of increasing performance) resulted in significantly better scores (see Table 7). Although this result was expected, it is unexplainable at this time.

In order to evaluate the effects of sequence of outcome on solution attainment, and test hypotheses one through seven, the data were reorganized.



The percentage of solution attainment on trial four as a function of problem type and sequence of outcome are shown in Table 3. These percentages represent the proportion of the time a particular sequence of outcome and problem type resulted in correct-choice response on trial four.

TABLE 3

PERCENTAGES OF CORRECT-CHOICE RESPONSE ON TRIAL FOUR  
FOR PROBLEM TYPE AND SEQUENCE OF OUTCOME

Sequence	Problem Type			
	Color	Letter	Size	Position
ooo	61	64	53	71
+oo	71	70	53	77
o+o	69	58	61	63
oo+	77	85	80	70
++o	86	89	84	90
+o+	88	78	78	84
o++	84	86	67	74
+++	100	98	100	100

Since the outcomes in this study were not preprogrammed but were contingent upon the subjects' choices on each trial, the number of subjects in each sequence of outcome by problem type cell (see Table 3) could not be

controlled. Therefore, there is unequal subject representation in the data. Despite the unequal subject representation in each cell (and sometimes repeated representation), the sequence of outcome data were analyzed by a two-way problem type x sequence of outcome analysis of variance (see Table 4). A harmonic mean of 46.51 was used according to Winer (pp. 242-243). This analysis yielded significant main effects of problem type ( $F = 2.94$ ,  $df = 3/1504$ ,  $p < .05$ ) and of sequence of outcome ( $F = 19.13$ ,  $df = 7/1504$ ,  $p < .01$ ). The interaction was not significant. Using F-tests (Winer, p. 244), it was found that the variation of the simple effects of sequence of outcome was non-zero (significant at  $p < .01$ ) at all four levels of problem type. However, the variance of the effects of problem type was non-zero (significant at  $p < .05$ ) only at sequence +oo. Problem type was not significant at the other seven sequences. (Note: the numerical value of the significant-levels F for problem type was equal to the F from the analysis of variance). Although this analysis of variance was not the "usual" type because of subject representation, individual comparisons from the data on Table 3 were nearly identical to similar tests (F-ratios) using the data from Table 4.

Table 5 resummarizes the data of Table 3 to show the percentages of correct-choice response following each sequence of outcome by question and memory groups. Totals

TABLE 4  
SUMMARY OF ANALYSIS OF VARIANCE OF  
PROBLEM TYPE X SEQUENCE OF OUTCOME

Source of Variance	Sums of Squares	DF	F	P
A Problem type	1.40	3	2.94	<.05
B Sequence	21.39	7	19.13	<.01
AB	3.26	21	1.00	NS
within cell (error)	245.31	1504		
Total	271.36	1535		

for the question-once-per-problem and the question-after-each-trial groups, and a grand total for all conditions by sequence are included. These data show the traditional outcome effect; that is, as the number of rights increases from 0 correct (sequence ooo) to 3 correct (sequence +++), there is an increase in the probability of being correct on trial four. These data are consistent with Kornreich's (1968) and DePalma's (1969) studies. If the grand totals for these percentages across the question and memory groups are used in making individual comparisons between sequences, some significant differences are obtained. Sequence oo+ was significantly ( $p < .01$ ) greater than +oo and o+o. Sequence ++o differed significantly ( $p < .05$ ) from o++. And +++ was significantly different ( $p < .01$ ) from all the

other sequences. Subjects who were correct on trial one performed better ( $p < .01$ ) on problems than incorrect subjects.

TABLE 5

PERCENTAGES OF CORRECT-CHOICE RESPONSE ON TRIAL FOUR  
FOR QUESTION AND MEMORY GROUPS BY SEQUENCE OF OUTCOME

Sequence	<u>Groups</u>				<u>Totals</u>		
	<u>QAE</u>		<u>Q</u>		<u>QAE</u>	<u>Q</u>	<u>Grand</u>
	PP	NP	PP	NP			
ooo	67	62	57	61	65	59	62
+oo	66	70	70	62	68	67	67
o+o	66	55	67	63	60	65	63
oo+	82	82	72	77	82	74	78
++o	94	91	92	74	92	82	88
+o+	90	76	76	86	83	81	82
o++	81	76	74	82	79	79	79
+++	100	98	100	100	99	100	99

QAE = question after each trial  
Q = question once per problem  
PP = paper and pencil  
NP = no paper and pencil

In the experimental task 16 problems were given, so it was possible for each subject to get from 0 to 16 problems correct. Table 6 shows the frequency of subjects for each number of correct problems. The lowest number of

TABLE 6

## FREQUENCY OF SUBJECTS BY NUMBER OF CORRECT PROBLEMS

Number correct	QAE	Q
6	-	1
7	-	-
8	-	1
9	5	3
10	4	4
11	7	10
12	9	7
13	4	11
14	8	4
15	10	5
16	1	2
Proportion correct	$\frac{600}{768} = 78\%$	$\frac{581}{768} = 76\%$
Total	$\frac{1181}{1536} = 77\%$	
Means	12.5	12.1
Grand mean	12.3	

correct problems in the experiment was 6. There were 48 subjects in each question condition, so a total of 768 (48x16) problems were given to the subjects in each group. The proportions and percentages of correct, and the means are also given. These values correspond closely to those of earlier studies and indicate that there was no more interference in processing for subjects in these conditions than for those in Kornreich's (1968) or DePalma's (1969) studies. Whatever the effects of the experimental question were, they were not distinguishable from the procedural

effects of these other studies. However, during the experiment some subjects (question-after-each-trial) said they felt they were aided by responding aloud after each trial. Other subjects felt they had a "hard time doing the task," because of the questioning; or that they weren't always able to give a "reason" for their choice. As we shall see, such statements are more indicative of the subject's manner of processing than they are of the effects of the experimental question.

Table 7 represents the number of subjects giving correct-choice response on trial four as a function of problem number and problem type. This table shows that the best performance was on problem number 6 (color), while the worst occurred on problem 3 (size). As expected, there was no improvement over problems. (Note: if the answer the subject gave is compared with the stimulus to which he pointed, an interesting result is obtained. The subject does not always give the correct reason for his choice, even though he may point to the correct stimulus. If we examine the frequencies of such occurrences we find:

	Color	Letter	Size	Position
QAE (768 problems)	7	22	11	12
Q (192 problems)	8	5	5	11
Total	15	27	16	23

or 81 problems which were pointed to correctly, but given the incorrect reason for their selection! This number

would probably have been larger, but we did not always receive verbal responses on trial four in the question-once-per-problem condition, so there was no way of obtaining these numbers).

TABLE 7

NUMBER OF SUBJECTS GIVING CORRECT-CHOICE RESPONSE  
ON TRIAL FOUR FOR EACH PROBLEM AND PROBLEM TYPE

Problem #	Type	Number of subjects
1	L	77
2	L	78
3	S	56
4	P	79
5	C	70
6	C	85
7	L	70
8	C	74
9	S	63
10	S	80
11	L	72
12	P	69
13	P	76
14	S	69
15	C	82
16	P	81

Problem types = Color, Letter, Size and Position

The sequence of outcomes and the subjects' verbal responses to the experimental question (in the question-after-each-trial group) were also utilized to investigate subjects' hypothesis-sampling strategies. Table 8 shows the percentages of correct response on trial four and usage of hypothesis-sampling strategies according to

problem types. These data indicate that Win-Stay, Lose-Shift and Other were most frequently used, and most often correct, with Other being the "best strategy" overall.

TABLE 8  
PERCENTAGES OF CORRECTNESS AND USE OF  
HYPOTHESIS-SAMPLING STRATEGIES  
FOR PROBLEM TYPES

Problem types	Strategy									
	WSLF		WFLS		WFLF		WSLS		OTHER	
	%c	%u	%c	%u	%c	%u	%c	%u	%c	%u
Color	88	40	75	2	72	15	73	6	82	47
Letter	82	35	46	6	73	21	100	5	78	42
Size	75	35	50	1	64	23	81	6	67	45
Position	80	39	75	2	79	20	100	5	85	42
Totals	82	37	57	3	72	20	88	5	78	44
<div><div></div><div>7219</div></div>										

W = win (right)  
L = lose (wrong)  
S = stay (keep hypothesis)  
F = shift (change hypothesis)  
%c = percent correct  
%u = percent of time used



## DISCUSSION

The solution analysis shows that of question, memory aid and problem type only the latter had a significant effect on subjects' performance. There were no differences between the performances of the question-after-each-trial and the question-once-per-problem subjects, or between the no-paper and the paper-and-pencil groups.

Size problems were shown to lead to much lower performance than any of the other problem types. This supports the hypothesis that problem type is an influential factor on performance, but it remains unexplainable at this time.

From the responses the subjects gave to the experimental question, it was observed that (see Table 8):

- 1) Subjects do change hypotheses after "right"s as well as after "wrong"s.
- 2) Subjects don't always pick hypotheses consistent with information received from an error trial.
- 3) Subjects may think they are being consistent with prior information when, in fact, they aren't. Thus, subjects do consider hypotheses which fail to pass a consistency check.

4) Subjects give the same hypothesis frequently during a problem--even after having been "wrong" with it on a previous trial. Therefore, the subjects sample with replacement from the hypothesis-pool during the selection process. These observations support predictions two through five.

In the sequence of outcome analysis of variance, both sequence of outcome and problem type were shown to have significant effects on performance. Since the interaction was zero, these main effects are presumably additive. From the individual F-tests (Winer, p. 244), it was found that the variation of the simple effects of B (sequence) was not zero at all levels of A (problem type). This was expected, because as the number of "right"s on the first three trials increases, the probability of correct-choice response on trial four increases. Thus, the variation among the effects of the sequence should be quite high. The variance of the effects of problem type, however, was non-zero only at level  $B_2$  (sequence +oo). This can be explained by the extremely low level (relatively) of performance for the size problems at this sequence. There was a difference of 24 percentage points between the value for the size problem (at +oo) and the highest value at +oo. All of the variance related to factor A can be accounted for by size problems (see pp. 23-25). Thus, the relationship of sequence of outcome to performance seems to be quite complex.

The traditional outcome effect can be observed in Table 5. As the number of rights increases from 0 (in ooo) to 3 (in +++) there is a corresponding increase in the probability of correct-choice response on trial four. However, the individual comparisons among sequences reveal a complex relationship between sequence of outcomes, number of rights, and performance. It seems that for two or more rights on the first three trials, being right on trial one is more important (with respect to performance) than being correct on trials three or two, respectively. For one right, being correct on trial three leads to better performance than being correct on trials one or two. So for two or more rights, primacy is more influential than recency; and for one correct recency is more important than primacy.

It is also possible that the outcome effect is actually the result of the number of transformations the subject must make during the problem. For every "wrong" outcome, the subject must "transform" this information in terms of "right" information, that is, he must determine what the "wrong" information means in terms of "right" information. If we arrange the sequences in order of increasing number of transformations we would have: ooo, o+o, +oo, oo+, +o+, o++, o++, ++o, and +++. It is also assumed here that sequences o+o and +o+ will be more difficult than the other sequences with the same numbers of

rights. In both sequences, there is an interruption in the consistency of information from one trial to the next. In the other six sequences at least two similar outcomes (similar information) follow one another. The data from this study do not quite satisfy this prediction. Instead we find, in order of increasing probability of solution: ooo, o+o, +oo, oo+, o++, +o+, ++o and +++. So sequences o++ and +o+ have exchanged positions. Perhaps, the facilitative effect of primacy was stronger than the detrimental effect of the transformation-interruption.

Although we have discussed the effect of sequence on performance, we have not included the relationship of sequence of outcome to the subject's processing. By examining the outcomes the responses given by the subjects in the question-after-each-trial group, we were able to observe different percentages of correctness (how often the response on trial four was correct) and use (proportion of the time a particular strategy was employed on the problems for all subjects) (see Table 8). From this table it is clear that the best single strategy is WSLF (Win-Stay, Lose-Shift). This strategy was used on 37% of the problems, and when used resulted in correct trial-four response 82% of the time. The remaining strategies combined were used by subjects on only 19% of the problems, and were correct 72% of the time. The remaining workable strategy was actually not one of these (Win-X, Lose-X)

types, but a "combination." That is, the subjects from the Other strategy did not use a specifiable strategy on those problems. This "strategy" was 78% correct and utilized 44% of the time. So overall, it was the most effective "strategy."

These data indicate that it doesn't matter particularly if the subject has a consistent strategy or not, but rather than he uses the information he receives correctly. This statement was verified by some observations of the paper-and-pencil subjects. After a wrong outcome, some of these subjects actually wrote down the incorrect (pertaining to "wrong" outcome) information. The incidence of this phenomenon varied from subject to subject (and sometimes from problem to problem in a single subject!). Such subjects found it very difficult to solve the problems, especially if they wrote down the incorrect information and on the following trials treated this information as correct information. Of course, such processing cannot lead to correct solution of the problem. Other subjects only recorded some of the information they received from the outcome. These subjects would then need more than four trials to solve a problem, so they were unsuccessful in attaining problem solution. Another interesting observation was that subjects wrote down (processed) more than one hypothesis at a time, yet they gave only one hypothesis as the reason for their choice

(in response to the question). This confirmed Levine's earlier work.

It is assumed that the data obtained from this paper-and-pencil group are representative of subjects' mental processing. It seems reasonable to conclude that subjects who can use correct ("right") and incorrect ("wrong") information effectively will solve the problems despite the experimental conditions (question and memory). Subjects who have trouble processing information (especially "wrong") for one of the reasons described (or any other) may be affected positively or negatively by the same methodology. The processing and hypothesis-sampling of such subjects merit further study.

## REFERENCES

## REFERENCES

- Andrews, O., Levinthal, C., and Fishbein, H. The organization of hypothesis-testing behavior in concept-identification tasks. American Journal of Psychology, 1969, 82, 523-530.
- Bourne, L. and Guy, D. Learning conceptual rules II: The role of right and wrong instances. Journal of Experimental Psychology, 1968, 77, 488-494.
- Bourne, L., Dodd, D., Guy, D., and Justesen, D. Response contingent ITS in concept identification. Journal of Experimental Psychology, 1968, 76, 601-608.
- Bower, G. and Trabasso, T. Concept identification. In R. C. Atkinson (ed.) Studies in Mathematical Psychology. Stanford: Stanford University Press, 1964.
- DePalma, D. unpublished senior thesis, Lehigh University, 1969.
- Dodd, D. and Bourne, L. Test of some assumptions of a hypothesis-testing model of concept identification. Journal of Experimental Psychology, 1969, 80, 69-72.
- Dominowski, R. Role of memory in concept learning. Psychological Bulletin, 1965, 63, 271-280.
- Erickson, J. Hypothesis sampling in concept identification. Journal of Experimental Psychology, 1968, 76, 12-18.
- Falmagne, R. Construction of a hypothesis model for concept identification. Journal of Mathematical Psychology, 1970, 7(1), 60-96.
- Falmagne, R. A direct investigation of hypothesis-making behavior in concept identification. Psychonomic Science, 1968, 13(6), 335-336.
- Kenoyer, C. and Phillips, J. Some direct tests of concept identification models. Psychonomic Science, 1968, 13(4), 237-238.



- Kornreich, L. B. Strategy selection and information processing in human discrimination learning. Journal of Educational Psychology, 1968, 59, 438-448.
- Levine, M. Hypothesis behavior by humans during discrimination learning. Journal of Experimental Psychology, 1966, 71, 331-338.
- Levine, M. Latency-choice discrepancy in concept learning. Journal of Experimental Psychology, 1969, 82, 1-3.
- Merryman, C., Kaufmann, B., Brown, E. and Dames, J. Effects of "rights" and "wrongs" on concept identification. Journal of Experimental Psychology, 1968, 76, 116-119.
- Nahinsky, J. and Slaymaker, F. Sampling without replacement and information processing following a correct response in concept identification. Journal of Experimental Psychology, 1969, 80, 475-482.
- Nunnally, J. Psychometric Theory. New York: McGraw-Hill Inc., 1967.
- Restle, F. The selection of strategies in cue learning. Psychological Review, 1962, 69, 329-343.
- Richter, M. Memory, choice and stimulus sequence in human discrimination learning. Unpublished doctoral dissertation, Indiana University, 1965.
- Rogers, S. and Haygood, R. Hypothesis behavior in a concept-learning task with probabilistic feedback. Journal of Experimental Psychology, 1968, 76, 160-165.
- Rourke, D. and Trabasso, T. Hypothesis sampling and prior experience. Proceedings: 76th Annual Convention APA, 1968, 47-48.
- Trabasso, T. and Staudenmayer, H. Random reinforcement in concept identification. Journal of Experimental Psychology, 1968, 77, 447-452.
- Trabasso, T. and Bower, G. Presolution dimensional shifts in concept identification: A test of the sampling with replacement axiom in all-or-none models. Journal of Mathematical Psychology, 1966, 3, 163-173.

Trabasso, T. and Bower, G. Memory in concept identification. Psychonomic Science, 1964, 1, 133-134.

Winer, B. J. Statistical Principles in Experimental Design.  
New York: McGraw-Hill Book Co., 1962.

**APPENDIX A**  
**EXPERIMENTAL DESIGN**

## APPENDIX A

### EXPERIMENTAL DESIGN

Question frequency	Memory aid	Number of subjects	<u>Problem Type</u>			
			Color	Letter	Size	Position
Q	NP	24				
	PP	24				
QAE	NP	24				
	PP	24				

**APPENDIX B**  
**QUESTION MATRIX**

# APPENDIX B

## QUESTION MATRIX

Questions asked after trial	<u>Problem Type</u>			
	S P C L	S P C L	S P C L	S P C L
1	*		*	*
2		*		*
3			*	*
4		*	*	*
1		*	*	*
2	*		*	*
3		*	*	*
4		*	*	*
1		*	*	*
2			*	*
3	*		*	*
4	*	*	*	*
1	*	*	*	*
2		*	*	*
3		*	*	*
4	*	*	*	*

NOTE: This matrix represents problem types (size, position, color and letter) vs the trial after which the question is asked.

MAY 5 1971

MICHIGAN STATE UNIV. LIBRARIES



31293103829085