EXPERIMENTS IN PROBLEM SOLVING

by

Gerald Marvin Weinberg

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in the
University of Michigan
1965

Doctoral Committee:

Professor Anatol Rapoport, Chairman
Associate Professor John H. Holland
Professor Irwin Pollack
Associate Professor William R. Uttal

PREFACE

The thoughts and work of this thesis represent a distillation of the contributions of many people over a period of years. Gordon Peterson, June Shoup, and all the other members of the faculty in Communication Sciences at the University of Michigan played a central role in this development. The contributions of the members of my doctoral committee--Anatol Rapoport, John Holland, Irwin Pollack, and William Uttal--were of special significance. It was also my particular privilege and good fortune to have had the advice and assistance of Paul Fitts, who was a member of my doctoral committee until his shocking and untimely death.

The use of computers was critical to this thesis. The PDP-1 which was used for the on-line experiments was made available by Ward Edwards. It was provided by the Electronics Systems Division of the Air Force Systems Command under contract AF 19 (628) - 2823. The University of Michigan Computing Center and the IBM Systems Research Institute supplied the computer time on the IBM 7090 and IBM 1401 which were used for the analysis programs. I also wish to thank Stanley Bielby at the Human Performance Center, Bernard Galler at the University of Michigan Computing Center, and George Payne at the IBM Systems Research Institute for their assistance

in overcoming the numerous small problems incident to using those computers. One other piece of indispensible equipment--an IBM Executary--was provided by the IBM OPD office in Dearborn, Michigan.

It was the generous award of an IBM Resident Study Fellowship which gave me the opportunity to come to the University of Michigan and undertake this program of study in Communication Sciences. Of all the people at IBM who have given me support and encouragement throughout this period, I owe special thanks to William Carpenter, Frank Beckman, and John McPherson, without whom this thesis could certainly never have become a reality.

To Sheila Abend and to the people at Professional Service Associates, I owe the technical assistance necessary to prepare this manuscript. Finally, I must thank my anonymous subjects, each of whom graciously and uncomplainingly volunteered an appreciable amount of time to address himself to the difficult task which these experiments presented.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Figure	Page

CHAPTER I

INTRODUCTION

A. The Problem of Problem Solving

A very popular area for recent work in Communication Sciences might be called "Problem Solving"--the study of adaptive systems with the ability to perform inductive symbolic behavior. As far as we know, this study includes only two kinds of systems--men and computers.

Appropriately, the work in this field has followed two main lines: first, the study of how human beings go about solving problems and, second, how machines might be constructed or programmed to solve problems. Within each of these two approaches, there is also a division between the theoretical and the practical. Thus, in the study of how human beings solve problems, research for its own sake would be exemplified by Bruner's work on "concept attainment" or Piaget's work on "concept formation," while Polya's works on problem solving are examples of the direct approach to "improving" problem solving techniques.

The studies of problem solving in men and machines are not at all independent. Certain results with cybernetic machines have strongly influenced such writers as Miller, Galanter, and

Pribram in their Plans and the Structure of Behavior (though perhaps prematurely). More particularly, researches into machine problem solving usually start with some conceptual model based on human performance data. These data may be primarily introspective, as in Samuel's checker-playing work or Friedberg's programming machine, or they may be based on a more or less extended series of observations, often taken especially for the purpose of constructing machine models, as in Newell, Shaw, and Simon's "protocols."

In general, however, such introspective or specially taken data would not be considered up to the standards of empirical rigor commonly accepted among modern psychologists. Furthermore, they do not really seem to get below the level of the subject's verbalizations, to the level of the processes which are often merely being masked by those verbalizations. A deeper level might be more appropriate for modeling machine simulations. On the other hand, the people working in machine problem solving argue that the great mass of rigorously taken psychological data is irrelevant to their models--either because it, too, does not penetrate the obfuscating layer of verbal behavior or because it penetrates that layer only by investigating situations which are inherently too simple to force the subject to display just that complex behavior which distinguishes "problem solving" from more general forms of adaptive behavior.

It appears, then, that there is a great need for fresh data on human problem solving behavior: data taken with due respect for both the rigor demanded by psychologists and the special needs of builders of problem solving models. Specifically, such data should be precise and operational. They should not be impressionistic, nor dependent on verbal report. They should describe situations not so desiccated by the requirements of data analysis that they no longer represent problem of interest, yet not so influenced by specific ideas of interesting problems that they are incapable of throwing light on more than one model. In other words, these data would begin to fill the gap between the rigorous but simplified experiments of psychophysics and the loose but complex experiments of classical psychology, while at the same time providing a link between the two lines of research on problem solving behavior.

B. A Technique for Investigating Problem Solving Behavior

We cannot reasonably expect that such distinctive data will be obtained by merely extending or elaborating existing techniques. We want, at one and the same time, to investigate more complex behavior and to descend at least one level of detail in the gathering and analysis of the data about that behavior; thus, we will tend to complicate the work involved in at least two ways. The use of computers naturally suggests itself, and although computers have already been widely used in psychological research, for the most part they have been confined to doing work that is not different conceptually from work mechanized by more specialized apparatus. Let us examine how using computers--both to run subjects and to analyze the data collected from such runs--enables us to perform a wide class of interesting experiments.

A class of problem solving tasks (which is imbedded within a problem solving situation) may be described in the following abstract way:

a. The subject sees a sequence of (symbolic) instances selected from a set Y which is in turn selected from a larger set, U, which is common to all problems of a given type.

b. The subject makes a sequence of responses, each selected from a set, R, also selected from a larger set, V, which is common to all problems of a given type.

c. The subject "solves" the problem by matching as closely as possible his responses with a set of "correct" responses, or more generally, by obtaining "payoff" from the system at or better than some prescribed rate. He may or may not be able to "explain" what he is doing.

A general computer program capable of conducting any one of a very large class of problems has been constructed. The experimenter has under his control the following variables:

a. The set U.

b. The mapping, M_u, from (0,1) into U which selects the set Y.

c. The set V.

d. The mapping, M_v, from $(Y^* \times R^*)$ into V, which selects the correct response at each instant of time.

e. The mapping, M_y, from $(Y^* \times R^*)$ into Y, which determines the sequence of instances seen by the subject.

f. The mapping, M_p, from $(R^* \times V^*)$ into the real numbers, which determines the "payoff" S is to receive.

g. The mapping from the various payoffs to the indications of them fed back to S.

h. The rate, or "pacing" of the experiment.

Now, the number of experiments made available by such a general apparatus is extraordinarily large, even with a very small computer such as a PDP-1. Furthermore, it is extremely easy to generate new experiments--the computer itself may even be programmed to do this automatically. Thus, the experimenter is faced not with the problem of which experiments he can manage to set up and perform, but with the necessity of choosing which experiments he really wants to do--given that all are equally easy.

For example, if the mapping, M_v, is one-to-one, the experiment becomes one in rote memory. But given a typical set, U, such as the one chosen for the work of this thesis which has 2^{40} elements, and a response set, V, with perhaps 2^7 elements, there are about $(2^{4000}!)$ <u>precisely specifiable</u> rote memory experiments which can be performed--not considering the variations in payoffs, feedbacks, and rates. Because of the precise specification of these experiments, the experimenter can systematically vary parts of the specification without losing track of the relationship of one experiment to another. In fact, he is able to quantify this relationship according to various theories about the significance of different aspects of the experiment.

As an example of this ability to quantify, the set, Y, can be chosen so as to have a precise "distance" (or distribution of distances)--in the coding sense or in any other defined sense--between each pair of its members. This distance could be systematically varied in order to study its appropriateness as a measure of interference leading to variations in difficulty in the rote learning (or other) problem.

Actually running the subjects on-line at the computer not only enables us to broaden the class of experiments (because R may be included in the definitions of the mappings), but also permits the most extreme thoroughness and precision in recording the data taken. Because of the precision with which the experimental structure can be specified, it makes sense to record the data with more precision than is usually demanded in psychological work. Such precision, however, demands more thorough analysis of the data, yet if the work is to be exploratory, a great flexibility in methods is also needed. Ideally, one would like to have a large computer with a random access file containing the data from a number of experiments and available for trial analyses through an on-line programming system. Lacking this, the next best alternative is to have the large computer with a highly flexible programming language and operating system. This we do have; and though it slows down the pace of the exploratory research, it at least makes it reasonably possible.

To summarize, then, let us look at the overall procedure by which an investigation is conducted. Once the general programs have been checked out, the experimenter can specify new experiments by constructing appropriate tables and compiling them on the computer. The resulting tape, when read by the general programs, transforms the computer into an experimental apparatus for performing the specified experiment. In exploratory work, the experimenter might run a few subjects on this apparatus, yielding a magnetic tape with all the performance data. (These data may be supplemented by observations by the experimenter on the external behavior of the subjects or by verbal reports by the subjects of their impressions of the experiment; but this step is not necessary, though it is performed in this thesis.) This tape is then available for computer processing, using whatever special programs the experimenter might wish to add to or select from a set of general routines already available. If the work is exploratory, a number of cycles of developing new analysis techniques and measures, programming them, and looking at the results will take place. Finally, new experiments will be suggested by these results, and the entire process beings anew.

C. The Experimental Apparatus

1. Overall Organization

Figure 1-1 is a data flow diagram of the overall experimental and analysis procedure. To prepare for a running of one or more experiments for a subject, the experimenter loads the general experimental program into the PDP-1 and readies the equipment,

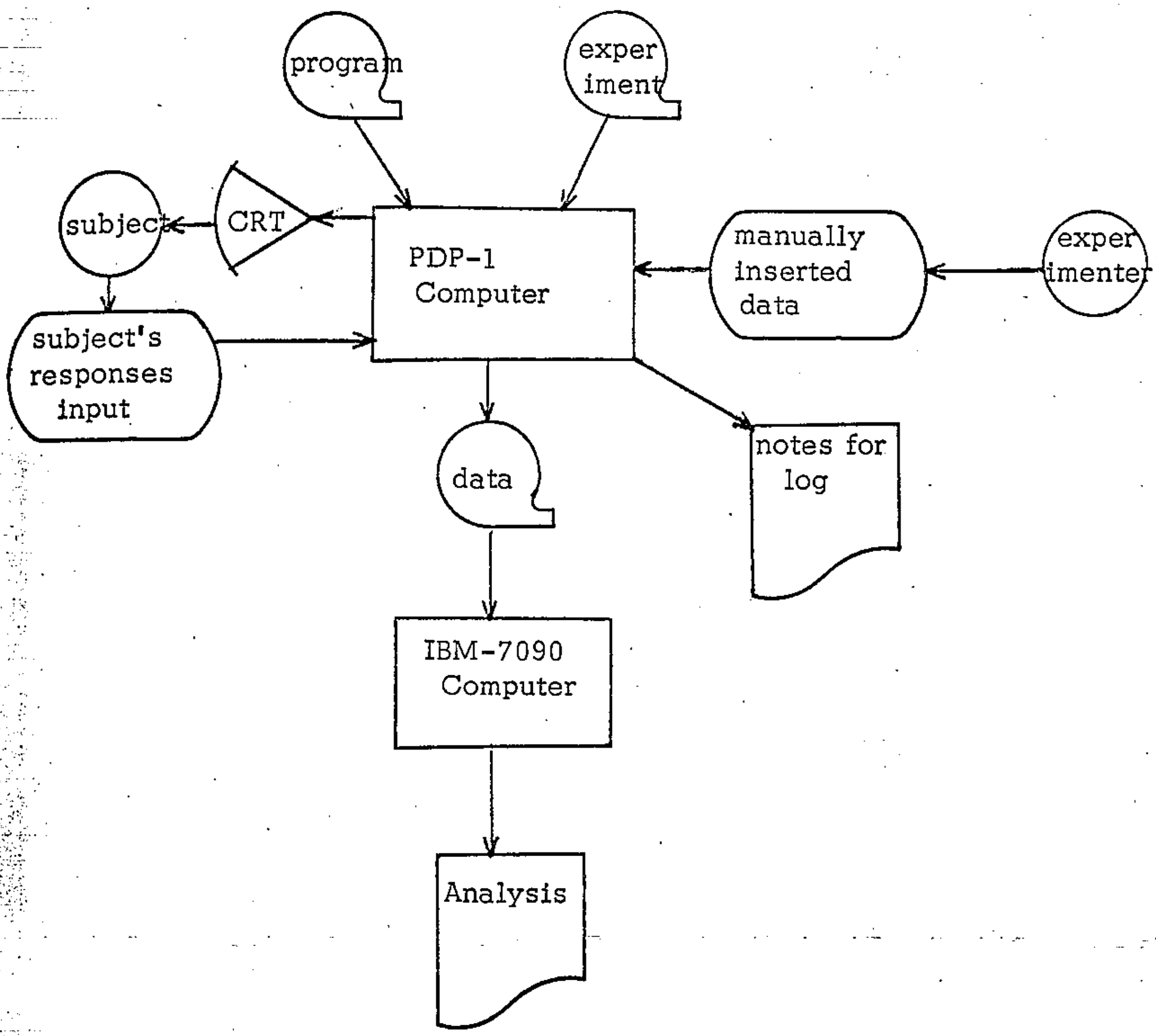


Fig. 1-1.--Overall experimental and analysis procedure

including mounting the magnetic tape for data recording. Each individual experiment is begun by loading its specifications from a previously prepared paper tape and entering the subject's name, the date, and any other pertinent information through the typewriter. These data are recorded on the magnetic tape, and as the experiment is carried out, a file is created on that tape containing all the experimental data, ready for later analysis. The system is designed to permit any number of individual experiments with the same or different subjects to be run consecutively without additional machine setup. Thus, a data tape of several files may be written.

During the running of an experiment, the typewriter produces a typed log of which may be recorded such items as, when the subject asked to pause for rest, when there was machine difficulty, how many trials each experiment takes, as well as the information which identified the experiment. The experimenter may also type any notes he wishes on this log.

2. Organization of the On-line System

Each experiment is divided into three major phases. The main part, of course, is the experimental trials themselves, but this portion is preceded by two others: initial setup and practice. Initial setup permits the experimenter to record the subject's name, the time of day, the date, any unusual conditions, and so forth. At the same time, it permits him to make adjustments to any of the experimental variables, such as the rate of pacing the trials, and to load new program portions which will define the new experiment.

After the initial setup, the subject is given whatever instructions are in effect for this experiment. The experimenter then starts a practice run, in which the subject can see how the stimuli look and what his responses are to be and can make any adjustments necessary for his comfort. No reinforcement is given during this practice. When the subject feels that he is ready to begin, the experimenter starts the experiment.

Once the experiment starts, everything is completely automatic until it terminates. After termination, the system is ready to run another experiment with the same subject or the same experiment with a different subject. The data from one or more experiments is accumulated on a magnetic tape, ready for analysis by automatic means.

The detailed structure of the experimental phase is given in Chapter II.

3. Organization of the Analysis System

Figure 1-2 is a data flow diagram of the analysis system, which uses the IBM 7090 as the analysis instrument. As a first step, the experimental tape is processed by the Data Compression Program, producing a data listing, and a pair of duplicate punch card decks for each experiment. This compression step is not a logical necessity, for the analysis could proceed directly from the magnetic tape; but it is performed for reasons of efficiency and safety. Ideally, of course, these duplicate, error-protected data sets should be produced by the on-line computer as the experiment is being run.

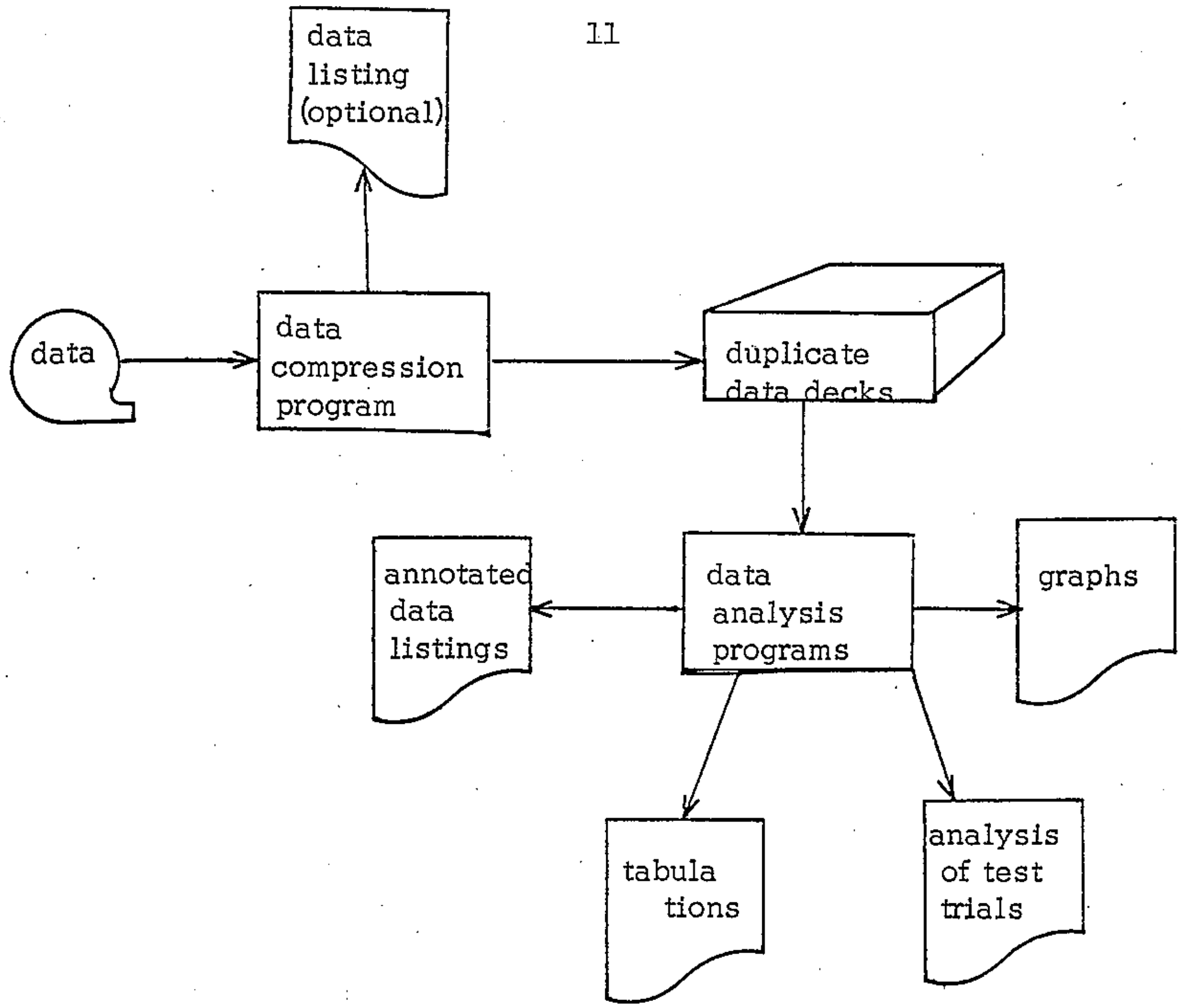


Fig. 1-2. -- Organization of the analysis system

Another compromise with reality was made in choosing to convert the magnetic tape to card decks. Although in theory the magnetic tape should be more efficient at storing the experimental data, the University of Michigan Computing Center is set up in such a way that, for the type of computations being done, using card decks is more efficient and more flexible.

The data analysis programs are built into a modular structure which permits the experimenter to choose any combination of a variety of different analyses for each experiment or, if he desires, to add special new analyses to the system. These programs are described in detail in Chapter III.

D. The Scope of the Experiments of This Thesis

The main emphasis of this work is on obtaining a more precise description of the human being viewed as an abstract, symbolic problem solver. To do this, every experimental design decision has been made in such a way as to minimize the effects of the non-symbolic environment and of inter-subject variation.

Under non-symbolic environment, we can group all the effects of variations in stress, instructions, motivation, and physical conditions, as well as social-psychological effects of using teams of subjects. Every attempt has been made to make subjects as comfortable as possible; motivation has been left to the inherent interest of the task; instructions are simple and relatively standard; and only individual subjects have been used.

Of course, even the most standardized set of conditions is subject to varied interpretation on the part of the subject, so the environment will never be constant over different subjects or even over the course of time with a single subject. Some of this variation will be eliminated by focusing the study on a sequence of problems solved by a single subject, and some of the motivational variation may be dominated by the intense motivation induced by the problem solving environment itself.

Inter-subject variation includes the effects of language; culture; personality; previous training; innate ideas about such things as causality, incongruity, stereotypes, a priori probabilities, and psychological experiments; physical capacity; and verbal ability. Subjects have been screened for any obvious impediments or mental disturbance. The task is designed to be well within the range of normal physical and sensory skills and to require very little in the way of verbal performance. Furthermore, the intensive study of individual subjects permits the cancellation of much of this variation by contrasting different experiments performed by the same subject.

In the course of analyzing the experiments, insights into the role of the non-symbolic environment and inter-subject variation will, of course, occassionally arise even though the experiments were designed to minimize such effects. No attempt will be made to conceal such insights; but, on the other hand, no attempt will be made to make them logically rigorous. The main concentration of our analyses will be on answering questions which fall into three classes--Decision Structures, Strategies, and Verbal Behavior. The following are typical of the general types of questions we shall investigate:

1. Decision Structures

    a. To what extent can human decision processes be modelled with deterministic computer type flow diagrams or fixed Boolean expressions?

    b. Are certain structures favored over others?

    c. What is the inter-experiment influence of decision structures?

    d. Are certain features of stimulus or response favored or disfavored in decision structures?

    e. What are the dynamics of the evolution of decision structures? Is there, for example, any combination, shuffling, or discarding of structural parts over time?

2. Strategy

    a. What strategies do subjects adopt?

    b. In what sense--if any--are those strategies optimal?

    c. Are there compatability effects between stimuli and responses which the strategy must overcome?

    d. How do other <u>a priori</u> ideas influence the subject's strategy?

    e. What do strategies reveal about limits of human information processing capabilities, such as, memory capacity, speed of logical processes, and input-output rates?

    f. Are inductive steps quantal?

    g. What role does rote memorization play in strategy?

3. Verbal Report

   a. How accurately does verbalization reflect behavior?

   b. How do the strategy and decision structures influence verbalization?

   c. Does verbal facility increase over time?

   d. Does verbal facility simplify problem solving?

Naturally, one series of experiments cannot be expected to give final and conclusive answers to even one such question. Therefore, one of the important objectives of this thesis is to outline an integrated series of experiments using such a generalized apparatus which would constitute a systematic attack on obtaining more precise answers.

CHAPTER II

THE STRUCTURE OF AN EXPERIMENT

A. General Description

An experiment consists of a series of trials, which is terminated when (a) a terminating criterion is met, or (b) the experimenter decides to terminate it manually--as when the subject is too discouraged to continue.

A trial consists of the following parts:

i) presentation of the stimulus to the subject.

ii) response by the subject, or, if no response if given within the prescribed time interval (5 seconds) automatic stepping to part iii and recording "no response."

iii) presentation to the subject of the response he just made, the response he should have made to be "right," and the stimulus pattern, which remains on the screen. After one second, the reinforcement is terminated and the next trial begins.

During the trial, the computer is calculating what the next trial will be and recording, after the response, the results of the trial. These tasks, however, are executed with such speed that the subject only perceives a continuous sequence of steps i, ii, and iii, with no breaks between them. When the experiment is terminated, the screen goes blank and a terminating message is typed out.

The entire process of a trial as seen from the computer's point of view is shown in Figure 2-1. The <u>test trial</u> is a special type of trial in which the reinforcement step is omitted. It permits the system to test the current status of the subject by mutilating the stimulus in various ways.

B. <u>Choice of value and noise bits</u>

If, in a given experiment, there are N different stimuli, then a string of $\lceil \log_2 N \rceil + 1$ bits is needed to specify a particular stimulus. If the stimuli are equally probable, of course, $\log_2 N$ is merely the measure of the information in a given stimulus presentation. In any case, choosing a stimulus from among the set of possible stimuli in an experiment is equivalent to choosing a string of bits of a certain length, and this is exactly what is done by the computer program.

Just how the choice is made is subject to variation from experiment to experiment. Thus, within the set of variable routines is a subroutine whose function is to produce a string of bits which determines the next pattern to be shown. Since the response associated with each string is also subject to experimental variation, this routine also produces a string of bits designating the string's value.

In other words, this routine contains and executes two mappings: (1) the mapping from $(R^* \times Y^*)$ into Y, and (2) the mapping from $(Y^* \times R^*)$ into V. Actually, these mappings may be much more general than this. For example, the class of

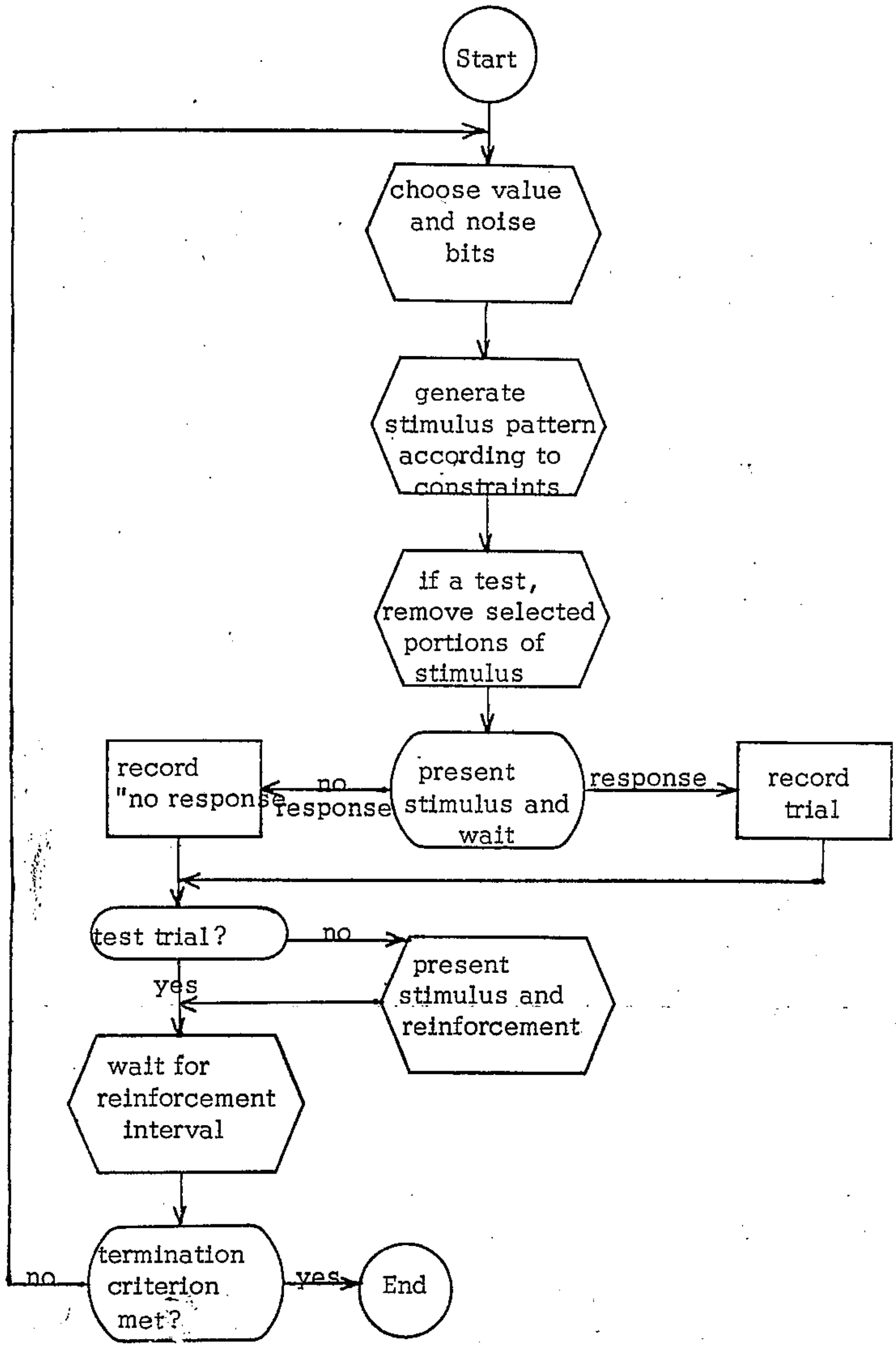


Fig. 2-1. -- The structure of one trial

possible experiments includes those in which the next stimulus chosen depends not only on the previous stimuli and responses (with any degree of randomness), but on such factors as the previous response time, or the average response time up to that point. Whatever variation of sequencing used, this routine is all that needs modification. In this work, only one method of sequencing is used--namely, random sequencing where Y is independent of (R* x Y*)--so no attempt was made to build in more general sequencing routines.

The choice of value associated with each pattern is also subject to a great many variations. Not only can the value be associated with sequences of patterns, but it can be associated with a single pattern in a vast number of ways. Each bit of the value can be specified by a logical expression relating all of the bits in the pattern specification. If there are k bits in the pattern specification, we can calculate how many unique expressions there are for each value bit as follows:

Each unique expression has a unique normal form (conjunctive or disjunctive). Each term in a normal form may have each of the k bits appearing in it as a complement, not complemented, or not appearing. Thus, there are 3^k--1 different terms, exclusing the empty term. But each term may either appear or not appear in the expression, so there are $2^{(3^k-1)}$ expressions, including the one that does not depend on the pattern bits at all. Since each bit of the value may have its own expression, if there are j bits, there are

$$(2^{(3^k-1)})^j = 2^{j\cdot(3^k-1)}$$

different mappings possible. Even for k = j = 2, this number is very large (2^{16}), a fact not always appreciated by students of concept formation.

Naturally, in order for an experiment to represent more than rote memorization, there must be some structure to this mapping. In the experiments performed for this thesis, $k = 7$, $j = 3$, and the same mapping was always used. That mapping is the symmetrical tree structure specified by

value bit 1 = pattern bit 1 (called bit A)

value bit 2 = pattern bit 2 (called bit B)

value bit 3 = pattern bit 3 (called bit C)

In this mapping, four of the pattern bits do not enter into the determination of pattern value at all. Thus, they are referred to as "noise" bits, numbered 1, 2, 3, and 4. With this structure, it is the noise bits which cause patterns with the same value to have different appearances, thus the term noise bits is justified for them. Since there are four independent noise bits in each specification, there are 2^4, or 16, different patterns associated with each value. There are, of course, 2^7, or 128, patterns in all.

C. Appearance of Stimulus Patterns

Experiments may differ not only in their sequence structure and their mappings from pattern designation to pattern value, but also in the representation of the pattern to the subject. Ideally, we should like to choose a set of pattern elements which were *a priori* indifferent as far as each subject were concerned. That is, each element of the pattern would have, for the subject, the same subjective probability of being "meaningful." We know, however, that this is impossible to achieve. In fact, experiments such as Bruner's concept attainment tasks using pictures of adults and children have been able to demonstrate quite clearly that each subject brings preset categories to the experiment--categories which are not a matter of indifference to him. One of the possibilities which our general computer program opens up is that of pretesting subjects to find a set of approximately indifferent pattern elements for each individual subject, because each pattern element displayed on the cathode ray tube is determined merely by a subroutine. The pretesting could determine a reasonably "good" set for the subject by trying a large variety of elements from a library. This set would then be used in the actual experiments.

In order to simplify the experiments, however, a different path was taken, one more in accord with classical methods in psychology. Instead of choosing a new set of pattern elements for each subject, preliminary testing was done in order to find a set that was reasonably indifferent to a representative sampling of people. This method is perhaps better than it might appear because of the existence of common cultural experiences, which tend not only to create common cultural categories, but which create common cultural "anti-categories," that is, groupings of attributes that are given no special importance.

In order to provide the possibility of sufficiently complex tasks, it was necessary to choose a set of pattern elements which would allow a great variety of stimuli to be presented, all within the same structure. It proved difficult, however, to display more than seven or eight binary attributes in a single figure while retaining relative neutrality and discriminability simultaneously. For instance, it was suggested that human faces be used. Here, indeed, it is possible to code a fair amount of information in a rather discriminable form, due to the subjects' familiarity with the task of classifying faces. By combining, say, eight different foreheads, eight different eyes, eight different noses, eight different mouths, and a few other features, we could pack perhaps fifteen or twenty bits into a single computer-generated face. But surely we would here by probing more into the subject's preset category schemes than into his abstract categorizing abilities--a not unworthy task, but not the one undertaken in this thesis.

Actually, even the twenty bits of a single face seemed to present somewhat too narrow of a limitation. Forty bits or so seemed more appropriate to permit pushing experiments up into regions not previously explored, particularly since many of those bits would be used up in various forms of controlled redundancy. One way to achieve forty bits would be to have two twenty bit faces; but interesting as this technique might be to social psychologists, it presented an even worse problem of bias than would single faces. The principle of gaining information capacity through replication is a useful one, however. In fact, it probably represents the only practical way to display as many as forty or more bits at once. Thus, for these experiments an array of letters of the alphabet was chosen as the method of presenting the stimuli.

Let us examine this array in more detail, along with the reasons for the particular choice of elements and structure. In the first place, the number of elements in the array is also somewhat limited by requirements on legibility. With the available screen, no more than perhaps sixteen letter positions would be practical. Possibly more serious is the problem of positioning the letters, since we know that there are strongly preferred positions in any array. A linear string, for example, was ruled out, since it was too suggestive of a textual message--which would almost force the subject to scan it from left to right and perhaps even impose grammatical structures on it. In order to break up such scanning, a rectangular array was used, but this also presented problems of preferred positions. A two-by-two array seems to have relatively neutral position structure, but would require coding ten bits into each position. A four-by-four array would require fewer bits per position, but seems to divide quite sharply into an inside and an outside set of positions--with the corners occupying an important role slightly secondary to the center. A three-by-three array would be just about ideal--except for the center position--so the center position was left out, giving a square array, three characters on a side. The behavior of subjects in the actual experiments does nothing to indicate any significant bias in this arrangement.

With eight character positions in the array, five bits per character are needed to attain forty bit stimuli. One way to get these five bits would be to add six characters to the ordinary Roman alphabet, but such a procedure does not permit any single bit elements in the array. We should like to have such elements so we can see how and when a subject combines them to produce concepts. This indeed, has been the classical form of the concept attainment experiments. We should also like, however, to have some elements in the array which tend to convey more than a single bit per element, so we can see whether subjects perfer such larger chunks, whether they use them differently, and whether they ever "recode" several small elements into one big one, or vice-versa.

By such reasoning, we can obtain several equally good choices for division of the bits among the attributes of a character. The one division actually used in all the experiments is the following:

> Each character position is occupied by one letter chosen from a set of eight (3 bits); this letter may be capitalized or lower case (1 bit); and the letter may appear bright or dim (but quite visible) on the screen (1 bit).

This division gives us a total of five bits per position times eight positions, or forty bits in all.

The choice of the eight letters deserves some attention as well. Vowels were excluded since this was the easiest way to exclude words, which would evoke preset categories in a very strong way. The eight letters chosen had to form a set meeting the following requirements:

a. The upper case of each was clearly and easily distinguishable from its lower case form.

b. The letters had to be equally distinguishable from one another.

The eight letters chosen are shown in Figure 2-2, drawn essentially as they appear on the screen to the subject. Care was taken to make the letters meet the above two criteria while avoiding any artificiality. Subject's responses give no indication that these criteria were not met.

The eight letters are represented in an internal code as follows:

R = 000, M = 001, V = 010, Z = 011, D = 100, H = 101, K = 110, and T = 111.

This coding was chosen as carefully as possible to eliminate easily recognized substructures to this three bit assembly. For example, the phonetic similarity of letters can affect memorization and grouping, so the group (V, Z, T, D) was broken up so that these would be unlikely to code as a group in any experiments. Similarly, the alphabetic order was carefully broken. Still, it would be impossible to assure that every subject would have no preconceived ordering of this set of letters, but the experimental results show no apparent bias in this direction. Naturally, we cannot prevent subjects from learning groupings in one experiment and carrying them into the next; nor would we wish to if we could, as this type of behavior is one of the aspects of problem solving we wish to observe.

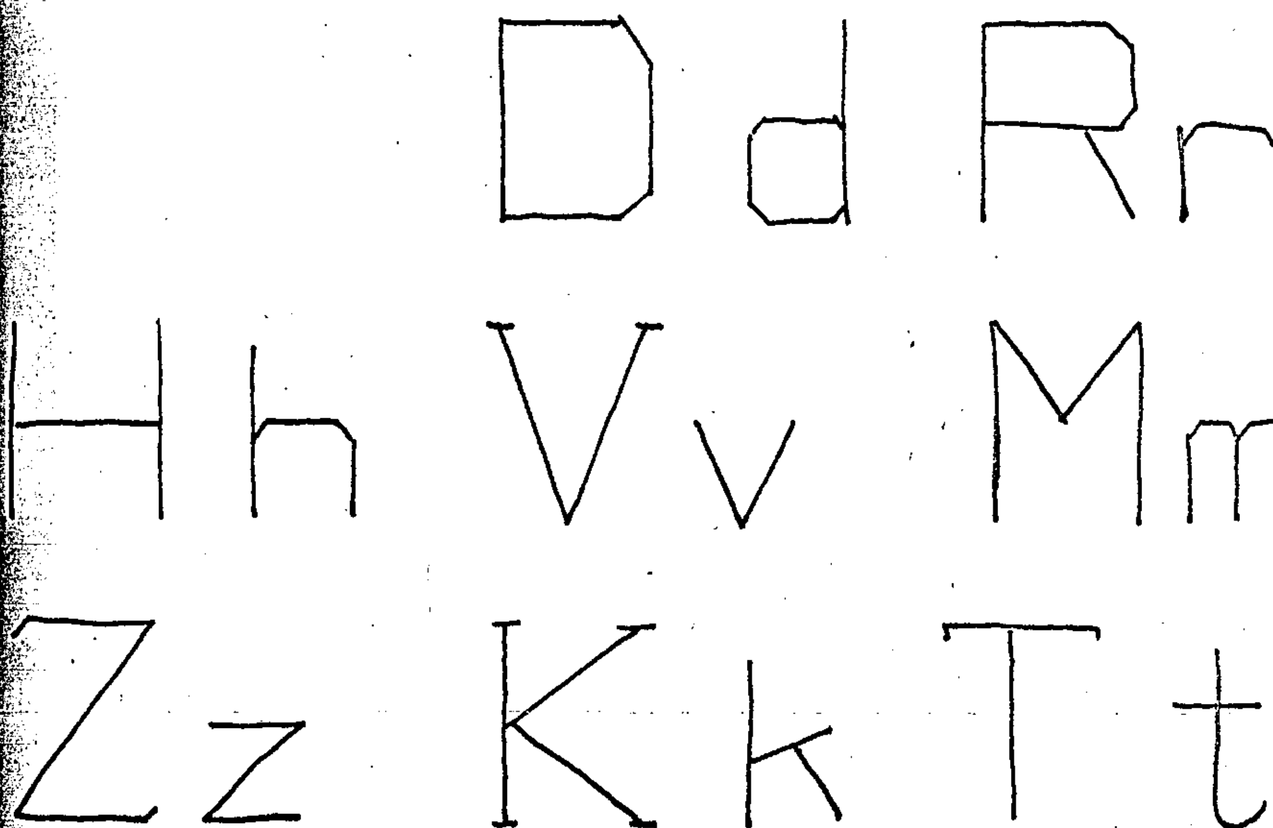


Fig. 2-2.--Designs for the eight stimulus letters appearing on the cathode ray screen. (Actual size is 3/4x 1/2 inches, approximately.)

To complete the coding of the five bits for each character position, the letter code is prefixed by a 0 for lower case and a 1 for upper. Then a 0 is prefixed to indicate low intensity or a 1, for high intensity. Thus, for example, the code 10101 indicates a bright, lower case h. In a similar manner, any one of 2^{40} distinct patterns can be coded into a 40 bit string and presented on the screen for the subject as an eight character array. Figure 2-3 shows how one such array might appear on the screen.

D. <u>Generation of Stimulus Patterns According to Constraints</u>

Since there are 2^{40} unique stimuli possible and only 2^7 are actually used in these experiments, there must be some mapping which chooses this 2^7 and associates them with the values chosen by the earlier subroutine. As before, each bit of the stimulus string can be specified by a logical expression involving each of the seven bits of the stimulus designator. This truly astronomical number of mappings can be substantially reduced without losing any essential flexibility by allowing each bit of the stimulus string to be set equal to one of the bits of the designator string or to its complement. All patterns are possible by this scheme, and it allows different experiments to be set up merely by composing a 5 x 8 table showing which designator bit (or complement) determines which part of each character. A typical table would look like this:

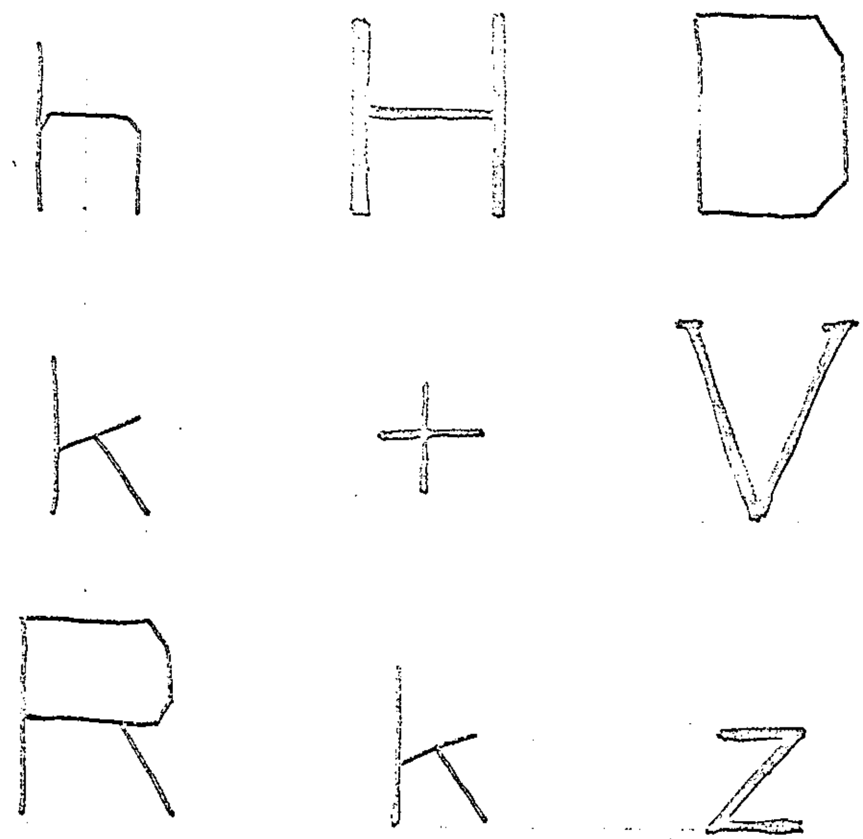


Fig. 2-3.--A stimulus pattern as it appears on the cathode ray screen. (Heavy lines represent bright lines on the screen. The + in the center of the array is always shown and provides both spatial orientation and a brightness reference.)

POSITION	1	2	3	4	5	6	7	8
INTENSITY	A	NA	B	NB	A	NA	3	B
CASE	NB	A	1	NA	B	NC	NA	A
LTRA	A	2	C	NA	C	A	B	N1
LTRB	C	NA	A	4	NA	N1	C	B
LTRC	1	NB	NA	NC	N2	NB	NA	A

LTRA represents the high order of the three bits specifying the letter; LTRB, the middle bit; and LTRC, the low order bit. The positions are numbered as follows:

1	2	3
4	+	5
6	7	8

N before a letter or number designates the complement of that bit.

To illustrate this mapping (which is actually Experiment 1), consider what it does to the string

```
ABC1234
0000000
```

The eight characters produced, in order, are

```
01000
10011
00001
11101
00011
11011
01001
00100
```

which leads to the stimulus array

R	z	m
H	+	z
Z	M	d

The entire set of stimuli for this mapping may be seen in Appendix I.

By using a subroutine to translate such a table into the actual transformations of the mapping, we can change from one experiment to another merely by changing the table. If we wish to have more than seven bits in the stimulus designator, we merely need additional names entered in the table. If we do not wish to have the entire array variable, we can have any entry we wish refer to a fixed bit X, or its complement, NX. Thus, if we wished to work with, say, 15 bit problems, we could fill all the columns beyond 3 with X's. The top row, then, as specified by columns 1, 2, and 3, would be the entire stimulus. Similarly, by filling rows with X's, we can specify problems where certain categories are not involved.

The details of construction of each table naturally depend on the experiment we wish to perform. One possibility is to have the table generated by the computer itself, within certain constraints. In this way, human biases on the part of the experimenter can be reduced. In fact, this technique was actually used so that the experimenter could be a "naive" subject in testing out the system.

A further advantage to computer generation of the constraint tables lies in the possibility of having the computer generate problems based on specific performance characteristics of the subject in previous problems. The problems could either be generated fresh or selected from an inventory of predesigned problems. The compactness of the specification lends itself to storing problems, and the simplicity of the technique means that the computer could change from one problem environment to another without the subject's being aware except through the stimuli themselves (and not, for instance, through some unusual delay or operation of some input device).

For these experiments, however, the constraint specifications were composed by hand in advance. They were constructed so as to permit the observation of performance differences related to the different redundancies of the three components of the response. Five different redundancy distributions were constructed, as shown in the following table:

	A	B	C	1	2	3	4
I	20	10	2	4	2	1	1
II	17	9	6	4	2	1	1
III	11	10	11	4	2	1	1
IV	6	9	17	4	2	1	1
V	2	10	20	4	2	1	1

Experiment 1 is of type II redundancy distribution.

Of course, many different tables could be constructed for a given redundancy distribution. In constructing these experiments, the noise bits were distributed both spatially and through the different categories of stimulus component. Thus, in the table given, each row and each column has at least one noise bit. The minimum number of noise bits required by this scheme, obviously, is eight, so this number was chosen and distributed among the four noise bits as shown.

These experiments were constructed so as to eliminate, as far as possible, the occurrence of "obvious" patterns--such as, a great many of the same letter of all bright letters. As will be seen, one of the experimental results is that it is far from obvious what is "obvious." The experimenter, looking at the array of patterns laid out in its entirety has a different point of view of the experiment than does the subject seeing the patterns sequentially. Furthermore, even in the sequential scanning of the stimuli, "obvious" seems to be a subjective phenomen. It would seem possible to construct, for each subject, sets of problems which would be obvious to him and to nobody else.

E. Test Generation

In the test trial, one or more of the stimulus components can be selectively eliminated, under automatic computer control. If the subject is "using" some of the information in the eliminated portion, his performance can be expected to deteriorate in specific ways. As we shall see, this technique often gives precise information on subject behavior not available through verbal report.

The elements which can be eliminated, of course, are the eight characters in the array. To eliminate a character, the computer merely sets a bit in the display subroutine for its position.

Such simplicity makes it possible to construct elaborate routines for determining which positions shall be tested, as well as when they will be tested. It would also be possible to eliminate characters on ordinary trials--where reinforcement is given--in order to experiment with various training procedures. Thus, for example, if a running analysis determined that the subject is "fixed" on certain components which did not contain the requisite information, the environment could guide him by eliminating those components and forcing him to look elsewhere.

In these experiments, however, only a very simple test procedure was used. After every 100 ordinary trials, a sequence of 32 test trials was run in which each character was elided four times. Then, at the end of the experiment, when the subject had mastered the task, one last set of 32 test trials was run. Since the subjects were getting no reinforcement during the test trials, they sometimes became impatient with them. This difficulty might be eliminated by less frequent trials (either on a fixed interval or by basing the appearance of trials on indications of changes in performance) or by providing some reward for performance in the tests. As it now stands, the subject's willingness to work seriously on the tests is only a matter of his voluntary cooperation with the experimenter. In spite of these difficulties, however, the subjects did generally take the test trials seriously, and they were an excellent source of the expected information.

F. Response Structure

In determining the method by which the subject should indicate his responses to the stimuli, both physical and abstract questions must be considered. Among the physical questions are such matters as subject comfort, fatigue, rough equivalence of responses, and ease of discrimination of responses. In order to get more detail on subject performance than had been obtained in earlier experiments, it was necessary to have more than just two responses--for these do not provide sufficient structure for discriminating between positive and negative recognition. Eight was chosen as a suitably high and convenient number.

The light-pen is a device which, when pointed at a lighted area of the CRT screen, may be made to register its position to the computer by the depression of a button on its side. Although perhaps not as easy to operate as an ordinary pushbutton, tests indicated that it could be used under our experimental conditions without undue fatigue or appreciable rate of error. The particular light pen used was perhaps not adequately maintained, so that occassionally we experienced minor difficulties with the pushbutton sticking in the "down" position. This trouble and an occassional careless pointing of the pen at room lights by the subject were perhaps the major sources of mechanical difficulty in these experiments, and could be remedied by using specially constructed response equipment.

The actual response arrangement is shown in Figure 2-4, surrounding the stimulus array on the CRT screen. The eight

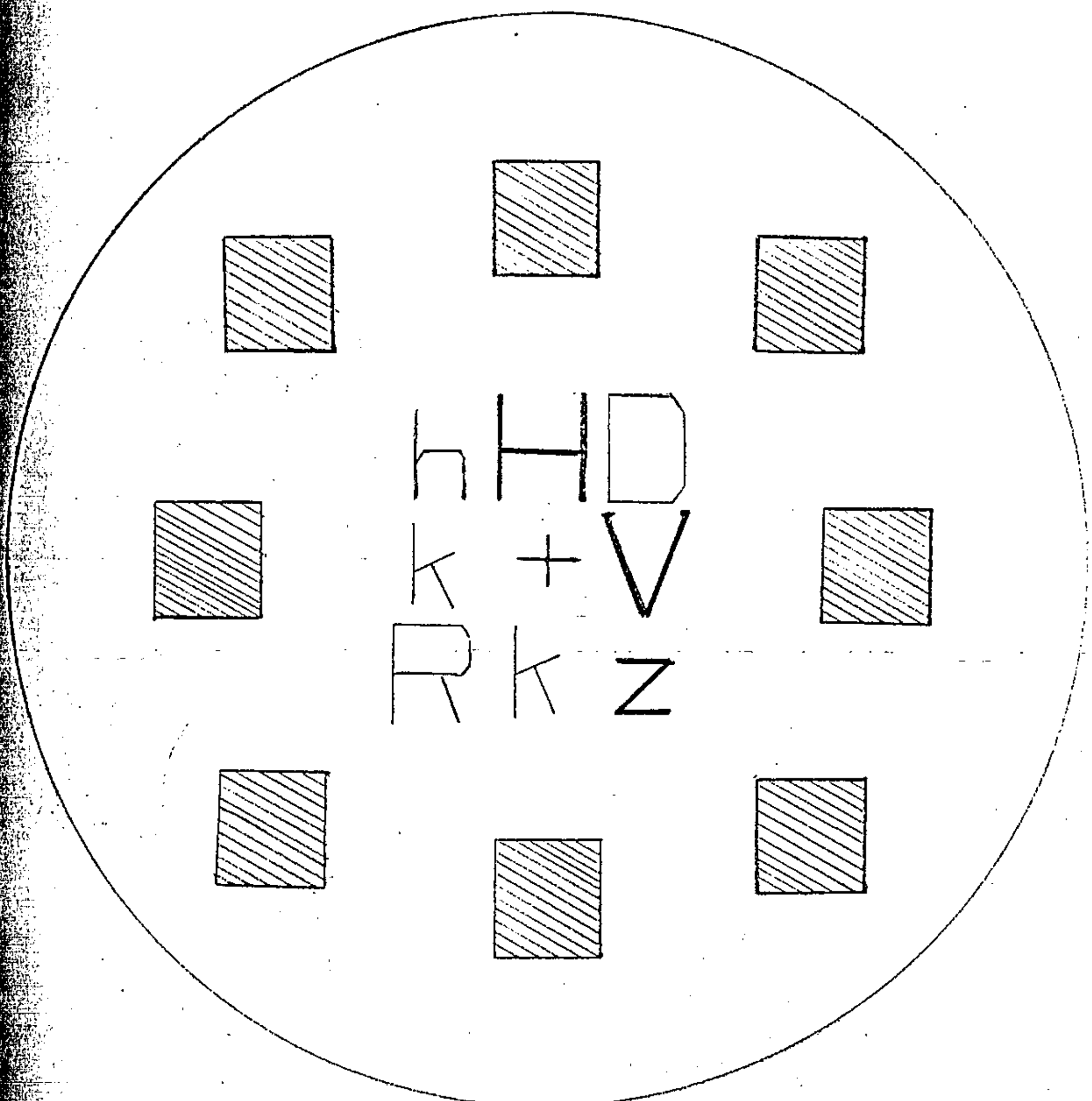


Figure 2-4 : Response Array Surrounding Stimulus Array areas are sufficiently well separated so that the probability of mistaking one for the other--even in a rapid response--is negligbly small. They are lighted to a brightness intermediate to the brightness of the bright and dim letters in the center array. Because both arrays have a common center, the subject need spend only a minimum time switching his attention from one to another. A response is made by pointing the light pen at one of the squares and depressing the button. If the subject points the pen at any other point of the screen when pressing the button, nothing happens. He must, however, release the button within one second (the interval before the next stimulus is presented) in order that a second response not be recorded.

The actual correspondence between the response squares and the values was held constant throughout the experiments. The topmost square corresponds to the value 000, the one to the right, 001, and so on in a clockwise order around the circle, terminating with the square to the left of the topmost square having the value 111. This permanent correspondence imposes a structure on the "response space," and one of the observations we wish to make is the extent to which the subject learns this structure--as opposed to preconceived structures--over the course of several experiments.

G. Feedback

In order for the subject to be able to learn to respond correctly, he must get some information about correctness of his responses. If the response space consisted of only two responses, only one type of feedback is available, namely, right or wrong. When there are more than two responses, however, a variety of different feedback schemes are available, since "not right" is not merely the complement of "right." Many studies indicated that, as might be expected, merely telling the subject "right" or "wrong" at each trial makes his task more difficult than telling him what was the right answer each time. In between these two extreme choices lie a number of interesting schemes which might be the eventual basis for further study.

One method which has frequently been used is "trial until right." Under this scheme, the subject responds and is told whether he is right or wrong. If wrong, he responds again, continuing until a right answer ends the trial. One advantage of this method is that it gives information on the structure of the subjects "confusion matrix," information which has not, however, been used in the experiments where this response technique has been employed. Clearly, it could be put to use in these experiments, given the depth of data analysis available. Its disadvantages, on the other hand, are the variable amount of information recorded on each trial and the inability to pace the experiment because of the great number of responses which might be made on each trial. Consequently, for these experiments, a method was chosen which gives only the first order elements of the confusion matrix, but which gives the subject the maximum information in the minimum time per trial.

Another method which does not appear in the literature, but which is made possible by our apparatus, is to feed back to the subject a measure of the degree of rightness or wrongness at each trial. This degree could be given in several ways. One way--probably most useful when the response space is large--is to tell the subject whether he is getting "warmer" or "colder" on successive attempts during each trial. Another method would be to give him some absolute measure, as might be done in this case by giving the code "distance" between the response and the value. All schemes of this type would give more information about the structure of the response space than do the simpler schemes. Where information on the structure of the response space is not explicitly given, the perception of that structure becomes a matter of individual variability, and is thus one of the factors we pay close attention to in the analysis of our experiments.

For these experiments, the feedback for each trial (except, of course, test trials where there is no feedback) informs the subject of what the "right" response is and reminds him of what his own response was relative to that response. This information is conveyed immediately after the subject indicates his choice of response by brightly illuminating the square he should have chosen. His own choice is also illuminated, but much more dimly so there will be no confusion between the two; and all other response squares are extinguished for the entire feedback period. At the same time, the stimulus pattern remains in the center of the screen for the subject's inspection. Thus, when the subject responds correctly, he sees one square very brightly illuminated; when he responds incorrectly, he sees one bright square and one dim square. When the feedback period is over, all response squares are uniformly illuminated and a new pattern appears in the center of the screen.

H. Data Recording

Perhaps the simplest way to describe the data recording is to refer to Figure 2-5, which is a sample of one "page" of data as printed directly from the magnetic tape created in one experiment. The data from fifty trials constitute a page, at the beginning of which all the indicative information about the experiment is recorded. This format permits readable data to be printed directly from the original tape as an aid to testing the analysis programs and as a protection against possible damage to the tape in handling.

Each line of the page after the headings is a record of a single trial. The trial sequence number is shown on the left, and the response time (in milliseconds) is shown on the right. The column labelled "pattern" records the seven bits of the stimulus pattern designator, although space for twelve bits has been allowed in the format. (The leftmost three bits are A, B, and C; while noise bits 1 through 4 are recorded in positions 6 through 9.) The "value" and "response" columns each allow for five bits, although only the rightmost three are used in these experiments.

EXPERIMENT EXPERIMENT 2
SUBJECT
DATE NOVEMBER 15, 1964

SEQUENCE	PATTERN	TEST	VALUE	RESP	TIME
000001	121112221111	11111111	11121	11121	002018
000002	122112122111	11111111	11122	11122	001948
000003	111112211111	11111111	11111	11211	002394
000004	111111122111	11111111	11111	11212	003023
000005	112112211111	11111111	11112	11111	002830
000006	212111211111	11111111	11212	11121	002726
000007	112111221111	11111111	11112	11111	002483
000008	121112222111	11111111	11121	11221	003432
000009	112111121111	11111111	11112	11111	001912
000010	122112122111	11111111	11122	11122	003063
000011	221112221111	11111111	11221	11122	002402
000012	221112221111	11111111	11221	11221	001725
000013	211112212111	11111111	11211	11121	002070
000014	111111211111	11111111	11111	11111	002285
000015	222111211111	11111111	11222	11211	002919
000016	121111121111	11111111	11121	11222	002505
000017	121112111111	11111111	11121	11121	002322
000018	221112221111	11111111	11221	11111	001901
000019	112111111111	11111111	11112	11111	002134
000020	111112111111	11111111	11111	11111	001953
000021	222112111111	11111111	11222	11221	002945
000022	212112111111	11111111	11212	11121	002128
000023	211112112111	11111111	11211	11212	001758
000024	122111122111	11111111	11122	11121	003344
000025	221111112111	11111111	11221	11212	002660
000026	112112111111	11111111	11112	11111	001684
000027	111111111111	11111111	11111	11111	001763
000028	122111222111	11111111	11122	11121	002518
000029	221112212111	11111111	11221	11221	002540
000030	212112211111	11111111	11212	11121	002300
000031	221112222111	11111111	11221	11221	001888
000032	122111211111	11111111	11122	11212	002459
000033	122112122111	11111111	11122	11122	001783
000034	212112211111	11111111	11212	11212	001843
000035	211111112111	11111111	11211	11212	001459
000036	111112211111	11111111	11111	11111	001762
000037	122111112111	11111111	11122	11212	003585
000038	111111212111	11111111	11111	11111	001510
000039	212111222111	11111111	11212	11121	002663
000040	112112122111	11111111	11112	11111	001584
000041	121111112111	11111111	11121	11121	002810
000042	111111221111	11111111	11111	11111	001618
000043	111112211111	11111111	11111	11111	001466
000044	112112222111	11111111	11112	11111	001462
000045	212111121111	11111111	11212	11212	002460
000046	221112122111	11111111	11221	11122	002310
000047	122111222111	11111111	11122	11212	002466
000048	121111222111	11111111	11121	11212	002583
000049	222112112111	11111111	11222	11122	002728
000050	221112221111	11111111	11221	11222	002459

Figure 2-5: Data listing

The "test" column indicates when the trial is a test trial and which character position or positions was blanked out on that trial. In Figure 2-6, for example, trial number 551, the 2 in the fifth position of the test field indicates that this was a test trial with character position five (middle row, right-hand column) blanked out. Also in Figure 2-6, we can see that the sequence number 999999 is used to terminate the last page of data for each experiment.

Using this format, then, it is possible for the computer to record every piece of data available to it on every trial. If necessary, these data may be supplemented by observations taken by the experimenter during the course of the experiment and by verbal comments of the subject, either taken continuously throughout the experiment or at convenient stopping times between and during the experiments. Because of the exploratory nature of these experiments, it was decided that such supplementary information would be important, so the experimenter was present throughout all of the experiments. Because of the heavy demands placed on the subject by the pacing of the experiment, verbal reporting during trials was not used. Instead, whenever the subject wished to rest--and also at the conclusion of each experiment, an interview with the experimenter was tape recorded and later transcribed for careful study. This transcription, then, also forms a part of the experimental data.

EXPERIMENT EXPERIMENT 1
SUBJECT
DATE NOVEMBER 15, 1964

SEQUENCE	PATTERN	TEST	VALUE	RESP	TIME
000551	121111212111	11112111	11121	11121	002045
000552	211112222111	11111211	11211	11211	002158
000553	111111211111	11111121	11111	11111	001675
000554	221111111111	11111112	11221	11221	003411
000555	111111112111	11111111	11111	11111	001644
999999	111111112111	11111111	11111	11111	001644

Figure 2 - 6: Data listing showing test trials

I. Terminating Procedure

Deciding when to terminate a run is an intrinsically difficult question. First of all, there are two cases which must be considered: terminating upon learning and terminating upon failure to learn. Both cases are difficult because they depend on having some criterion for deciding either that learning has taken place or that learning is not likely ever to take place; yet, these are two of the questions we hoped to answer using the experiments themselves.

Although we cannot say precisely how we would recognize learning of an entire experiment by the subject, we can establish some minimal tests. First of all, we want the subject to be responding correctly at some level significantly above chance. Secondly, we want each of the responses to be included in this performance level--that is, we do not want to be satisfied with a score of 88 per cent if that score is attained by getting seven out of eight of the responses correct and responding randomly to the eighth. In order to meet these tests in as simple a way as possible, the criterion used was a string of consecutive correct responses.

How many responses should such a string contain? If it is too short, it may not contain each of the responses at least once, or it may be achieve through guessing. If it is too long, on the other hand, momentary lapses by the subject may prevent him from <u>ever</u> finishing. Eventually, twenty consecutive correct responses was chosen as the criterion for termination. Subsequent experience with this criterion has shown it to be unsatisfactory in several ways.

The second question, that of unsuccessful termination, was left to the discretion of the experimenter and the subject. Provision was made manually to terminate the experiment if either one felt it would be undesirable to continue.

CHAPTER III

FORMS OF ANALYSIS EMPLOYED

A. The Objectives of Analysis Techniques

In an exploratory study, there are two main objectives which the various data analyses must serve. First, they must highlight events or conditions of possible interest; and second, they must not completely conceal anything. The second objective is rather easily satisfied merely by providing the experimenter with a complete data listing, which is his ultimate recourse when none of the other forms seems to tell what he wants to know. The first objective, on the other hand, lies at the core of the mysterious art of research and is likely to remain only partially satisfied even in the best designed experiments. One of the frightening things about the richness of this set of experiments is the way new insights keep turning up each time the 10,000 or so bits of each experiment are recombined in some new manner. How many times in the history of science have discoveries been left hidden for want of the right point of view, as when Uranus appeared on the photographs of several astronomers before its "discovery?"

In this chapter, the various kinds of analyses performed--both with and without the aid of a computer--will be discussed in order to indicate the kind of search strategy which has been used. Some of that strategy has been fairly conscious--as in the attempt to "symmetrize" the tabulations or to present data pictorially wherever possible. Most, however, has been a matter of trying out things that seemed plausible, seeing the results, pondering them, showing them to others and getting suggestions, waking up in the middle of the night with new ideas, and starting the entire cycle again. The current set of techniques bears little resemblance to the initial set, though here and there some vestiges of discarded ideas remain merely because they were not worth the trouble to remove them from programs.

The computer programming was pervaded by a philosophy of generality. Because so many unanticipated programs had to be written, everything that these routines might have in common was written once and made into a generally available subroutine. Extended to the data, this philosophy meant that every bit of data available to one routine should be available to all, and this was accomplished by compressing the data for each experiment so that a whole experiment would fit into the computer's main memory. Thus, for instance, if a program wanted to correlate the response on a given trial with the value on the seventh previous trial, it was merely a matter of proper indexing, rather than input-output manipulation. Ideally, this idea should be extended so that all <u>output</u> of each program be available to each other program, but this would have necessitated perhaps one hundred times the core storage available on the 7090. As a result, some computations are duplicated, and some--where routines first thought unrelated were later associated in some way--are left to be done manually.

B. Annotated Data Listings

Although the data listing is essential as a last resort and may be obtained directly from the data tape, its usefulness can be enhanced by printing it in a more redundant form. A typical page of the annotated listing appears as Figure 3-1.

As can be seen in the figure, the correct value (V), the actual response (R), and the four noise bits (NOISE) have been printed in their binary form the better to expose their structure. Furthermore, the column labelled V-R has been added to show (by 1's) the bits of the response which were in error. Scanning this column already gives a useful general picture of the changes in the subject's behavior, but the column labelled PICTORIAL is intended to show this in another way. Here, eight columns have been set aside to represent the eight values. For each trial, the value is marked with an X and the response given is marked with a +, except when the response is correct, when an * is used to mark the coincidence. Thus, a somewhat different running picture is gained by scanning these columns, and the experimenter's eye quickly adapts to the method of presentation.

EXPERIMENT EXPERIMENT 1
SUBJECT
DATE NOVEMBER 15, 1964

SEQ	PICTORIAL	V	R	V-R	ANNOTATION	NOISE	TIME
1	X +	0	10	10		11	1531
2	+ X	110	10	100	N	1	3172
3	X +	10	110	100		101	947
4	+ X	100	10	110		1001	950
5	+ X	111	100	11	N	1000	777
6	X +	100	111	11		100	870
7	X+	10	11	1		1111	1011
8	+ X	101	11	110		1000	1058
9	+ X	110	11	101		111	1173
10	X +	1	11	10		101	1071
11	X +	0	11	11	Z	101	1047
12	+X	100	11	111		1011	1112
13	+ X	110	11	101		100	992
14	+X	100	11	111		1110	1014
15	+ X	111	11	100		1011	1127
16	X +	101	111	10	ZP T	1011	1007
17	X +	101	111	10		11	1044
18	X +	10	111	101		1	1188
19	X +	1	111	110		1101	1091
20	X +	0	111	111		1000	1485
21	X +	100	111	11		1010	1204
22	X +	1	111	110		1011	1120
23	X +	100	111	11	V	100	1134
24	X +	0	111	111	V	10	1062
25	X +	1	111	110	P	1110	1051
26	+X	1	0	1	P T	110	1065
27	+X	1	0	1	S	10	1130
28	+ X	10	0	10		1100	1135
29	+ X	111	0	111		1000	990
30	+ X	10	0	10	N R	110	1105
31	*	0	0	0		0	940
32	X.......	0	XXX	XXX	T	1101	'OVER'
33	+ X	111	0	111		1011	171
34	+ X	11	0	11		111	1151
35	X+	1	10	11		1011	1259
36	+ X	110	10	100		1101	1125
37	X +	0	10	10		1110	1161
38	+ X	111	0	111	T	111	1311
39	+X	1	0	1		0	1345
40	X +	10	100	110		11	1303
41	X +	0	100	100		1100	1232
42	X +	10	100	110		1	1196
43	X+	11	100	111		101	1321
44	+ X	110	100	10	P	110	1281
45	+ X	110	100	10		1000	1200
46	X +	10	100	110	N B	1	1142
47	*	100	100	0		11	1265
48	X +	0	100	100		111	1479
49	+ X	111	100	11		100	1624
50	+ X	101	10	111	T	1010	1563
SEQ	PICTORIAL	V	R	V-R	ANNOTATION	NOISE	TIME

MMM	SEQ	PICTORIAL	V	R	V-R	ANNOTATION	NOISE	TIME
MMM	51	+ X	110	10	100	P	1101	1520
MMM	52	+ X	110	10	100		1001	1703
MMM	53	+ X	111	10	101		1110	1451
MMM	54	+X	11	10	1		1011	1282
MMM	55	+ X	100	0	100	T	1001	1136
MMM	56	+ X	101	0	101		1111	1381
MMM	57	+X	11	10	1		0	1403
MMM	58	+ X	111	10	101	P	10	1286
MMM	59	+ X	111	100	11		1100	1596
MMM	60	+ X	110	10	100		110	1369
MMM	61	+X	11	10	1		1011	1581
MMM	62	+ X	111	0	111	ZN R	1011	1378
MMM	63	X +	0	10	10	V	1001	1302
MMM	64	X*	1	10	11	TEST 1	1010	1288
MMM	65	.	0	0	0	TEST 2	110	1268
MMM	66	* X	100	10	110	TEST 3	0	1368
MMM	67	*X	11	10	1	TEST 4	1100	1494
MMM	68	‡X	1	0	1	TEST 5	100	1307
MMM	69	+ X	101	10	111	TEST 6	1001	1419
MMM	70	X*	1	10	11	TEST 7	1111	1140
MMM	71	.	0	0	0	TEST 8	1000	1609
MMM	72	* X	100	10	110	TEST 1	101	1493
MMM	73	‡ X	111	0	111	TEST 2	1011	1185
MMM	74	.	0	0	0	TEST 3	1100	1464
MMM	75	* X	111	10	101	TEST 4	10	1312
MMM	76	* X	101	10	111	TEST 5	0	1463
MMM	77	* X	101	10	111	TEST 6	1001	1541
MMM	78	*X	11	10	1	TEST 7	1100	1535
MMM	79	* X	100	10	110	TEST 8	1001	1585
MMM	80	+ X	110	10	100	TEST 1	1111	1507
MMM	81	*X	11	10	1	TEST 2	1011	2161
MMM	82	*X	11	10	1	TEST 3	1010	1865
MMM	83	*X	11	10	1	TEST 4	100	1627
MMM	84	‡X	1	0	1	TEST 5	11	1441
MMM	85	‡ X	101	0	101	TEST 6	1	1676
MMM	86	Y *	0	10	10	TEST 7	1001	1584
MMM	87	*	10	10	0	TEST 8	11	1573
MMM	88	+X	11	10	1	TEST 1	1000	1701
MMM	89	+ X	100	10	110	TEST 2	0	2002
MMM	90	‡ X	10	0	10	TEST 3	11	1411
MMM	91	+ X	110	10	100	TEST 4	10	2116
MMM	92	+ X	110	10	100	TEST 5	10	1563
MMM	93	*	10	10	0	TEST 6	101	1938
MMM	94	+ X	100	10	110	TEST 7	1111	1272
MMM	95	.	0	0	0	TEST 8	100	1540
MMM	96	+ X	110	10	100		0	1156
MMM	97	.	0	0	0		1	1760
MMM	98	* X	101	10	111		1111	1632
MMM	99	*X	11	10	1		110	1643
MMM	100	+ X	111	100	11	R	11	2294
MMM	SEQ	PICTORIAL	V	R	V-R	ANNOTATION	NOISE	TIME

Figure 3-1

In order to facilitate scanning yet provide a check to the scan when the test trials are met, the test trials are offset one column to the right. The response times (in milliseconds) are listed on the right, readily available yet out of the way of visual scanning.

The ANNOTATION columns provide a space for various occurences to be noted, which the experimenter would otherwise have to detect by tedious manual work. The letter Z, for instance, indicates that two stimuli in a row have the same noise bits, while the letter I indicates that the entire stimulus has been repeated. Two other structures which were tagged as potentially interesting are sequences of 3 or more identical values (S) or strings of three or more identical responses terminated by a different response (T).

In order to investigate the influence of "events in the small" on the learning process, a number of other notations were added to the data listing. Whenever a trial has an incorrect response, but the next time that value comes up it is correctly recognized, the trial is marked with a V. If the trial is wrong but the next time the same response is made it is made correctly, the trial is marked with an R. If both R and V notations are appropriate, the trial is marked with a B. The purpose of this notation was to draw attention to possible points at which behavior changed, in the hopes of identifying structures which were frequently coincident with them.

In the course of examining the R, V, and B trials, it was noticed that they were frequently of one of the following two forms:

1. Trial t - 1 repeated the value that had been missed on trial t.

2. Trial t - 1 presented as a value the response that the subject had given on trial t.

Because of the demands these experiments place on the subject's memory, any structure that helps him with his memory task is a potential point of learning. Form 1 represents possible assistance which might be expressed verbally as "No, try another one of those." Form 2 might be expressed verbally as "No, that's not an X, <u>this</u> is an X." Form 1 was designated as a <u>positive</u> aid (P), while form 2 was designated <u>negative</u> (N). With this notation, the data listing may be examined for correlates of these events, particularly the R, V, and B type events.

Events of types R, V, B, P and N would only be significant to a subject if his strategy were based on separating the individual responses and concentrating on some subset of them. Furthermore, if he is using such a "response separation" strategy, it will be meaningful to ask when he "learns" each response. In fact, the extent to which such a question turns out to be meaningful is a measure of the extent to which the subject is actually using a response separation strategy. We should like, therefore, to be able to identify--as closely as possible--the point at which the subject "captures" each response, both for its interest as an "event" and for its influence on the interpretation of the remaining trials. After capture of a response, X, events of types R, V, B, P, and N which involve X are not to be interpreted in the same way as before. In other words, if the capture is truly identified, further errors involving X must be interpreted as lapses or momentary confusions--unless they are persistent enough to say that the subject actually forgot X after having learned it. Therefore, we eliminate R, V, B, P, and N notations involving X for all trials after X has been captured.

The term "captured" rather than the term "learned" has been used, because a capture seems to involve two kinds of learning, which might be termed positive and negative (not to be confused with P and N type events). Positive learning of X occurs when the subject always recognizes X when it appears; that is, when X is the stimulus, he responds X. Negative learning occurs when the subject never responds X incorrectly, that is, he never makes the response X to a stimulus in class $Y \neq X$. If we identify the points of positive and negative learning, no correlations of interest seem to occur. Further investigation reveals that the trouble may lie in the inability properly to identify such points, because of the particular strategy adopted by the subject. For example, in the extreme case when he decides to respond X to <u>every</u> stimulus, these criteria would lead us to conclude that he had positively learned X and negatively learned all other responses. Although this case never actually happened, the response biases shown in virtually every experiment cause at least some difficulty in locating meaningful learning points.

In order to obtain a more consistent identification, a capture criterion was made such that capture was said to occur at the last error before

1. Positive learning had taken place,

and

2. Negative learning had taken place.

This criterion was deemed met when the subject responded X correctly four times in a row without, over the interval between the last error and the last of the four, either failing to respond correctly to an X stimulus or responding X to a $Y \neq X$ stimulus. Although this measure did seem more meaningful, attempts to apply it mechanically revealed one flaw. Because the sequence of values is generated at random, it often occurs that four X's come up in rather close succession. If, for instance, the subject had been unable to master the distinction between X and some other value, Z, four X's might come up without an intervening Z. In that case, the criterion misses by not being sufficiently critical.

In order to correct this flaw, one additional measure is added to the test, namely, that in the interval between the first and last of the four X's, each value must appear at least once. If not, the test must be extended to five, six, or more X's, until the subject has had an opportunity to display his knowledge of each of the values at least once. As we shall see, this objective measure, though crude enough to be applied mechanically helps to identify points of real significance in the experiments.

In order to aid meaningful interpretation, then, after a response has been captured notations relating to it are deleted. Furthermore, its representation in the PICTORIAL columns is changed so as to distinguish the behavior of already captured responses from those being actively pursued. A correct, captured response is denoted by (.), an incorrect one by (≠), and a missed, captured value by (Y), as shown in Figure 3-1, in which value 000 was captured at trial 63.

In addition to the basic annotated data listing, other listings may be useful for particular situations. Typical of these is the special listing shown in Figure 3-2, in which one particular value is selected from the others, in order to show in its entire history as an aid to analysis of strategy and decision structure. The listing is divided into three panels of identical format: the one on the left gives those trials on which the particular stimulus (in this case, 4, or 100) was responded to correctly; the center panel gives the trials on which it was responded to incorrectly; and the right panel gives those trials on which the response was given to an incorrect stimulus. Each panel tells the trial number, the response given (left and center) or value seen (right), the noise bits seen, the response time, and the number of the position deleted if the trial was a test. By following the course of each of the responses on these listings, many facts become apparent. For instance, in Figure 3-2, we can see that position 6 is quite important in the identification of 100 by looking at the errors on trials 194 and 202 as well as the greatly increased response time on trial 186.

EXPERIMENT EXPER NO 4A
SUBJECT
DATE DEC 20, 1964
HISTORY OF RESPONSE 4

SEQ	R.V	NOIS	TIME	TEST
17	100-	1101	1550	(0)
39	100-	1	1989	(0)
57	100-	11	1747	(0)
58	100-	1011	1888	(0)
75	100-	1010	2453	(0)
76	100-	10	2056	(0)
85	100-	100	1948	(0)
100	100-	1101	2885	(0)
102	100-	111	1956	(0)
104	100-	1000	2322	(0)
106	100-	110	1915	(0)
113	100-	11	1818	(0)
129	100-	1101	3912	(2)
135	100-	1111	2259	(8)
136	100-	100	1931	(1)
139	100-	110	2050	(4)
151	100-	100	2385	(8)
160	100-	10	2011	(0)
176	100-	10	1889	(4)
183	100-	1111	1424	(3)
186	100-	1101	3827	(6)
200	100-	1001	2023	(4)

SEQ	R.V	NOIS	TIME	TEST
27	101-	1001	3682	(0)
84	11-	1110	4451	(0)
128	110-	100	3089	(1)
194	1-	1001	4196	(6)
202	1-	100	4225	(6)

SEQ	R.V	NOIS	TIME	TEST
1	101-	10	4067	(0)
2	11-	1011	2954	(0)
4	11-	1111	1399	(0)
6	11-	1011	2202	(0)
11	1-	110	2922	(0)
13	1-	10	1496	(0)
14	11-	1011	1692	(0)
16	110-	1110	2462	(0)
23	1-	0	2257	(0)
30	1-	1000	2606	(0)
44	1-	1011	3913	(0)
189	1-	111	4341	(1)
196	11-	100	3148	(8)

Figure 3-2

C. Determining Response Times

Response time on any given trial can indicate many things. In order to make such inferences, however, it is usually necessary to compare the actual response time with other response times on nearby trials and with other response times to the same or similar stimuli or for giving the same or similar responses.

There are a number of problems associated with extracting a representative response time from the actual data. After much trial, the following method was adopted to find the time for responding to each value:

1. First, from all the response times to that value, the "no response" trials and trials with times so short that they must represent an accidental button pushing are eliminated.

2. This list of response times is then smoothed into a list with equally spaced members, each smoothed point being a sum of its eight nearest data points--weighted in inverse relation to distance from the smoothed point.

3. The standard deviation of this smoothed points is calculated, and the raw data is culled again to eliminate points which are more than two standard deviations away from their local smoothed point.

4. Steps 2 and 3 are then repeated, using the new culled list each time. A total of four smoothings has been found adequate.

This method tends to eliminate maxima and minima created by single extreme points, leaving a curve more representative of trends over a span of time. The points which are eliminated, of course, may be extremely interesting individual points, or they may be spurious points caused by inattention or accident. In any case, such points can be quickly spotted by comparing the data listing with the smoothed curves, or by printing a data listing in which the times are expressed as deviations from the smoothed curves.

From the eight individual smoothed curves, we can obtain a more realistic overall average time, since effects due to a local abundance or shortage of a particular value are eliminated. Figure 3-3 shows a rather typical overall response time curve--starting as it does with a rather fast response, then slowing down until a peak is reached, then steadily speeding up as the task is mastered. We shall be particularly interested in the location of maxima and minima of this curve, as well as of the response times to individual values, a typical one of which is shown in Figure 3-4. (Both curves are plotted in terms of trial numbers rather than sequence numbers, as the test trials have been excluded from the response time calculations.)

It should be remarked that the curves produced are not always so smooth. In some cases, a large percentage of the points are discarded by the smoothing--yet no smooth curve results. Investigation of these cases usually reveals that the subject has two different rules for identifying the same response, and we are attempting to smooth two curves into one.

D. Tabulations

When approaching the data for the first time, it is useful to have some general figures about its characteristics. One of

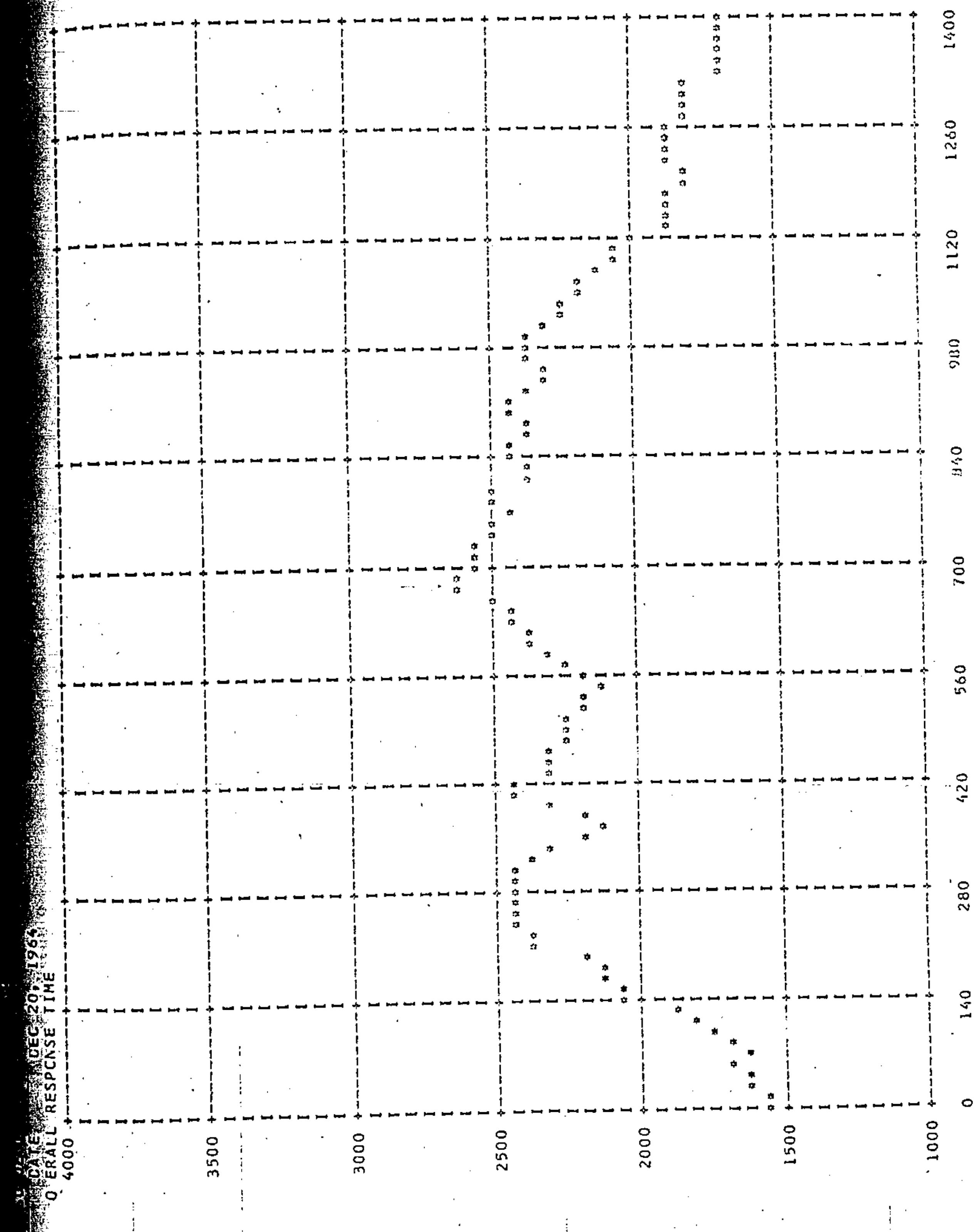

O'ERALL RESPCNSE TIME
RIAL NUMBER
0
140
280
420
560
700
840
980
1120
1260
1400
4000
3500
3000
2500
2000
1500
1000

Figure 3-3

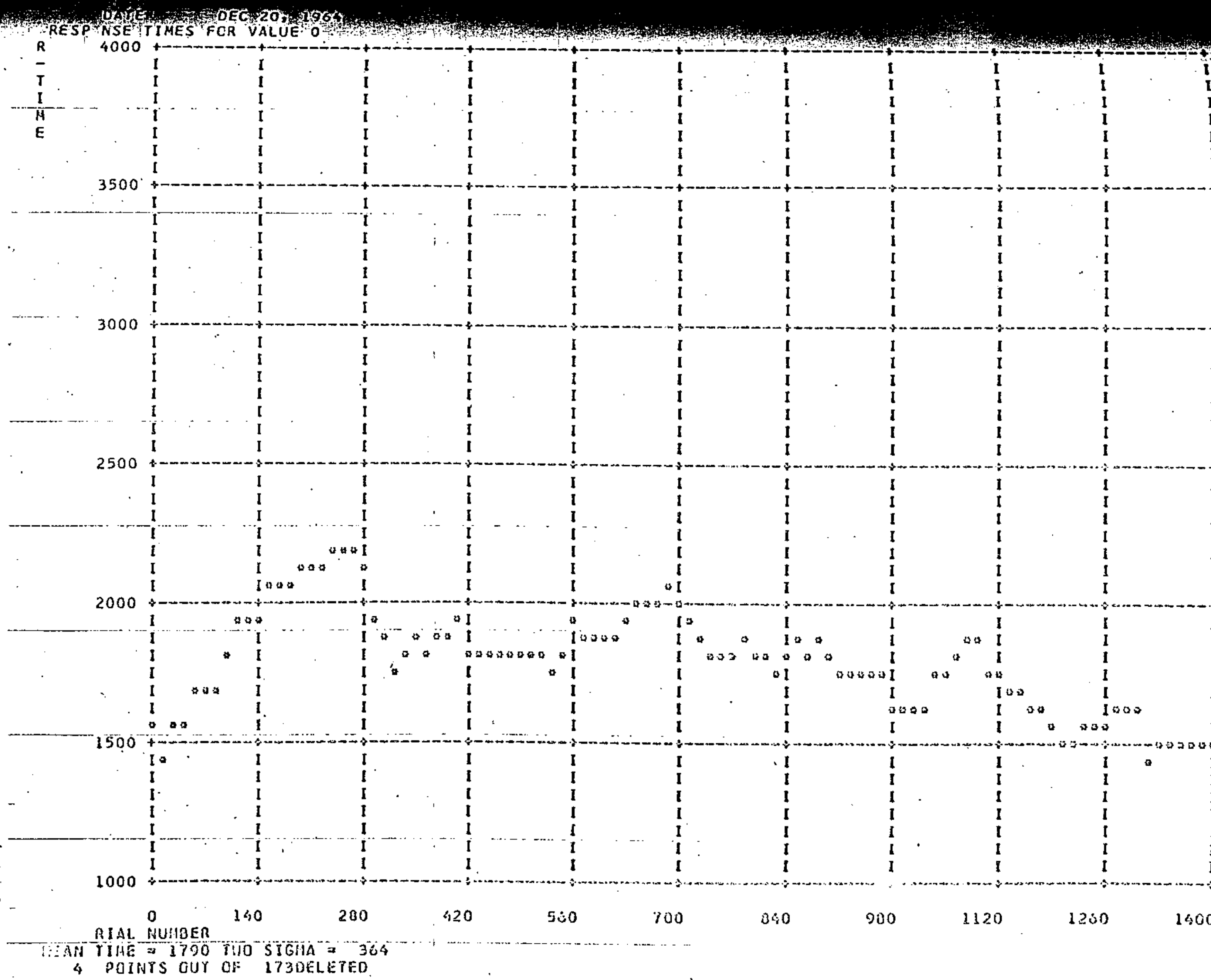

DATE DEC 20, 1964
RESP NSE TIMES FOR VALUE 0
R-TIME
4000
3500
3000
2500
2000
1500
1000
0
140
280
420
560
700
840
900
1120
1260
1400
RIAL NUMBER
MEAN TIME = 1790 TWO SIGMA = 364
4 POINTS OUT OF 173DELETED

Figure 3-4

the simplest ways to reduce thousands of numbers to just a few is to tabulate and take averages. Slightly more detail can be obtained by breaking the data into chunks and averaging and tabulating over each chunk. In order to facilitate such analysis, a general tabulating system was developed so that different tabulations could be made over varying intervals as well as in total.

The tabulating system is based on a pair of 10 x 10 matrixes, which really contain 8 x 8 matrices--8 being the number of responses. A ninth row and column are added to account for the cases where no response is given within the alloted time, and a tenth row and column for crossfooted column and row totals is provided to round out the 10 x 10 size. A pair of such matrixes is shown in Figure 3-5. As shown by the R*V heading, the rows represent the different responses and the columns, the values, for each trial. N heads the "no response" row and column, while T stands for the totals. The left hand matrix tabulates the number of times response R was given to value V, while the right hand matrix displays the average time (truncated to 100ths of a second) associated with the events tabulated in the corresponding position in the left-hand matrix.

For example, response 1 (actually 000) was given correctly (that is, to value 1) 40 times, and the average time for giving it correctly was 1.62 seconds. Response 6 (101), on the other hand, was given correctly only 9 times, averaging 2.64 seconds per response. Presumably, this difference indicates that 1 was learned well before 6, and, because of the response time differences, is also much easier for the subject to discriminate. We can also

EXPERIMENT FXPERIMENT 1
SUBJECT
DATE NOVEMBER 15, 1964

R*V

TOTALS

	1	2	3	4	5	6	7	8	N	T
1I	40*	3	2	2	1*	1-	0	4 I	0I	53I
2I	0	38*	4	2=	6	3	0	0 I	0I	53I
3I	2=	2	14*	8	4	12+	13+	5 I	0I	60I
4I	1*	1-	2	29*	6	5	8	1-I	0I	53I
5I	2	2	5	1*	21*	5	8	8 I	0I	52I
6I	0	0	7	6	7	9	13‡	2=I	0I	44I
7I	0	2=	4	5	0	4	13‡	2 I	0I	30I
8I	2	4	1*	1	3	4	2-	29*I	0I	46I
NI	1*	0	1	1	0	0	0	0 I	0I	3I
TI	48	52	40=	55	48	43	57‡	51 I	0I	394I

MEAN ELEMENT = 6MEAN DIAGONAL = 24

ERAGE TIMES

	1	2	3	4	5	6	7	8	N	T
1I	162+	118	112-	213‡	113	138-	0	96*I	0I	153I
2I	0	199	189	226‡	181	176=	0	0 I	0I	196I
3I	123=	150	265*	203	185	217	221	199 I	0I	218I
4I	104*	107-	174	233‡	181	155	197	112 I	0I	205I
5I	135	191	185	132*	207	258‡	234	189 I	0I	207I
6I	0	0	256	251+	222*	264*	256+	264+I	0I	252I
7I	0	217*	239	220	0	237	255‡	227 I	0I	240I
8I	127	126	118	219‡	106*	189	186-	199 I	0I	181I
NI	500*	0	500	500	0	0	0	0 I	0I	500I
TI	164=	185	233	231	193	219	234‡	191 I	0I	206I

Figure 3-5

observe that response 6 was actually given 13 times to value 7, indicating one of the sources of difficulty may have been confusion between the two. Similarly, response 3 was given 13 times to value 6, indicating another source of confusion. By using a set of cutout templates, we can even break down these confusions according to the bits A, B, or C, that they represent.

The total rows and columns give us additional information. For instance, e_{TT} tells us that there were 394 informational trials (test trials are not included) and that the overall mean response time was 2.06 seconds. The average response times for each row and column gives us an impression of the relative difficulty of each response--in this case, for example, responses 1, 2, and 8, seem to have been the easiest, followed by 3, 4, and 5. The tabulations seem to verify this impression.

Figure 3-6 shows another pair of matrixes from the same experiment, this one representing a partial tabulation covering the second 100 informational trials. The excess of the mean diagonal element over the mean element gives us an immediate impression of the degree of learning which has already taken place by this time. Examining the left-hand matrix, we see that this learning seems largely accounted for by 1 and 2, and to some extent, 8, thus confirming our impression from the overall tabulations, and permitting us to focus our examination of the data listing--if we wish--on just those areas where things are happening. The T column shows a considerable response bias, giving us some clue as to the subject's

XPERIMENT EXPERIMENT 1
SUBJECT
DATE NOVEMBER 15, 1964

V

OTALS

	1	2	3	4	5	6	7	8	N	T
1I	13*	0	0	1*	0	0	0	0 I	0I	14I
2I	0	12*	2+	1=	5+	2	0	0 I	0I	22I
3I	0	0	2‡	1=	1-	2	2-	1-I	0I	9I
4I	0	0	0	3‡	1=	3+	2	0 I	0I	9I
5I	0	2-	0	0	1=	0	0	3‡I	0I	6I
6I	0	0	2	6*	5	1*	5+	2 I	0I	21I
7I	0	0	1*	2‡	0	2	2	0 I	0I	7I
8I	0	0	0	0	0	1=	2	7*I	0I	10I
NI	0	0	1*	1	0	0	0	0 I	0I	2I
TI	13	14	8=	15‡	13	11	13	13 I	0I	100I

MEAN ELEMENT = 2MEAN DIAGONAL = 5

AGE TIMES

	1	2	3	4	5	6	7	8	N	T
1I	136*	0	0	311*	0	0	0	0 I	0I	149I
2I	0	210*	150*	195-	173-	179	0	0 I	0I	193I
3I	0	0	246+	284	233	182=	229	306*I	0I	238I
4I	0	0	0	237	278*	177=	199	0 I	0I	213I
5I	0	191*	0	0	197	0	0	217‡I	0I	205I
6I	0	0	224	251	212	144*	236	264‡I	0I	232I
7I	0	0	243	228=	0	259‡	238+	0 I	0I	242I
8I	0	0	0	0	0	267*	186*	190-I	0I	197I
NI	0	0	500*	500	0	0	0	0 I	0I	500I
TI	136=	207	248	264‡	203	198	222	217 I	0I	211I

Figure 3-6

strategy which we can confirm by examining the data listing. Notice, incidentally, that the off-diagonal zeros in column 1 indicate positive learning of response 1, while $e_{14} = 1$ indicates that perhaps negative learning occurred little later. Column 2 compared with row 2 indicates that positive learning of 2 is well ahead of negative learning, while the opposite may be true for response 8.

Sometimes, the largest smallest element in a row or column can be especially significant. Therefore, a notation has been added to signal such pivotal elements: + meaning column maximum; ≠, row maximum; -, column minimum (other than zero); and =, row minimum (other than zero). Where a peak, valley, or saddle point exists, it is marked with an *.

Other tabulations can be made using the same set of tabulating routines. Figure 3-7, for example, shows the same 100 trials tabulated according to response vs. previous response (R*PR), and also vs. the previous v lue (R*PV). The top tabulation is intended to reveal any sequential bias in the subject's strategy. Does he, for instance, tend to keep trying the same response when he has no idea what to do (creating large diagonal elements); or does he have a tendency to follow a more complex response pattern (such as following response 2 by response 6 which might be indicated by $e_{62} = 6$)? Of course, as learning progresses, the subject's freedom to choose responses arbitrarily is reduced, so these tabulations are really of value only in the early trials. This observation holds also for the lower tabulations, which attempt to reveal

EXPERIMENT EXPERIMENT I
SUBJECT
DATE NOVEMBER 15, 1964

R*PR

TOTALS

	1	2	3	4	5	6	7	8	N	T
1I	3‡	3	2+	2	0	3	1*	0 I	0I	14I
2I	4*	2=	2	3+	2+	3	2+	4+I	0I	22I
3I	1*	3‡	2	0	0	1-	0	0 I	2I	9I
4I	1=	3‡	0	1-	0	2	1	1-I	0I	9I
5I	1*	1-	1-	0	1-	1	1	0 I	0I	5I
6I	2	6*	1=	2	2	6+	0	2 I	0I	21I
7I	1=	2‡	0	0	0	2	1	1 I	0I	7I
8I	0	2	0	1=	1	3‡	1	2 I	0I	10I
NI	1*	0	1	0	0	0	0	0 I	0I	2I
TI	14	22‡	9	9	6=	21	7	10 I	2I	100I

AVERAGE TIMES

	1	2	3	4	5	6	7	8	N	T
1I	139-	194‡	126-	109*	0	141-	134	0 I	0I	149I
2I	225‡	187-	173	219+	161*	167	207	191 I	0I	193I
3I	202=	249	263‡	0	0	256+	0	0 I	203I	239I
4I	287*	237	0	160=	0	172	181	235 I	0I	213I
5I	168=	197	279*	0	213	183	190	0 I	0I	205I
6I	264	265*	213	163=	264+	201	0	238 I	0I	232I
7I	243	240	0	0	0	228=	246+	266*I	0I	242I
8I	0	261‡	0	156	216	189	151*	180-I	0I	197I
NI	500*	0	500	0	0	0	0	0 I	0I	500I
TI	232	237‡	235	163=	213	187	197	206 I	203I	211I

MEAN ELEMENT = 2MEAN DIAGONAL = 2

R*PV

TOTALS

	1	2	3	4	5	6	7	8	N	T
1I	3*	0	1*	2	3+	2	2	1-I	0I	14I
2I	3	1*	3+	1-	2	3+	3+	6*I	0I	22I
3I	1*	1	2‡	2	2	C	0	1 I	0I	9I
4I	1=	1	0	5*	0	1-	0	1 I	0I	9I
5I	1=	0	0	0	2‡	2	1-	0 I	0I	6I
6I	2	7*	0	3	1*	3	3	2 I	0I	21I
7I	1=	2‡	1	0	1	0	2	0 I	0I	7I
8I	0	1=	0	2	3‡	0	2	2 I	0I	10I
NI	1*	0	1	0	0	0	0	0 I	0I	2I
TI	13	13	8=	15‡	14	11	13	13 I	0I	100I

AVERAGE TIMES

	1	2	3	4	5	6	7	8	N	T
1I	139-	0	169	158-	194*	121-	114*	131-I	0I	149I
2I	218	236-	159*	245‡	218	222	159	174 I	0I	193I
3I	202	306*	195=	246+	232	0	0	284+I	0I	238I
4I	287*	278	0	176=	0	238+	0	235 I	0I	213I
5I	168=	0	0	0	205	231‡	190	0 I	0I	203I
6I	264‡	259	0	180=	206	221	210	244 I	0I	232I
7I	243	240	246+	0	248*	0	238*	0 I	0I	242I
8I	0	267‡	0	186	201	0	145=	219 I	0I	197I
NI	500*	0	500	0	0	0	0	0 I	0I	500I
TI	231	260‡	223	190	211	207	176=	202 I	0I	211I

MEAN ELEMENT = 2MEAN DIAGONAL = 3

Figure 3-7

any tendency for previous values to influence behavior. Here, for instance, the slightly larger diagonal elements might indicate a "following" tendency. Of particular interest here are the response times, since there seems no *a priori* reason that a subject should respond .84 seconds faster after a 7 than after a 2, for instance. If this tendency is shown in other intervals, there may be some significance to it.

E. Correlations

One of the subject's most critical tasks is picking relevant features out of the stimulus. Presumably, he does this by trying one or more features as determinants of his response and then accepting or rejecting them as relevant based on his success in using them. It is possible, however, that the subject uses a response strategy which is more or less independent of his hypothesis testing strategy. If this be the case, we will have little indication of what he is doing from an examination of the data. If clear signs can be found that he is reflecting his thinking in his behavior, they will support the intuitive hypothesis that behavior and thought development are linked. Once this has been substantiated, we would hope to be able to find out more about the structure of this process of searching out relevance by examining the subject's responses before he has completely mastered each, or any, response.

Although forty bits are displayed in the pattern, they are really a redundant representation of the seven stimulus selection bits. If the subject is basing his current responses on selected stimulus features, some of those seven bits should show correlations with some of the three bits of his response. By plotting running correlations on the twenty-one combinations of response bit with stimulus bit, we may be able partially to reconstruct the hypothesis testing strategy the subject used.

Figure 3-8 shows a typical correlation plot, relating the three response bits to noise bit 3 over the course of the experiment. Each point represents a correlation taken over an interval centered at that point. The interval is chosen so as to compromise between the smooth curve a long interval will give and the more localized picture which we can get by choosing a smaller interval. By trial, the following rule for choosing the interval was obtained:

1. Divide the total number of trials by 100 and add one to the quotient. This gives the interval which will be indistinguishable on a plot with 100 horizontal divisions. In this example, 1561 trials gives an interval of 16.

2. A minimum of 7 such intervals will be used, 3 on each side of the computed point as well as the central interval.

3. If the experiment is so short that this technique will not yield at least 50 points, increase the number of intervals in steps of two until at least 50 points will be used. In our example, 7 x 16 = 112, so no extension is needed. Thus, each point shown is related to the three points on either side of it.

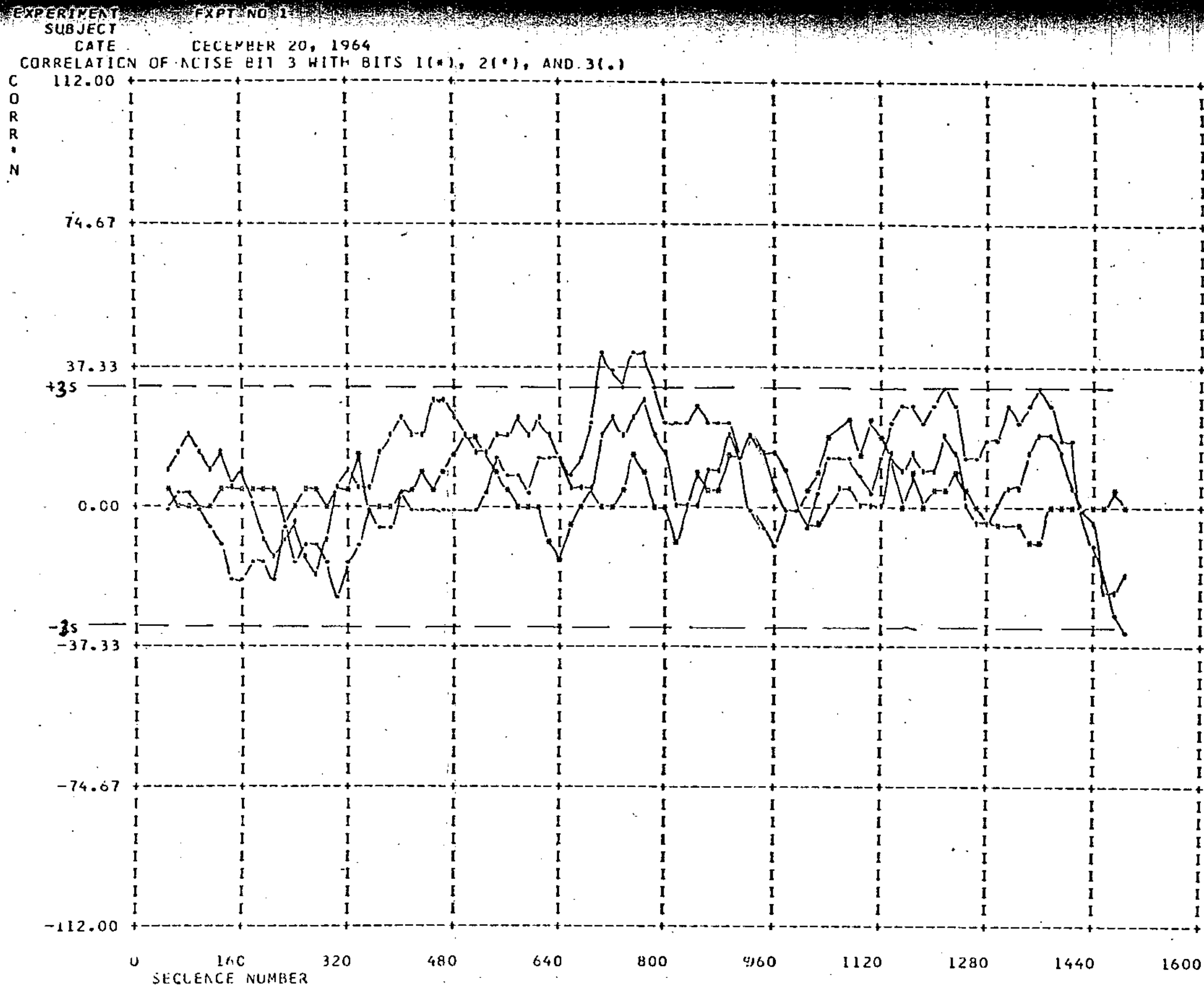

EXPERIMENT FXPT NO 1
SUBJECT
DATE DECEMBER 20, 1964
CORRELATION OF NOISE BIT 3 WITH BITS 1(*), 2(*), AND 3(.)
CORR. N
112.00
74.67
37.33
+3s
0.00
-3s
-37.33
-74.67
-112.00
0
160
320
480
640
800
960
1120
1280
1440
1600
SEQUENCE NUMBER

Figure 3-8

Since the individual components of the stimulus and the response each have only two values, we can compute the correlation in the following way:

1. Let $z_1 = \partial_{ab}$, where a is the response bit, b is the stimulus bit, and ∂ is the Kronecker delta.
2. Then $C = 2\sum_i z_i - N$, where N is the size of the interval and the sum ranges over all trials in the interval.

For random responses, C should be near zero. If a and b are always identical, C = +N; and if a and b are always opposite, C = -N. Since the physical manifestations of the bit values in the stimulus are essentially arbitrary, negative and positive values of C are equally significant.

We can calculate just how significant any particular value of C is by relating C to the outcomes of a series of Bernoulli trials in the following way:

1. Let $C' = C/N$, so $1 > C' > -1$
2. Let $C'' = (C' + 1)/2$, so $1 \geq C'' \geq 0$
3. But C" is the mean of N Bernoulli trials, with outcomes 1 and 0. If the responses are random, then $p = q = 1/2$ for these trials.
4. Approximating these trials using the normal approximation, we get $s'' = \sqrt{pq/N} = 1/2\sqrt{N}$.
5. Reversing our transformation, we get $s = 2Ns'' = \sqrt{N}$.

Thus, retaining the normal approximation, we can expect about one deviation of about 3s from the mean (which is zero) in every set of one hundred points, which is just the number we have on the graph.

For our example, $N = 112$, and $\sqrt{N} = 10.6$. 3s is therefore about 32, and a line is drawn at ± 32 on the graph. We notice immediately that the correlation with response bit 3 reaches or exceeds these lines 9 times, so we may reasonably conclude that there is some significant correlation of these bits. Another way of spotting significant correlation is by considering the probability of the curve staying on the same side of the mean for m consecutive points. Because the points do not represent independent samples of trials, it is not easy to compute significance levels; but it is easy to make trial calculations using random data. In such random data, an average sequence of correlations reaches or crosses zero 18 or 19 times. Furthermore, a run of as many as 15 consecutive points on the same side of zero occurs less than once in every three sequences. In our example, the correlation with bit 3 reaches or crosses zero only nine times and that with bit 2, ten times. Furthermore, the bit 3 correlation has one-sided runs of 15, 25, and 25; while that of bit 2 has one run of 31 consecutive positive points. But 1, on the other hand, presents a curve which is quite typical of random data.

Our graph, then, seems to present the following picture: The subject--either consciously or unconsciously--is letting one or more manifestations of this bit influence his response. Moreover, he persists in returning to this bit, even though it is not related to the correct values, throughout the experiment. In addition, when he once "fixes" on it, it takes him an inordinate amount of time to get rid of its influence, far more than enough time to accumulate substantial contradictory evidence.

In order to construct a more specific picture, we have to relate this graph to other evidence. In this case, it is particularly interesting to note that this bit--noise bit 3--is represented only once in the stimulus--as the intensity of character 7 (see Appendix I). Thus, any correlation with this bit must be a correlation with that specific feature of the stimulus. Among other things, this example shows that it is quite possible for the subject to isolate single features of the stimulus, among the forty features presented to him. In the case of other stimulus bits, the correlation curves may not give such unambiguous information; but often the results of test trials narrow down the key item by pointing to the character position it occupies.

When the correlation is with one of the significant stimulus bits, we get additional information from the curves. For instance in Figure 3-9 (which is from the same experiment as Figure 3-8) we see the curve for bit 1 finally emerge from the other curves between trials 640 and 800. The sharpness of this rise--especially in view of the smoothing which is applied--permits us to localize the attainment of this partial concept quite precisely. Further examination of the curves shows evidence that the subject had

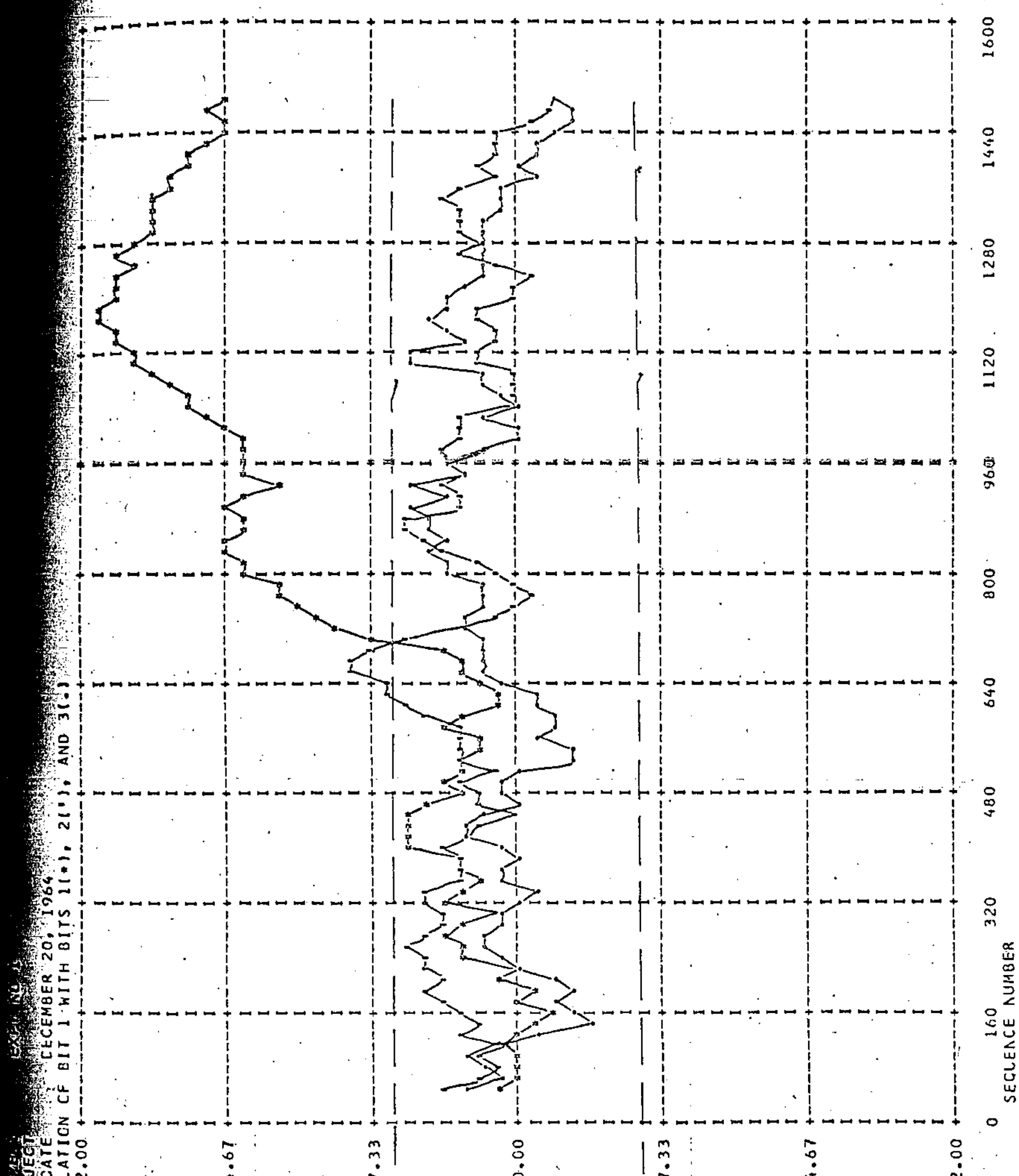

SUBJECT
DATE : DECEMBER 20, 1964
CORRELATION OF BIT 1 WITH BITS 1(*), 2('), AND 3(.)
CORR. N
112.00
74.67
37.33
+3S
0.00
-3S
-37.33
-74.67
-112.00
0
160
320
480
640
800
960
1120
1280
1440
1600
SEQUENCE NUMBER

Figure 3-9

partially (perhaps unconsciously) picked up the significance of bit 1 much before this emergence--for starting at trial 244, the bit 1 correlation is never again negative or zero. Moreover, there is a good indication that the subject was also entertaining some sort of hypothesis which related bit 1 of the stimulus to bit 2 of his response. In fact, just before the emergence of the correct hypothesis, bit 2 shows its strongest correlation, as if it were receiving an explicit test. All of these features are quite typical of a number of the correlation curves for different experiments.

One other rather typical feature can also be seen in this example. After bit 1 has been very strongly separated, bits 2 and 3 of the response still seem to remain somewhat correlated with it--especially bit 3. When we look at the other graphs for this experiment, we see that bit 3 never was learned by the subject. These correlations with bit 1 could either represent a "dragging" of bit 3 by bit 1--that is, an _a priori_ preference for one of a pair of indistinguishable responses--or attempts by the subject to make hypotheses on the basis of other manifestations of bit 1 in the stimulus. We can throw some light on the question by examining the tabulations of responses vs. values to see if such _a priori_ preferences of the right type have been shown. Also, a correlation curve which does not consistently stay above or below zero would tend to favor the interpretation that the subject is testing and rejecting hypotheses.

The shape and width of the unilateral portions of the correlation curves also characterize the subject's behavior in a more general way. Rapidly rising, sharp, and rapidly falling curves probably indicate explicit hypotheses, consciously tested and rejected. Longer, flatter curves, on the other hand, would seem to indicate less explicit testing and the inability unequivocally to reject contradicted hypotheses--perhaps because they are less explicit. Curves which are long <u>and</u> high (though inappropriate) or regions which show no correlations of any significance on any of the correlation curves probably indicate breakdowns of the subject's ability to work on the task. Long, high curves indicate that the subject is unable to keep himself from making responses which he must clearly see are inappropriate; while failing to correlate anything would seem to indicate an inability to make any hypotheses at all. Of course, a particularly quick subject might well be able to reject hypotheses so fast that no significant correlations would be visible while he was testing, but such a subject should then show an eventual sudden emergence when he finally hit upon a correct hypothesis.

F. Analysis of Performance

The classical measure of learning is the learning curve--a plot of level of performance versus time. We can, of course, produce such curves--not only for totally right responses (Figure 3-10), but for parts of those responses as well. Figure 3-11 shows typical curves for the learning of each bit, and Figure 3-12 shows

EXPERIMENT EXPT NO 1
SUBJECT
DATE DECEMBER 20, 1964
PERCENT CORRECT RESPONSES

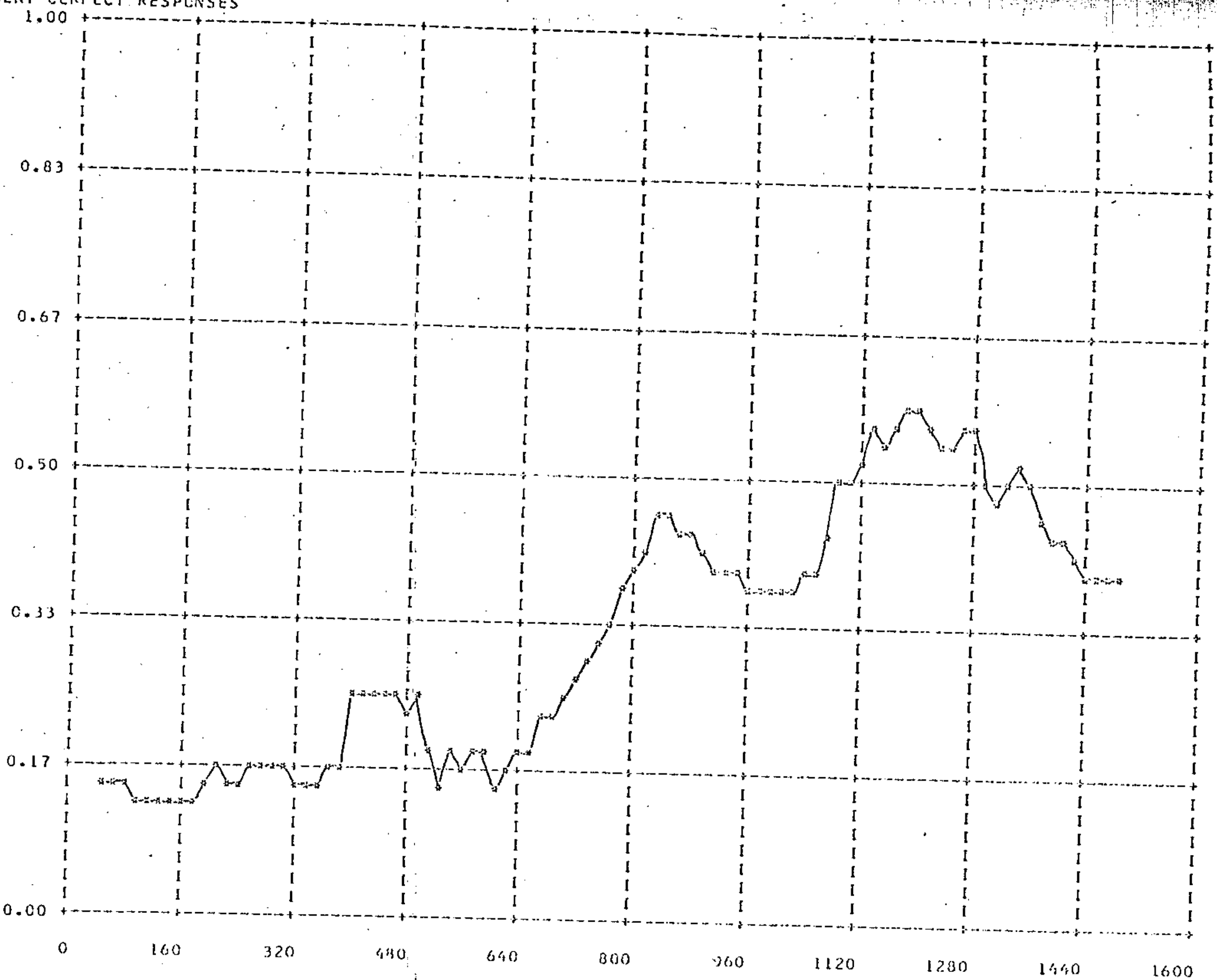


Figure 3-10

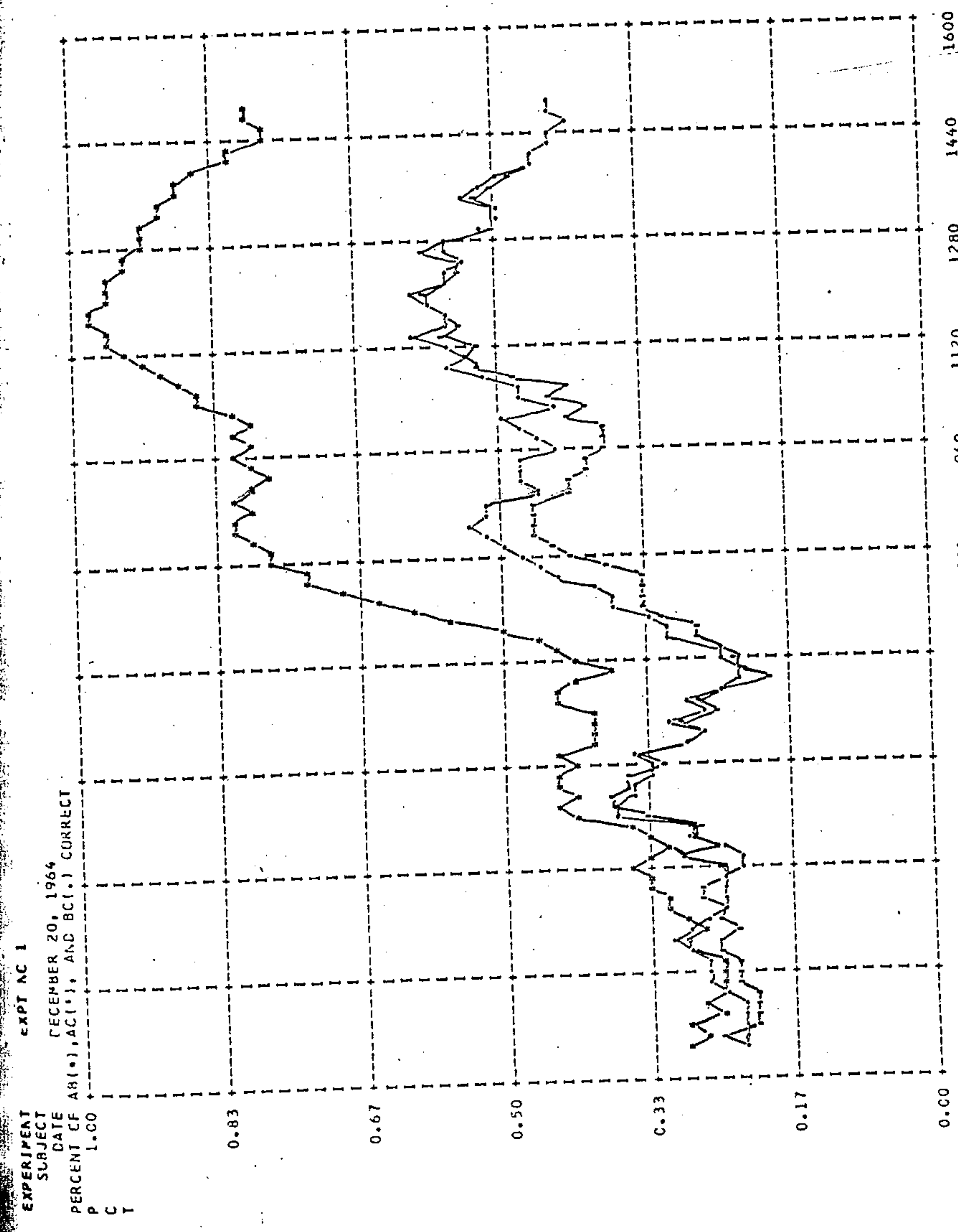

EXPERIMENT EXPT NC 1
SUBJECT
DATE DECEMBER 20, 1964
PERCENT OF AB(•), AC(*), AND BC(.) CORRECT
P
C
T
1.00
0.83
0.67
0.50
0.33
0.17
0.00
0
160
320
480
640
800
960
1120
1280
1440
1600

Figure 3-11


EXPERIMENT EXPT NO 1
SUBJECT
DATE DECEMBER 20, 1964
PERCENT OF A(*), B(°), AND C(.) CORRECT
P C T
1.00
0.83
0.67
0.50
0.33
0.17
0.00
0
160
320
480
640
800
960
1120
1280
1440
1600

Figure 3-12

learning curves for each of the three bit pairs. The intervals over which the percentages are calculated are the same as those used in the correlation calculations.

Our ability to get below the simple learning curve does not end with these decompositions, however, for there are almost numberless ways we could break down the subject's performance. For example, here are a few of the ways we might plot performance curves:

1. Performance on each separate response--8 or 16 curves.
2. Performance on each separate bit when each bit had a specific value--18 curves, or 42 curves if noise bits are included.
3. Performance on each separate bit when each bit pair has a specific value--36 or 252 curves.
4. Performance on each bit pair when each bit or bit pair has a specific value--54 or 294 curves.
5. Performance on each bit when each other bit or bit pair is correct or incorrect--18 curves.

Clearly, we have to exercise some selectivity if we are to be able to see any trees at all in a forest of computer output.

Experimentation with a number of different types of performance graphs has taught several guiding principles:

1. The more specific are the conditions placed on a graph, the more the graph comes to represent the smoothing technique, rather than the data. To illustrate this principle, consider the extreme case in which we might try to graph the performance on each individual stimulus--128 graphs in all. In a typical experiment, we might have between 500 and 1000 trials, meaning that each unique stimulus would have been presented about four to eight times--with some perhaps never having been presented at all. That would hardly be sufficient data on which to construct a curve of performance over time.

2. On any given experiment, most specific curves are completely irrelevant, because they have no meaningful interpretation in terms either of the structure of the experiment or of the subject's perception of the experiment.

3. By looking at the general performance graphs, the tabulations, and the correlation graphs, it is usually quite apparent which specific tabulations might be relevant to a particular experiment. Thus, for instance, if bit C shows no learning at all, it can hardly be worthwhile to look at the relation of the learning of C to other specific events. The same argument can be made for a bit which is learned almost perfectly, so that the only bits of interest are those which show a reasonably broad plateau in their performance curve. Sudden rises in performance level do not have enough data points to reveal their structure, and continuously rising curves have a structure that is presumably changing throughout. The details of this change might be extracted by looking at the data listing, but would surely be lost in any smoothing of the data.

Observations such as these have led to the evolution of a general method of proceeding with the analysis of each experiment: the performance and correlation curves are used for focus attention on specific areas of interest, the tabulations are used to try to reveal the intermediate structure of the interesting phenomenon, and the annotated listing is perused for specific events or regularities. Originally, it was thought that a separate program would be written to extract specific performance curves after a preliminary perusal of the data had been made. Perhaps if a more accessible computer system had been available, this plan would have proved workable; but under the circumstances, it is more convenient and direct to examine specific phenomena by manual methods using the general data reductions already in hand.

Because of the symmetry and simplicity of structure of much of the experimental apparatus, manual extraction of important data is often easier than it might first seem. In particular, because of the precise printing of the computer outputs, templates can often be used to simplify the extraction of particular data. As an illustration, suppose we wanted to see whether the performance on bit A was influenced by the value of noise bit 2. It is a simple matter to construct a template with two windows so spaced that one exposes the first column of V-R and the other, the second column of NOISE on the data listing. A third window can be made to follow, say, the sequence number, for proper alignment. Then, by moving the template down the page, we immediately reveal just those cases we wish to examine, either for tabulation or for marking for later study.

Templates may also be used with great effect on the tabulations, although hypotheses involving noise bits cannot be tested there. The group of three templates shown in Figure 3-13 can be used either singly or overlaid to group the tabulations according to bit or bit pair values or errors. As a sample of what can be done, the bit A template, when placed as shown over the R*V matrix shows all the errors where bit A = 0 and is wrong in the lower left window; the errors where bit A = 1 and is wrong in the upper right window; and all bit A errors in the two windows together. By turning the template over, the windows can be put in the upper left and lower right, thus showing where bit A is correct. By overlaying the bit A template in this new position

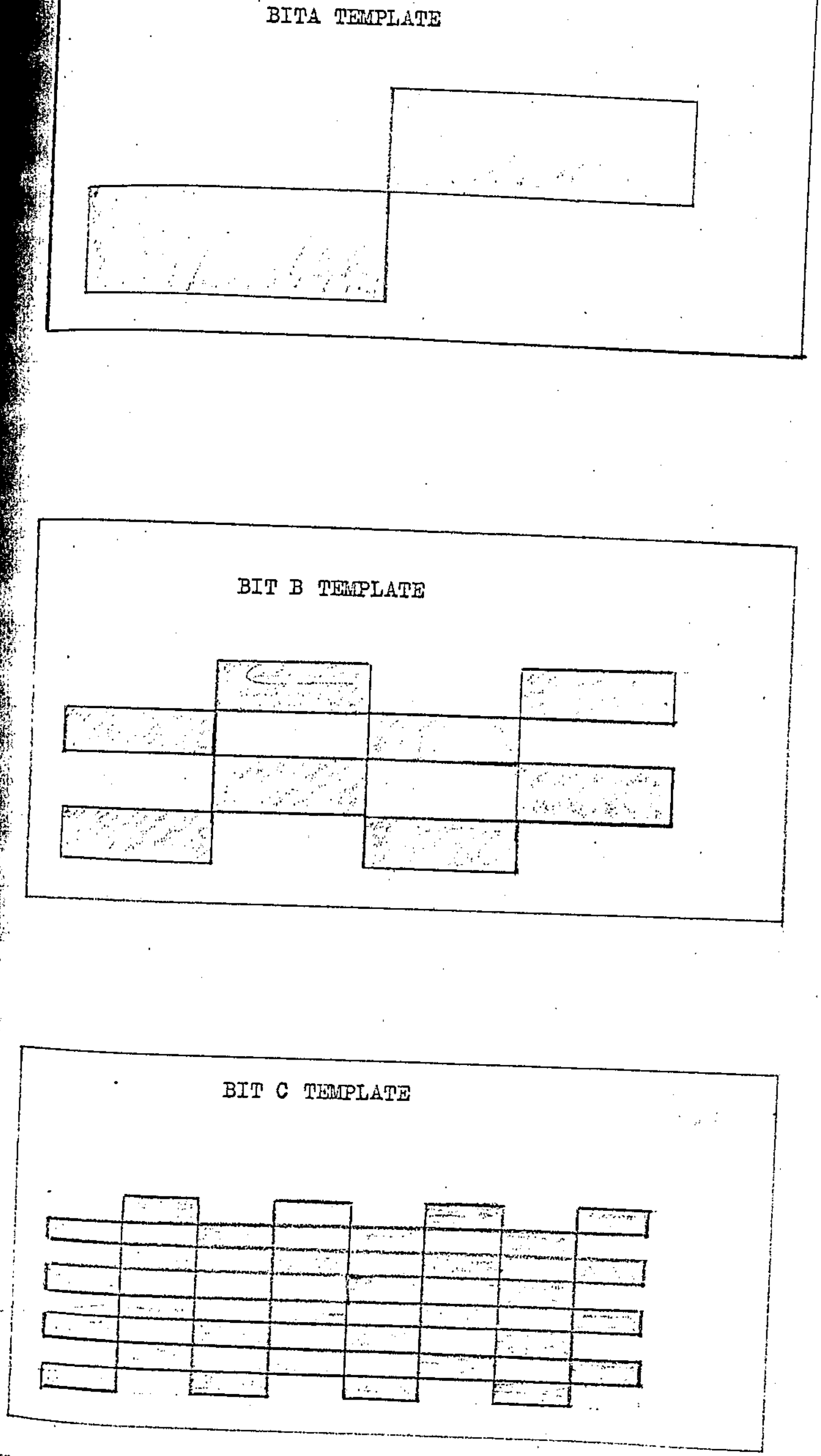


Fig.3-13.--Templates for analyzing tabulations.
(Shaded areas represent cutouts.)

with the bit C template as shown, we can immediately see all the cases where bit A is right and bit C is wrong. The extension to all sorts of other cases should be clear from these specimens. In addition, special templates for less regular conditions can be made--and a few minutes with celluloid, masking tape, and a razor blade can prove much more rewarding than a week's worth of trips to the computing center.

G. Gestalt Factors

Correlational data relating individual bits of the stimulus and response cannot tell us certain things about the subject's behavior. In particular, if the correlation for a particular bit rises, does it indicate a general tendency to get that bit right in all responses or does it indicate perfect behavior on some responses and random behavior on others? In other words, is the subject learning by learning each response part and then putting them all together, or is he learning to recognize one class of patterns at a time and then (perhaps) later abstracting the structure from what appears to be shared among several learned responses? In classical psychological terms, to what extent is he learning by Gestalts?

One way of attacking this question is to look at the tabulations of response vs. value. For a given set of trials, we can extract from the tabulations the following factors:

1. The percentage of totally correct responses, PABC.

2. The percentage of responses in which each pair of bits was correct (including those in which all three were correct), PAB, PAC, and PBC.

3. The percentage of responses in which each bit was correct, PA, PB, and PC.

Now, if these percentages represented the true probabilities of each event, we could construct a number of models for measuring the extent of Gestalt contribution to learning. Since they are the best estimates we have for those probabilities, we use them as if they were the true figures and take the results with a grain of salt. In order to gain some measure of protection against random data fluctuations, two rather different models were used, which we can call the product-moment model and the g-model.

1. The Product-Moment Model

What we are trying to measure in both of these models is the extent to which performance on all three bits combined exceeds the performance which we would expect from the individual bit rates of performance. If there were only two bits involved, we could use the well-known product-moment correlation directly; but since there are three bits involved, we have to decide how to pair them. Shall we measure how much the performance on ABC exceeds that expected from the performance on AB together with that on C? Or should we group AC, B? Or BC, A? Obviously there is no reason to prefer one to the other so we choose to use a simple average of the three, to help smooth out irregularities in the data.

The formula for the product moment correlation coefficient is

$$r = \frac{N\Sigma XY - \Sigma X \Sigma Y}{\sqrt{|N\Sigma X^2 - (\Sigma X)^2| \; |N\Sigma Y^2 - (\Sigma Y)^2|}}$$

In our case, however, the variables only take on the values 0 and 1, so that ΣXY = the number of times X and Y are both 1 (right), and

$$\Sigma X^2 = \Sigma X, \ \Sigma Y^2 = \Sigma Y.$$

The equation for r may thus be simplified to read

$$r = \frac{N\Sigma XY - \Sigma X \Sigma Y}{\sqrt{\Sigma X(N - \Sigma X)\ \Sigma Y(N - \Sigma Y)}}$$

Dividing numerator and denominator by N^2 we get

$$r = \frac{\Sigma XY/N - (\Sigma X/N)\ (\Sigma Y/N)}{\sqrt{(\Sigma X/N)\ (1 - \Sigma X/N)\ (\Sigma Y/N)\ (1 - \Sigma Y/N)}}$$

But in terms of our notation for the various percentages of right responses,

$$PX = \Sigma X/N$$
$$PY = \Sigma Y/N$$
$$PXY = \Sigma XY/N$$

Thus,

$$r\ (X,Y) = \frac{PXY - PX \cdot PY}{\sqrt{PX \cdot PY \cdot (1 - PX)\ (1 - PY)}}$$

a formula which holds regardless of whether or not PX and PY are simple events, as long as PXY counts the percentage of trials on which both events, X and Y, occur. Thus, we may use this formula to calculate pair product moments for each pair--in order to measure "partial Gestalts"--or to get a measure of the overall Gestalt by the formula

$$f(A,B,C) = (r(AB,C) + r(AC,B) + r(BC,A))/3$$

2. The g-model

Another way of deriving a measure of Gestalt from the given data is based on a more explicit model. Let us assume that there are four "real" quantities concealed in the data, a,b,c, and g. a is the probability that the subject would have gotton bit A correct, independent of whether or not he would get it correct because of the operation of a Gestalt. b and c are similarly defined for bits B and C; while g is the probability that the entire response will be made correctly, independent of the chance of getting it right because of the coincidence of the independent probabilities of A, B, and C. Using these quantities we can write down four equations directly which relate them to one another and to the observed data, assuming that it represents a perfect sample:

$$PABC = a \cdot b \cdot c + (1 - a \cdot b \cdot c) \cdot g$$

$$PA = g + (1-g) \cdot a$$
$$PB = g + (1-g) \cdot b$$
$$PC = g + (1-g) \cdot c$$

We can rewrite these equations in the form

$$g = (PABC - abc)/(1 - abc)$$
$$a = (PA - g)/(1 - g)$$

and so forth. In this form, the equations may be solved by developing a polynomial and finding its roots; but using a computer, it is simpler to solve them directly by iterating until the values coverge to stable points.

3. Comparing the Two Models

The g-model has the advantage of yielding, in addition to g, the measure we were seeking, estimates of the individual bit learning. It does not, however, take separate account of the pair Gestalts, but lumps them into g. In this way, it differs from the product moment model, in which f measures that part of performance not accounted for by both the single and double bit performance. Thus, we would generally expect g to be greater than or equal to f--within the accuracies of the data--and if there is any significant difference between them, we may reasonably attribute it to significant pair correlations.

A third model was tried which extended the g approach to try to separate the pair factors directly, but the equations did not always converge using actual data. In yet another model, the product of the three pair product moments was used as a measure of the correlation among the three bits; but this technique had the disadvantage of magnifying, rather than reducing, small data deviations. It should be noted that f and g themselves present certain difficulties in this regard. In particular, as any of the individual percentages gets close to one, the product moment calculation becomes extremely sensitive to slight fluctuations and is not a reliable measure, whereas the g calculation is more sensitive at low performance levels. We must therefore be wary of making conclusions based on momentary excursions of the value of f or g, especially when near random or perfect performance. In the middle ranges, however, f and g should not be so sensitive. Furthermore, they have the important property that--in those ranges at least--they are not dependent on the actual level of performance. It is possible, for example, to have the same value of g when PABC = .6 as when PABC = .9, or, on the other hand, to have g = 0 or g = .3 at two different places where PABC = .7 or some other constant value. Consequently, they may be said to be measuring something that the performance measures are not.

Upon applying these two models to some of the tabulated data, certain results emerged which seemed to indicate their usefulness. As a result, the system was expanded to include running calculations of both g and f, the three pai product moments (X=r(A,B), Y=r(A,C), Z=r(B,C)) and a,b, and c. These were calculated and plotted on the same overlapping interval scheme used in the correlation calculations. A typical set of graphs obtained are shown in Figures 3-14, 3-15, and 3-16.

H. Test Trials and Reconstruction of Decision Procedures and Strategies

The main purpose for using test trials is to isolate more precisely the parts of the stimulus pattern which the subject is using to make his discriminations. We must be alert, however,

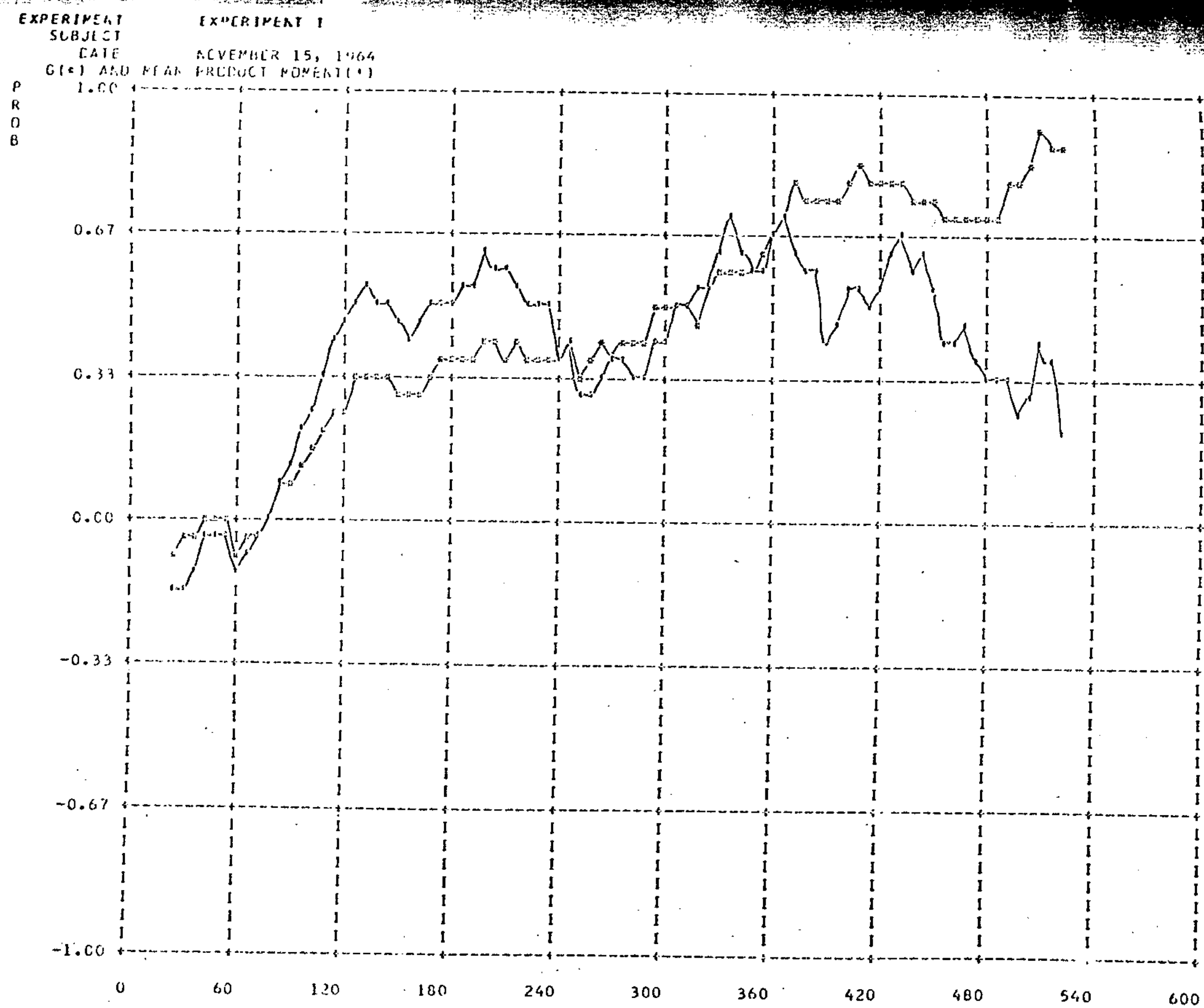
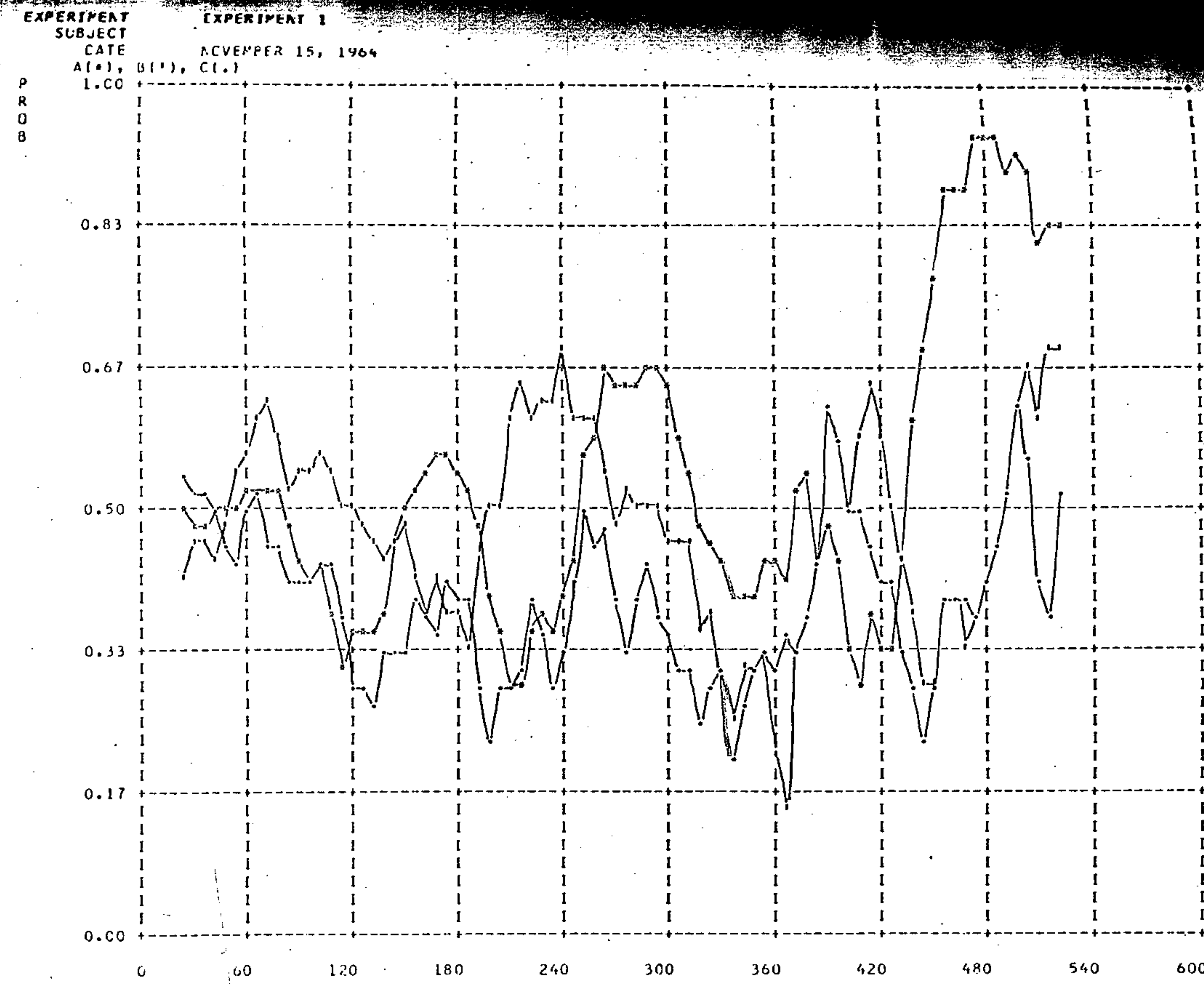

EXPERIMENT EXPERIMENT I
SUBJECT
DATE NOVEMBER 15, 1964
G(*) AND MEAN PRODUCT MOMENT(*)
PROB
1.00
0.67
0.33
0.00
-0.33
-0.67
-1.00
0
60
120
180
240
300
360
420
480
540
600

Figure 3-14



EXPERIMENT EXPERIMENT 1
SUBJECT
CATE NCVEMBER 15, 1964
A(*), B('), C(.)
P
R
O
B
1.00
0.83
0.67
0.50
0.33
0.17
0.00
0
60
120
180
240
300
360
420
480
540
600

Figure 3-15

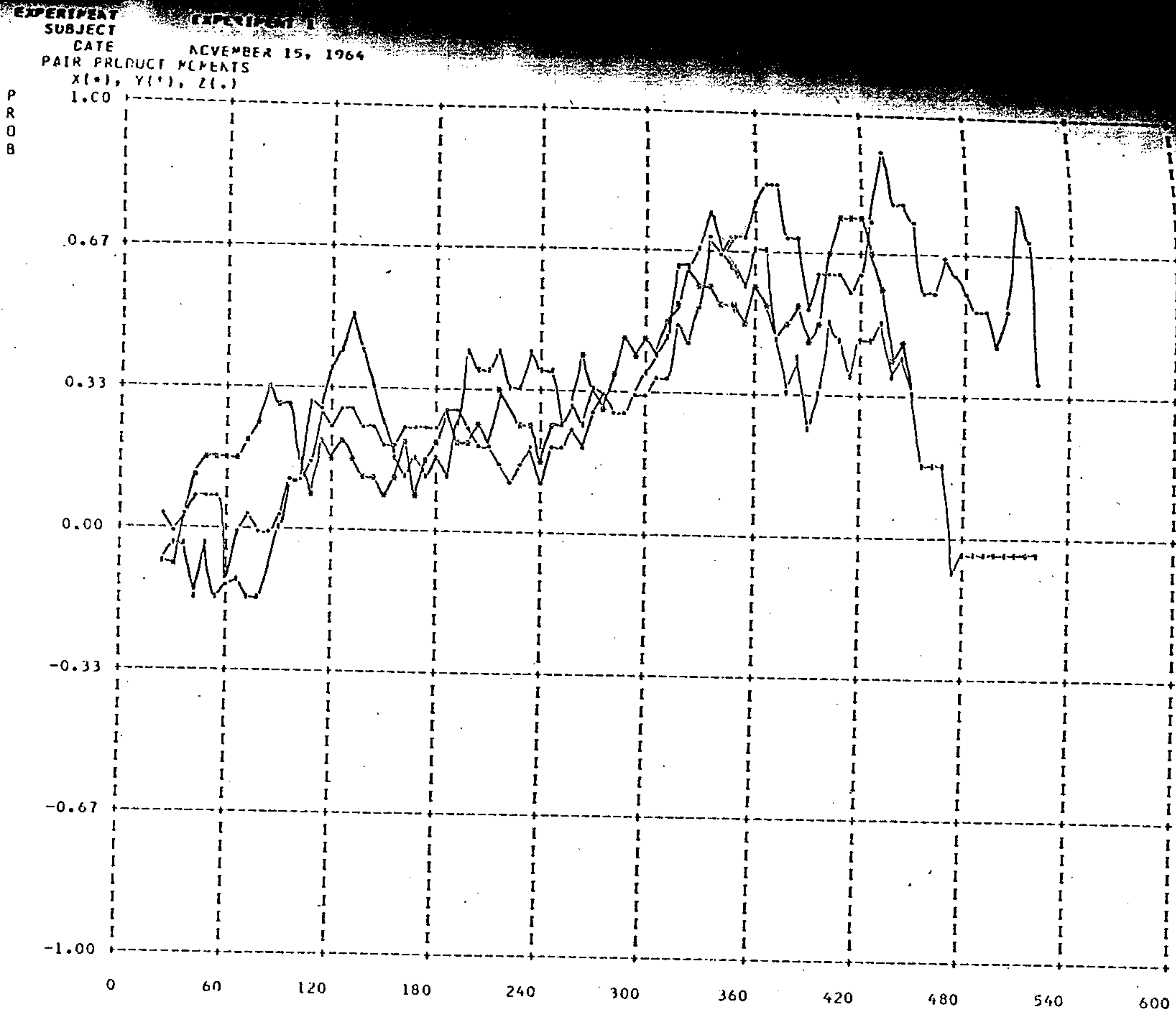

EXPERIMENT
SUBJECT
CATE ACVEMBER 15, 1964
PAIR PRCDUCT MCMENTS
X(*), Y('), Z(.)
PROB
1.00
0.67
0.33
0.00
-0.33
-0.67
-1.00
0
60
120
180
240
300
360
420
480
540
600

Figure 3-16

to secondary information available from them. For example, certain models of concept learning predict that the presentation of stimuli without reenforcement will interfere with concept learning, while other views indicate that stimulus familiarity may be enhanced by such presentations. Although no special tests of these hypotheses are made, other measures of behavior can be examined in the vicinity of the test trials, particularly looking for changes in behavior which seem to originate during those trials.

1. Test Tabulations

The data taken during the test trials is tabulated for each individual set of 32 trials, using the tabulation subroutines. As shown in Figure 3-17, we tabulate for each deleted character position (column) the number of errors (rows 1 through 3) and the number of non-errors (rows 5 through 7) in each of the three response bits (A = 1 or 5, B = 2 or 6, C = 3 or 7). This same structure is, of course, reflected in the response time matrix, which can be a valuable source of information in test trials, too. The mean element and mean diagonal are not significant in these tabulations.

There are several types of inferences we can make from these tabulations:

a. We can compare the mean response time for test trials (e_{TT}) with the mean time for other trials at the same point in the experiment. In this case, the mean is

PERIMENT EXPERIMENT 1
SUBJECT
DATE NOVEMBER 15, 1964
TEST ENDING AT
553

1-3

OTALS

	1	2	3	4	5	6	7	8	N	T
1I	0	0	0	1*	0	C	0	0 I	0I	1I
2I	1*	0	0	0	0	0	0	0 I	0I	1I
3I	0	0	0	0	0	1*	0	0 I	0I	1I
4I	0	0	0	0	0	0*	0	0 I	0I	0I
5I	4*	4*	4*	3=	4*	4+	4*	4*I	0I	31I
6I	3=	4‡	4	4+	4	4	4	4 I	0I	31I
7I	4‡	4	4	4	4	3=	4	4 I	0I	31I
8I	0‡	0	0	0	0	0=	0	0 I	0I	0I
NI	0‡	0	0	0	0	0=	0	0 I	0I	0I
TI	12*	12	12	12	12	12	12	12 I	0I	96I

MEAN ELEMENT = 2MEAN DIAGONAL = 2

RAGE TIMES

	1	2	3	4	5	6	7	8	N	T
1I	0	0	0	233*	0	0	0	0 I	0I	233I
2I	271*	0	0	0	0	0	0	0 I	0I	271I
3I	0	0	0	0	0	215*	0	0 I	0I	215I
4I	0	0	0	0	C	0*	0	0 I	0I	0I
5I	219	183*	236*	261*	249*	229	170*	243*I	0I	222I
6I	202-	183	236	254‡	249	229	170=	243 I	0I	221I
7I	219	183	236	254‡	249	233+	170=	243 I	0I	223I
8I	0	0	0	0‡	0	0	0=	0 I	0I	0I
NI	0	C	0	0‡	C	0	0=	0 I	0I	0I
TI	219	183	236	254‡	249	229	170=	243 I	0I	223I

Figure 3-17

223 (2.23 seconds) compared with 2.39 seconds for the previous 94 trials, a figure obtained from the other tabulations.

b. We can examine the mean times for each missing character position (row T, columns 1-8). In this case, columns 2 (1.83 sec) and 7 (1.70 sec) may be significantly lower than the other times, indicating that these may have little to do with the subject's discriminations. This effect, however, may be due to chance variation in that better learned values may have been presented on these trials.

c. If a significant number of errors were made (as is not the case here), we could compare response times for correct vs. incorrect, as a whole or bit by bit.

d. When the test trial series in question is the one at the end of an experiment (as this one is), we can assume that in the absence of the deletions, the subject would have performed perfectly. Under that assumption, any errors made can be attributed to the missing information. To the extent that these assumptions are correct, then, the errors indicate exactly which positions were used to identify each bit. In this example, the error in e_{14} indicates that bit A is associated with some attribute of character 4; e_{21}, that bit B is associated with some aspect of character 1; and e_{36}, that bit C is associated with some aspect of character 6.

e. By referring to the constraint specifications for this experiment (Appendix I), we can usually make a more explicit identification of the discriminating attributes. For instance, the only place bit C enters into the composition of character 6 is in the case of the letter there. Similarly, the only place bit B enters into 1 is in the case. Although bit A enters into both the case and the ltra bit of the letter in position 4 and thus cannot have its determinant uniquely identified, the fact that the other two bits are determined by case is a strong indication that bit A is also. Thus, we can conclude with fair certainty that the subject has built his discrimination on the capitalization of the left hand column (positions 1, 4, and 6) of the stimulus array. In this case, that analysis is in exact accord with the verbal report.

f. To some extent, we can make inferences from the number of errors made, although we must take care because of the small sample size. Since each character position is deleted four times in each series, random behavior with respect to a bit when that character is deleted should result in an average of two errors. Viewed another way, the chances of any number of errors can be calculated as for Bernoulli trials, so that, for instance, zero errors should occur by chance only once in 16 times. If, then, the number of errors is lower than chance, it may indicate that the subject is to some extent taking information about that bit from more than one source, so that his performance is not so disturbed by the elimination of a single character. In this case, though the errors are lower than expected, we would not be justified in making such a conclusion based on this evidence alone. Supporting evidence might be obtained from examining other test trials to see if the same trend were in evidence.

2. Test Analysis Through Examination of Data Listings

When the subject is not performing at or near perfection, we may have to make a more detailed analysis of the test tabulations to correct for errors he might have been making even if he had full information. Also, we may have to make corrections to the response times based on the differences in the responses tested by each deletion. These analyses and adjustments can be made by examining the data listings, as shown in the following examples:

a. We noted earlier the low response times when positions 2 and 7 were deleted. By examining the data listing, we find that position 2 was blanked for values 4, 0, 3, and 2, which have an average response time of 2.32, while position 7 was blanked for 0, 4, 2, and 0, which have an average response time of 2.18. To be more precise, then, we should transform the average times to a deviation from the expected mean, which would yield -.49 for 2 and -.48 for 7. In fact, if we carry out this transformation for all eight positions, we get the adjusted times:

(1) = +.07
(2) = -.49
(3) = -.04
(4) = +.12
(5) = -.29
(6) = +.17
(7) = -.48
(8) = -.11

Considering the empirical nature of these numbers, they give a surprisingly good validation of the conclusion we reached through examining the error rates alone, namely, that positions 1, 4, and 6 are the positions most relevant to the subject.

b. The same adjustment can also be made to the overall mean time, in this case yielding a deviation of about -.13 for the test trials.

c. Where the subject has been making certain errors consistently before and after the test trials, the error tabulations can be adjusted for errors probably not caused by the deletions. Figure 3-18 shows the test trial tabulations for the test trials preceding the final ones, shown in Figure 3-17. Here, the subject was not yet performing to the terminating criterion. In particular, the subject was consistently confusing

EXPERIMENT — EXPERIMENT 1
SUBJECT
DATE — NOVEMBER 15, 1964
TEST ENDING AT
490

W=1-3

TOTALS

	1	2	3	4	5	6	7	8	N	T
1I	0	0	0	1*	0	0	0	0 I	0I	1I
2I	4*	0	0	1=	0	0	1-	1-I	0I	7I
3I	1*	0	0	1	0	2*	1	1 I	0I	6I
4I	0=	0	0	0	0	0‡	0	0 I	0I	0I
5I	4‡	4*	4*	3*	4*	4+	4+	4+I	0I	31I
6I	0	4‡	4	3=	4	4	3	3 I	0I	25I
7I	3	4‡	4	3	4	2=	3	3 I	0I	26I
8I	0	0‡	0	0	0	0=	0	0 I	0I	0I
NI	0	0‡	0	0	0	0=	0	0 I	0I	0I
TI	12*	12	12	12	12	12	12	12 I	0I	96I

MEAN ELEMENT = 2MEAN DIAGONAL = 1

AVERAGE TIMES

	1	2	3	4	5	6	7	8	N	T
1I	0	0	0	264*	0	0	0	0 I	0I	264I
2I	264	0	0	258=	0	0	267*	265+I	0I	264I
3I	362*	0	0	258=	0	300+	267	265 I	0I	292I
4I	0‡	0	0	0=	0	0	0	0 I	0I	0I
5I	264‡	204*	239*	257-	218*	245	238	207 I	0I	233I
6I	0	204	239	259‡	218	245	228-	180*I	0I	226I
7I	232-	204	239	259‡	218	191-	228	188=I	0I	221I
8I	0	0	0	0‡	0	0	0	0=I	0I	0I
NI	0	0	0	0‡	0	0	0	0=I	0I	0I
TI	264‡	204=	239	259	218	245	238	207 I	0I	234I

Figure 3-18

responses 5 (101) and 6 (110) both before and after the test trials. If we examine the data listing, we find that the two errors in the 8 column were caused by precisely this error, which yields an error pattern of 011. Furthermore, we find this same error occuring three other times in the trials, when 1, 4, and 7 were deleted. If we adjust the error figures under the assumption that these errors were not caused by the deletions at all, we get the following error figures:

1	2	3	4	5	6	7	8
0	0	0	1	0	0	0	0
3	0	0	0	0	0	0	0
0	0	0	0	0	2	0	0

which, though based on a slightly smaller sample, are precisely consistent with the results obtained in the next series of test trials, which were shown in Figure 3-17. From this, we can infer that the subject had already established his final basis for discrimination at this earlier time. In a similar manner, we can trace back through still earlier tests to narrow down to the point where the basis of discrimination was first used.

d. Since the adjusted error figures in the previous step are much closer to expectation than were the figures for the final set of test trials, we also have strengthened the evidence for a possible spreading of the basis of discrimination as the experiment proceeds. This phenomenon, too, can be traced back through previous test trials.

e. The methods of analysis so far given depend on the assumption that the subject uses the same stimulus component each time he discriminates the same bit. Although we know that this is possible, because of the structure of the experiments, there is no way the subject knows this, a priori. Thus it is possible--and indeed it happens--that the subject will use different parts of the stimulus pattern to get the same unit of information, perhaps depending on other bits to decide which part to use. No simple tabulation will reveal such a structure to us, but detailed examination of the data listing of the test trials can often tell us a great deal. Let us look at an example of such an analysis:

Subject 2 developed the strategy of looking at only the brightly lit letters of a stimulus pattern. This strategy often led him to highly complex rules, as the rules for each response depended on a different set of discriminants. In the case of Experiment 3 this strategy led him to make a quick discrimination of the set (000,010,100,110) from the set (001, 011, 101, 111) according to whether the letters in positions 7 and 6 (the lower left corner) were bright or not. When he learned to discriminate among the members of the first set, the information did him little good for learning to discriminate among the second, for position 8 was not bright for the second set and was a pivotal position for the first. Thus, he had great difficulty learning the second set and, in fact, probably never had a clear idea of what he was doing with it. About the first set, on the other hand, he had a very clear decision structure for choosing among its members. Because of the two sets of rules, however, the simple evidence from the test trials seemed confused. In such a case, we resort to a detailed analysis of the data listing--perhaps using special listings which emphasize certain points critical to this experiment. Using such techniques, the reconstruction of the decision structure shown in Figure 3-19 was created. Space does not permit giving all the details of that reconstruction, but we can show their nature by reasoning back to the data from the diagram.

Looking at the diagram, we can form several predictions of what the data should show if the diagram is correct:

a. The discrimination of bit C should not be disturbed very much by deletions, because one of the two discriminating characters will always be present.

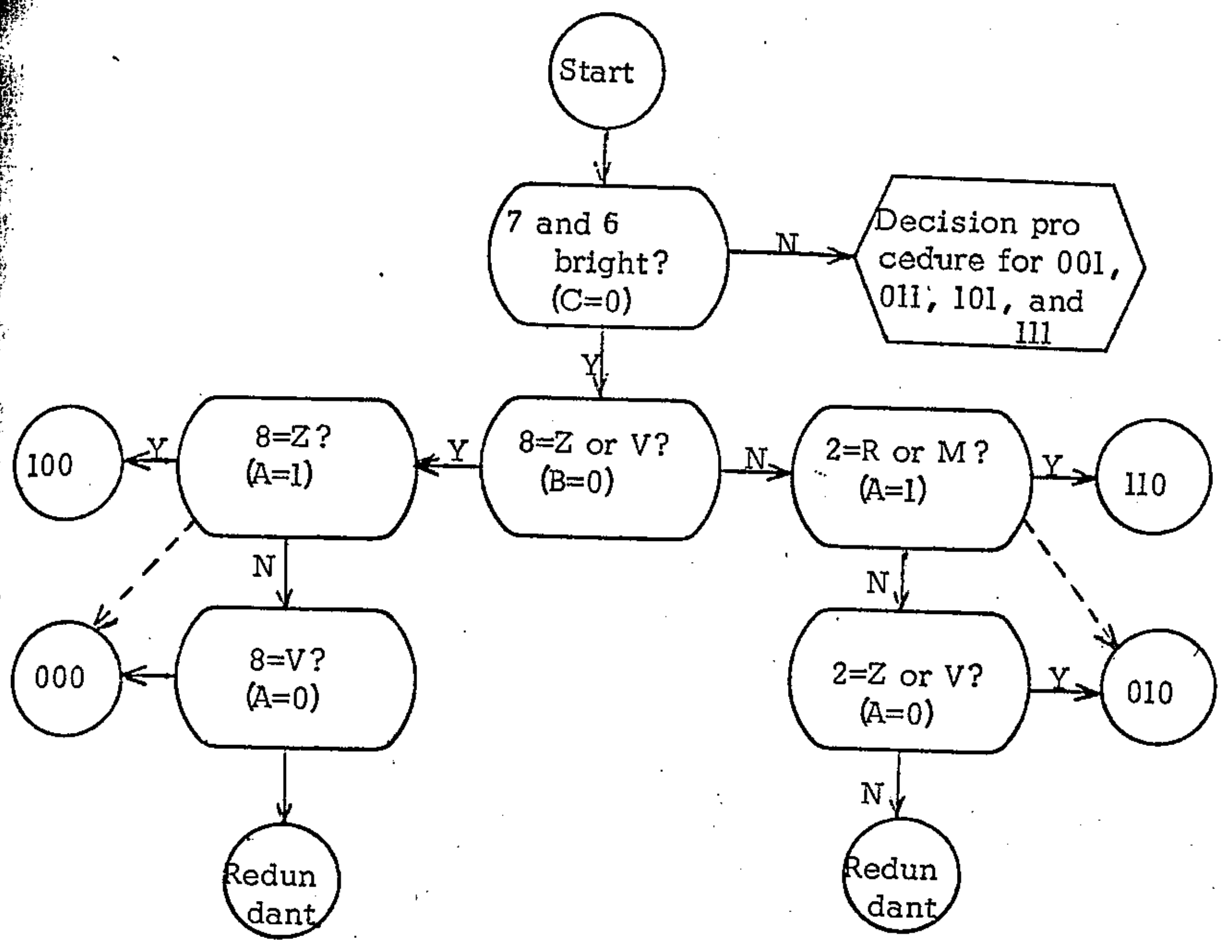


Fig. 3-19. -- Decision diagram for subject 2 on experiment 3

b. When position 8 is deleted, the subject should display difficulty with 000 and 100. He should not (necessarily) show difficulty with 110 and 010, since the evidence of position 2 will still be shown (and the test on position 8 is redundant for the right side of the diagram anyway, as position 2 is only bright when bit B = 1).

c. When position 2 is deleted, the subject should display difficulty with 110 and 010. He should not show difficulty with 100 or 000, since position 2 does not enter their discrimination at all.

When we examine the data from the tests following the mastery of this set of responses, we find the following results:

a. In 352 trials, bit C is missed only 10 times, and 9 of those times, one or more other bits were in error, indicating that a more general confusion may have been the cause. In fact, none of the errors came when 7 or 6 was deleted.

b. When position 8 was deleted, the following results obtained for the first set:

Value	000	010	100	110
Errors	3	0	2	0
Occurences	7	5	8	11
Response Time	2.51	1.96	3.81	1.68

c. When position 2 was deleted, the following results obtained for the first set:

Value	000	010	100	110
Errors	0	3	0	2
Occurences	8	4	3	2
Response Time	1.89	3.92	1.55	4.89

Clearly, all three predictions are amply fulfilled. Therefore, the diagram seems to represent an objective statement of the decision procedure in use, although certain modifications would also be consistent with the data.

I. Verbal Reports

Because verbal reports were thought to be at most a secondary source of data, not as much care was given to their design as was given to the rest of the experimental procedure. This decision was somewhat unfortunate, because, as the early experiments were analyzed, it became clear that there might be interesting relationships between the verbal reports and the objective data. After the first subject--whose remarks were taken down as notes--verbal reports were recorded on an IBM Executary and later transcribed for study.

No special set of questions was used for obtaining the verbal reports. Generally, the subjects were encouraged to give the following information:

1. How did they <u>feel</u> about what they were doing?
2. How did the experiments compare with one another?
3. What had they learned?
4. How did they proceed to learn?
5. What difficulties were they experiencing?
6. What additional comments could they make about the experiment?

In addition, they were asked about their physical comfort and other physical aspects of the task. Some of these answers were used to improve conditions for future experiments.

In retrospect, it might have been valuable to have more explicit information on the strategies they thought they were using and the exact things they had learned at each interview point. The reason such precise questioning was not used was a fear that it would impose structure on the experiment that the subject had not extracted for himself. If this were done, we could not observe changes in the subject's perception of that structure as he proceeded through the experiments.

Consider the following specimen of one interview:

Subject: ....and I look at the middle letter in the center vertical column. The middle letter in the whole board....

Experimenter: The middle letter? There isn't any letter in the middle.

S: Well, there's nine things, aren't there?

E: There are eight things.

S: There are eight things? Then what am I looking at?

E: There is a plus in the middle.

S: Oh, a plus, yes.....The one above the plus? ......

This was the third experiment for this subject, and he did not yet know the general structure of the stimulus--a fact of some interest, especially since he had learned one of the first two quite easily.

Unfortunately, even in the above exchange, the experimenter has given him some information which may help him to modify his image of the stimulus and the rule he was using. It is extremely difficult to strike the right middle path between precision of report and naiveté of subject.

Be that as it may, there is much information to be gleaned from the verbal reports, particularly since we can form such a precise idea of what was going on without them. As a consequence, we are able to identify at least the following types of situations--which we shall discuss more fully later:

1. The subject reports correctly on what he is doing.

2. The subject's report is in contradiction to what he is doing.

3. The subject is unaware of something he is doing.

4. The subject shows a change in method of talking about the structure of some part of the experiment.

5. The subject experiences emotional reactions which differ for different experiments.

CHAPTER IV

RECONSTRUCTION OF EXPERIMENTS

In this chapter we shall examine each individual experiment in the order performed by each subject. Whenever we see some substantially new phenomenon, we shall spend some time showing how deductions about that phenomenon were made from the data. As the same phenomena appear repeatedly, however, we shall only allude to them, for to repeat each reconstruction would take hundreds of pages. Consequently, the reconstructions get shorter and shorter as fewer and fewer new phenomena are introduced.

Each reconstruction follows the same format. First, a brief general description of the behavior is given. This is followed by a description of the decision structure which the subject was using at the end of the experiment, insofar as it can be derived from the data. Next, the process by which the decision structure was developed is reconstructed; and, finally, interesting points from the subject's verbal report are related to the reconstructions.

A. Subject 1

1. First Experiment (Experiment 1)

This was the subject's first experiment, and she was slightly confused at the beginning about how quickly she had to respond.

She was responding at about a one second rate for about the first 100 trials, until the experimenter decided to remind her that she had five seconds to respond. She then slowed down to about a 2.5 second rate. At this point, her performance began to improve almost immediately above the random level. It continued to improve almost linearly throughout the 522 trials until criterion was reached, in all performance measures--bit, bit pair, and overall.

a. Decision Structure

The subject learned a very explicit set of rules based on the capitalization of the three letters along the left side of the stimulus, namely, 1, 4, and 6. (See Chapter III--Test trials--for details of part of the analysis.) Decision structure, however, does not seem to have been the minimal tree which would have been possible using these indications but rather a serial string of decisions, resulting in the successive testing for entire patterns (Figure 4-1a).

The last two responses shown in the string were never actually learned up to the capture criterion, although the subject seems to have been well on the way to them when the experiment terminated. If we tabulate the response times for each response at the end of the experiment, we get the following table:

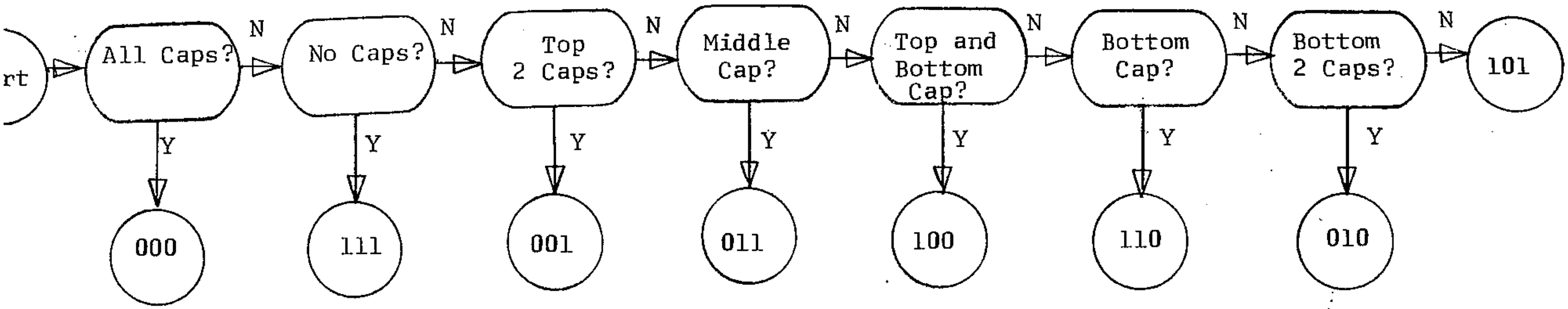


Figure 4-1a: Serial decision tree

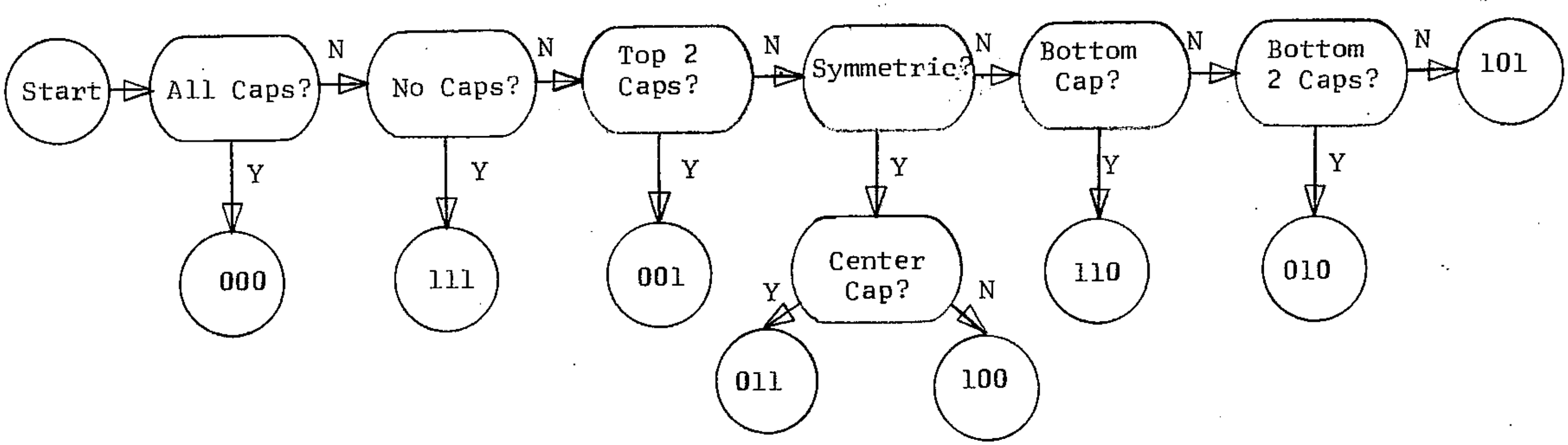


Figure 4-1b: Reconstructed decision tree

Response	Time	Capture Point	Order Learned	Revised Order Learned
000	1.70	63	1	1
111	2.00	272(190)	3	2
001	2.15	261	2	3
100	2.20	365	5	5
011	2.25	364	4	4
110	2.55	421	6	-
101	2.60	-	-	-
010	2.70	-	-	-

One anomoly in the matching of the order learned and the ascending response times is response 110. Upon investigating the annotated data listing, the reason for this anomoly becomes clear. The capture of 110 takes place on trial 421. Immediately before trial 421, the subject had been confusing 110 and 010; but between 421 and 445, she seems to have resolved that confusion long enough for 110 to meet the capture criterion. Immediately following this period, however, she began confusing 110 and 101 and continued to do so until the experiment was terminated. In fact, the last error made--at trial 502--is responding 101 to 110. It seems then, that she "had" 110 for a short time, but could not "hold" it. Thus, 110 really falls into a group with 101 and 010, none of which were really mastered and all of which were confused with one another.

The other anomoly, 111, seems actually to have been captured at 190, being given correctly eight times in a row after that. It happened, however, that five of those times were in the midst of a test, immediately after which 111 was missed once. Its official capture was thus postponed, but it might more realistically be considered the second response captured. The revised order for learning is given in the last column, and is strongly correlated with the time sequence.

We cannot say unambiguously that the ascending time sequence indicates the decision order, however, since the earliest learned responses are also the most practiced. Thus, we do not know with certainty that the order of the decisions is exactly as shown in Figure 4-1a, and the subject herself is vague on the order after the first three--whose order is well established in the verbal report. These first three, on the other hand, so seem to indicate that the order of learning determines the order of the decision structure--or at least influences it strongly.

b. Strategy

The subject's overall strategy seems to be one of isolating the individual responses and mastering them one at a time. It is rather clearly shown by the relative precision of definition of capture points (which would be fuzzy if the subject were not operating primarily on a response basis); by the slowing down of responses to a particular stimulus class just before it is captured; by the spreading of the capture points, rather than the clustering which should be encountered if more general decision principles were being isolated; and by the high g and f measures.

Within such a strategy, there are many variations possible. Perhaps the most interesting question is how the order of responses to be isolated is chosen. In this case, knowing the criteria the subject was using, we can reason back to some conjectures on choice procedures. By far the earliest response learned was 000--which in the subject's method of looking at the stimulus, was all capitals. In a rule based solely on capitalization, the case of all capitals is one of the two easiest rules to discriminate from the remainder, since placement of the capitals does not have to be remembered. In fact, the subject reports that she "noticed" this structure quite early, and that it drew her attention to the left-hand column. On the other hand, response 000 is perhaps the most conspicuous response in the array, being the topmost response and positioned on the center line. Thus, it may have been singled out for attention first, though we cannot make any conclusions about this choice on the basis of a single experiment.

Now, if our reasoning about all capitals is correct, the next easiest case should be all small letters, or 111, and this is indeed the case. If position in the response array were the determining criterion, we might expect 100 to be next (because it is symmetrically opposite to 000, on the bottom) or we might expect 111 or 001 because they are at the top, too, and next to 000. The fact that the third capture is response 001 seems to support the second of these hypotheses.

Reasoning once again on the basis of stimulus appearance, it seems that the simplest of the six responses with mixed lower and upper case would be the two symmetrical ones, LUL (011) and

ULU (100), for these do not require the subject to remember the precise placement of the cases. Reasoning on the basis of response position, however, we might pick 100 as a likely candidate, but 011 seems ruled out. 110 and 010 might be picked because they are the next two down in the symmetrical chain from the top.

A comparison of these arguments with the actual capture sequence shown, makes it seem likely that both response and stimulus "conspicuousness" contributed to the isolation sequence, the actual sequence being partly determined by random events that were "helpful" to the subject. There is a good deal of evidence that if such local "events" do not actually determine the sequence of captures, they at least contribute strongly to the determination of exactly where the capture takes place.

For example, looking at the list of capture points, our attention is drawn to 365 and 366; for this is the only place in the experiment that two responses are captured close together, the others being spaced 70 to 90 trials apart. In the thirty or so trials preceding 365, the subject's behavior was marked by a frequent confusion of 011 and 100, as would be expected on the hypothesis that the two symmetrical stimulus cases had been singled out for study at this point. Trial 364 was, in fact, such an error; and 364, 365, and 366 looked like this:

Trial	V	R	
364	100	011	N
365	011	011	
366	100	100	

Such a structure, of course, gives the subject a chance to test her hypothesis without the usual burden on memory of intervening irrelevant cases. (It is also interesting to note that the subject characterized this experiment in the following words: "I liked it very much; I had the feeling that the machine was a person helping me because at times it would give a good case for a hypothesis I was testing...I sometimes felt it was helping me.") If we look at the other capture points, we find, for example, that before the capture of 111, it was being confused with 100, and that at the capture point the structure was

Trial	V	R	
190	111	100	N
191	100	001	
192	111	111	
193	111	111	

and that just prior to this, there was the structure

184	111	100	N
185	111	111	

001 does not seem to have any such structure at its capture point, and, in fact, seems to have been gradually worked out over the course of about 150 trials. 110 also lacks any clearly helpful structure at its "capture" point, but its capture is not too definite anyway. 000 also shows nothing unusual about its capture point, but like 001, does not seem to have been consistently confused with any other response beforehand.

The confusion of one response (or one stimulus) with another can give us some clue as to the development of the decision structure. The confusion of 100 and 111 does not seem to make much sense until we realize that at that time, the subject may have been simultaneously considering response 100 because of its position (in relation to 000, the lone learned response) and 111 because of its appearance (three small letters, in relation to 000's three capital letters). This idea is supported by the fact that the confusion between the two is always a matter of giving <u>response</u> 100 to <u>pattern</u> 111, and never the reverse, as if the subject were reasoning that "if all caps is the top, then all small should be the bottom." In the verbal report, the subject said that the relationship between the placement of single capitals and the appropriate response was "not what you might expect," indicating that she did, indeed, expect that there should be some "logical" relationship between stimulus and response.

Because the responses were easier to distinguish than the stimuli, once the subject was able to distinguish one of the stimuli unequivocally, it would be easier to change her hypothesis about what response to attach to it than to look for some other stimulus class to associate with the more arbitrarily chosen response. This would answer our question, then, about why 111 was the second response learned. It would also explain why 100 was then dropped from consideration until later, and perhaps why 001 was then chosen to be solved next. Actually, 001 was being "worked on" at the same time as 111, but did not have such a clear stimulus relationship to 000.

The confusion of 100 and 011 which occurred later seems quite likely to have been on the basis of their similar stimulus structure, rather than on their placement as responses in the response array. The fact that the capture of one coincides with the capture of the other (as contrasted to the confusion between 100 and 111) is the best indication for this hypothesis. If this is the case, we should probably modify the decision structure of Figure 4-1a to that of Figure 4-1b. Thus viewed, this experiment no longer represents a pure serial strategy, but a basic serial strategy overlaid with a little stimulus generalization, or parallel strategy.

Within this overall strategy, the subject seemed to be using another sub-strategy, or tactic, which has interest of its own. The response biases shown in the earlier trials invite us to look at more local behavior. On doing so, we find that the subject is often using a tactic which may be thought of as a narrowing of criteria. When narrowing, the subject makes the response on which she is presently focusing to any stimulus for which there is any doubt. By successively eliminating stimuli which gave wrong answers, she is able to eventually narrow down to a remainder class which includes only the proper class for that response. Narrowing can be detected by observing a high rate of giving one response which diminishes until it reaches coincidence with the stimulus rate. It may also show step by step elimination of one confusion after another involving that response.

One other characterization of this experiment should be noted, and that is the relative sharpness of the peaks on the correlation graphs corresponding to incorrect hypotheses. Sharp high peaks should indicate very explicit hypothesis making, which seems quite consistent with the rapid exclusion of all characteristics but the capitalization of three of the letters. The correlation curves also show, however, strong tendencies to relate two unrelated bits over long periods at quite high levels. These seem to be due to biases among the responses which are not yet learned. In particular, if we examine her statement about the placement of single capitals, we indeed find that for the overall experiment, in the cases that 110 was missed (the single capital is on the bottom in 110) 29 out of 44 were given responses 011, 100, and 101, the three <u>bottom</u> responses. For 011, which has the capital in the middle, 13 of 26 wrong responses were given as 010 or 110, the two <u>middle</u> responses. We should expect 101 would have given her the most difficulty, in this regard, as she had learned the three <u>top</u> responses and knew that they were not 101, the case with the capital at the top. Indeed, response 010, which was as high as she could go under the circumstances, was chosen more often than 101 itself. Thus, we have a rather clear example of the persistence of an <u>a priori</u> idea of the compatibility of stimulus and response interfering with the learning process.

c. Verbal Report

A few other items in the verbal report seem worthy of mention. Our hypothesis that the subject was looking at the structure of the response space in terms of the position of the first learned response is supported by the use of such expressions as "the square to the right of the top" or "the square to the left of the top" as well as by more uncertainty in speaking about the elements farther removed from the top square.

The reconstruction of the decision structure in Figure 4-1b is given further support by the way the subject described the rule she was using. Every response was given in terms of the number and placement of capitals--except 100, which was described as "middle small" in contrast to "middle capital" for 011. 000, for instance, was described as "all capitals," and 111 was described as "no capitals." Although 101, for example, was described as "top capital," 010 was not described as "top small," but as "bottom 2 capitals." Unfortunately, the experimenter asked for the rules in order starting with 000, so we have no way of knowing what order the subject would have given them in if she had responded freely. We do, however, have additional confirmation of our hypothesized order from the logical inconsistency which would result if, for example, the rule "top two caps" were applied before the rule "all caps." Confirmation of the identification of the strategy as serial is given by the subject's statement that "I learned them in order." Unfortunately, she did not state in what order she learned them.

Generally, the subject expressed the feeling that the experiment was "disturbing" and "threatening", particularly at the beginning. As she began to "get some things right", however, "it got to be more fun." She was reluctant to change the part she was looking at, because it had "worked" so well so far.

2. Second Experiment (Experiment 2)

This experiment was marked by a very rapid start in which the subject was already far exceeding chance before the first 20 trials were finished. After reducing the choice to adjacent pairs of responses, however, she entered a long period of seemingly fruitless exploration--being unable to resolve the pairs. Then, one by one, the pairs were resolved, though it seemed to get harder with each successive pair. Finally, after about 750 trials, the last pair seems to have been sufficiently well mastered to permit reaching the terminating criterion. After such a fast beginning, this experiment turned out to be much longer and more difficult than the first--even though they are informationally isomorphic.

One interesting feature of this experiment is the extreme steadiness of the overall average response times--even though the response times to different stimuli vary substantially over time.

a. Decision Structure

The set of rules learned for this experiment were not nearly so explicit and non-redundant as for the first experiment. As in the previous experiment, the subject quickly confined her attention to three adjacent character positions, this time, the three across the top--1, 2, and 3. In this case, however, there was no available complete rule which used only capitalization of these three letters. (The subject could have used such a rule had she chosen the bottom three positions--6, 7, and 8--but 1, 4, and 6, the three she had used previously, all reflected bit A and would have been of very little use in this experiment.) The best reconstruction of the final decision structure is shown in Figure 4-2. Examination of this structure shows that it is much more redundant and complex than the structure she learned for experiment 1 (Figure 4-1b).

Response 111 was not actually learned up to the capture criterion, but apparently only because insufficient opportunities were available to get the four correct responses before the end. Tabulation of response times, capture points and order learned yields the following table:

Response	Time	Capture	Order	Pair Capture
000	1.40	229(303)	1	19
001	1.40	304	2	19
100	1.75	442	5	96
101	1.85	447	6	96
011	2.00	358	4	106
010	2.10	350	3	106
111	2.25	(750)	8	108
110	2.35	742	7	108

Looking at the capture points and response times, the strong pairing is immediately apparent. The only exception seems to be 000-001, but when we examine the region between 229 and 304, we find that 000 was presented much more frequently than 001. 000 meets the criterion for capture, but only because a single case of 001 is correctly identified. Then, 000 and 001 are confused several more times, until at 303-304 they are definitely both captured. It appears that they were partially captured at 229-231 and partially at 303-304.

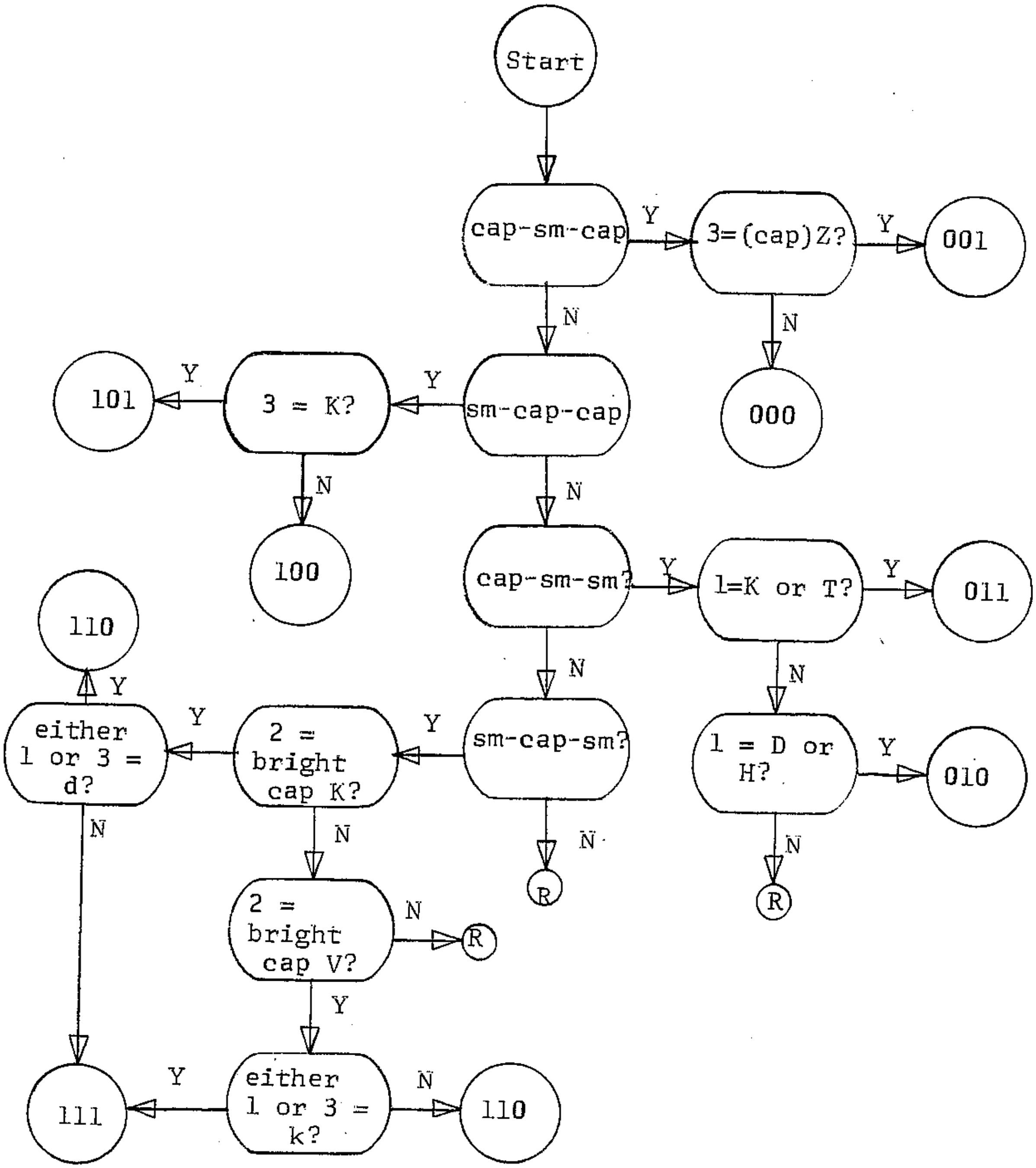
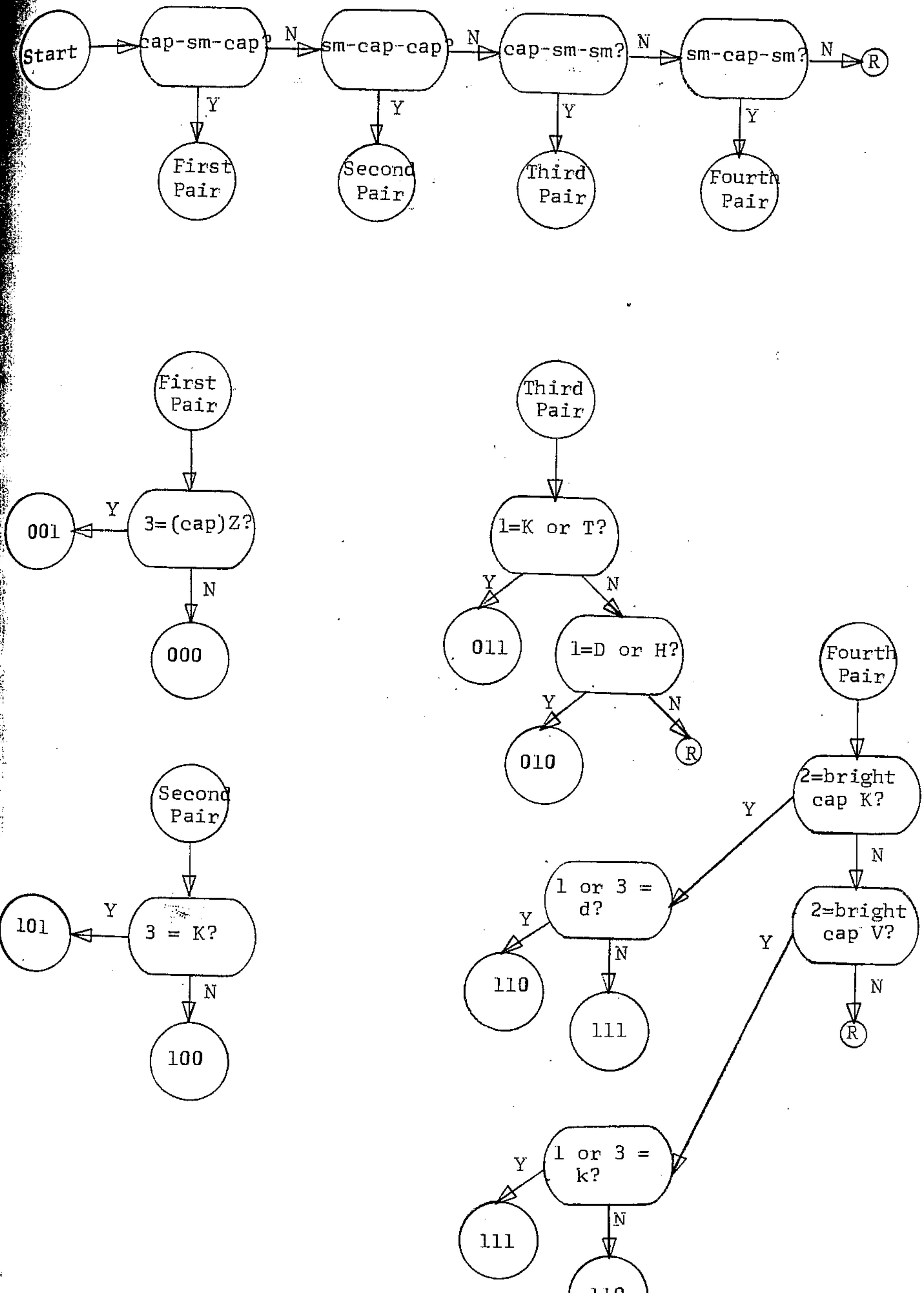

Start
cap-sm-cap
Y
3=(cap)Z?
Y
001
N
000
N
sm-cap-cap
Y
3 = K?
Y
101
N
100
N
cap-sm-sm?
Y
1=K or T?
Y
011
N
1 = D or H?
Y
010
N
R
N
sm-cap-sm?
Y
2 = bright cap K?
Y
either 1 or 3 = d?
Y
110
N
111
N
R
N
2 = bright cap V?
N
R
Y
either 1 or 3 = k?
Y
111
N
110

Figure 4-2

Examination of the tabulations and other evidence shows that this pairing resulted from a complete mastery of bits A and B before C was learned at all. In other words, the subject learned a decision structure which first isolates four pairs and then tackles the resolution of each pair as a separate problem. Viewed in this way, the decision structure may be represented as in Figure 4-3. Breaking the decision structure into two phases seems justified on the basis of several types of evidence. First of all, test trials show that the subject was perfectly able to get bit A correctly when either position 1 or 2 was deleted; but she continued to the end insisting that "all the capitals are always important" and always verbalizing in terms of a three capital string. This evidence suggests the existence of the serial string as an entity which cannot be consciously decomposed by the subject. Also supporting this view are the occurrences of certain errors after such errors had long disappeared through the mastery of the pairings (which can be expressed in terms of "pair captures," using the same kind of capture criteria, but not considering interpair confusions as errors). For instance, long after 101 has been mastered, it is confused several times with 010. In all of these cases, the 101 was displaying an H in position 2, and "H" is one of the decision criteria for 010 and not for any other response. In other words, 010 and 101 seem to be "interfering", which they could not do if the decision structure were a firmly united tree. Similar interference--but much more extensive and pronounced occurs between the pair (011, 010) and the pair (111, 110) seemingly based


Start
cap-sm-cap?
N
sm-cap-cap?
N
cap-sm-sm?
N
sm-cap-sm?
N
R
Y
First Pair
Y
Second Pair
Y
Third Pair
Y
Fourth Pair
First Pair
3=(cap)Z?
Y
001
N
000
Third Pair
1=K or T?
Y
011
N
1=D or H?
Y
010
N
R
Second Pair
3 = K?
Y
101
N
100
Fourth Pair
2=bright cap K?
Y
1 or 3 = d?
Y
110
N
111
N
2=bright cap V?
N
R
Y
1 or 3 = k?
Y
111
N

on the multiple appearance of the letter K in their decision structures. It is as if the subject is concentrating on the sub-decision and "forgets" on which primary branch she is working.

b. Strategy

As we might expect, the subject seems to have been looking for a "similar" rule to the one which had been successful in experiment 1, and seeking it by similar strategies. The significant aspects of the first experiment were the following:

1. The narrowing of the stimulus field to three (adjacent) positions.
2. The dependence on capitalization clues.
3. The search for conspicuous stimuli.
4. The ordered isolation of responses.

Apparently, the subject could easily eliminate the same three positions previously used--1, 4, and 6--since they only showed two variations--cap-sm-cap and sm-cap-sm. On the other hand, she apparently made her decision to narrow to the top three without much searching, for the bottom three would have permitted her to make a complete capitalization rule which, though different from the rule of experiment 1, would have undoubtedly made the entire experiment quite simple for her. Perhaps she did not notice this possibility because the conspicuous stimulus case (all caps) did not correspond to the conspicuous response case (top square) as it had in the first experiment. The capture of the 000, 001 pair at the 19th trial indicates quite clearly that the subject had decided to concentrate on 000 as a beginning strategy. In fact, after seeing only two instances of 000--on trials 3 and 4--she correctly identifies the 000,001 pair on trials 5, 7, 9, and 14--and only missing capture at trial 5 by responding 000 to 110 on trial 18. From that time on, she responded 000 to every 000 and 001 except one until trial 231 when she made the response 001 for the first time--and made it correctly. In fact, response 001 was never made incorrectly in the entire experiment.

For a while after isolating the 000, 001 pair, the subject seems to be concentrating on responses 010, 100, and perhaps 110--making those in great preference to the others. Since these are the three responses perpendicularly related to 000, it seems likely that she already had the idea that there was a pairing involved and was trying to master it first. This bias becomes even stronger after the pairs are captured--for instance, response 101 is never given between trials 191 and 447 (its capture point)--and we have already seen that 001 was handled in the same way. The other pairs are not quite so extreme, but 010 shows ratio of about three to one over 011. Only 110 and 111 do not show this bias so clearly, perhaps because 111 is also associated with 000 by adjacency--which tends to conflict with the perpendicular association of 110.

As in experiment 1, the subject seems to take advantage of local structures which conserve memory, both in capturing the pairs and in finally resolving them. A speciman of a pair capture is illustrated by the structure:

Trial	V	R
95	101	010
96	100	101

Before this time, 010 was consistently being given as a response to 100 or 101. Clearly, if 101 and 100 are considered to be the same stimulus (sm-cap-cap), this structure corresponds to the P type structure. Each point of pair capture shows such a structure--either N or P type. Further, the capture point of 010 (and 011, since it was their resolution which prevented capture) looks like this:

Trial	V	R
348	011	011
349	010	011
350	010	010

with the added feature that 348 and 349 had identical noise bits. The capture of 110 (and 111) was at a point where the <u>identical</u> stimulus, 110-0010 was presented on two successive trials. The capture of 000 was immediately preceded by a series of 18 trials in which 001 was presented six times and 000, four times. The capture point of 001 is immediately preceded by a 000; but the capture point of 100 (and 101) does not show any such memory-conserving structure. The capture of 100 and 101 may represent more of a "reasoning out" than a "noticing," especially since by that time, 011 had already been captured (giving the element "K") and so had 000 and 001 (giving the element "position 3").

Once the subject had become aware that the capitalizations of 1, 2, and 3 were not going to provide sufficient information for complete resolution, she was faced with the choice of where to seek further information. She apparently chose to try extracting information from the other features of the same positions rather than keeping her strategy of looking only at capitals--which would have required the examination of other positions. At a pause about halfway through the experiment, she said, "I'm confining myself to the top three letters, and I think the rule is there. It may involve something else, but I don't want to go looking at something new." Apparently the first thing she noticed was the frequent occurrence of a pair of capital Z's in symmetrical positions on both sides of the top row. This structure occurs for 001, but only when noise bit 2 = 0. When noise bit 2 = 1, only the right-hand Z appears. Indeed, if we examine the region between 229-231 and 303-304, we find that 001 is presented six times. The three times bit 2 = 0, it is gotten correctly; the three times bit 2 = 1, it is missed. Here we have clear-cut support for our conjecture that only part of the distinction between 000 and 001 was captured at 229-231. Then, at 303-304, the criterion for 001 is widened to include the single Z case as well as the double Z case. Actually, this apparent widening is a narrowing, since the subject recognizes 001 as an exception to 000. Thus, when 000 is narrowed, 001 appears to be widened.

At the pause, the subject made the following statements:

Experimenter: Can you describe what rules you do know now?

Subject: Well, number one (000) is always cap-small-cap with one exception--if the last capital (position 3) is a Z, then it is a two (001).

Experimenter: What else?

Subject: Number two is capital Z, small letter, capital Z.

Here, then, we have an interesting relation of speech and behavior. When the subject describes 000, she gives the correct rule (which she has apparently been following for some time). When, however, she is asked to describe a rule for 001, she attempts to make a positive identification and reverts to a partial rule that she used on the way to discovering the whole rule (as an exception to 000). Her only correct understanding of 001 is in relation to 000, and her speech reflect the history of obtaining that understanding.

When we examine the other capture points in more detail, we find similar occurrences. The capture of 010, 011 and 349 turns out to be another partial capture that accidentally meets criterion. Before that point, 010 is always chosen to represent the pair. After that point, 011 is correctly differentiated if bit 2 = 1 (which puts a K in position 1) and always wrong if bit 2 = 0 (which puts a T in 1). The capture seems to have been completed at 468. In the verbal report at the pause (which is right in this area), the subject reports that 011 is distinguished from 010 by the presence of a dim capital K (which she wrongly asserts is in the middle position, 2) and "may be a T also, but I'm not sure." She also states that "K screws up the whole rule," by which she may be alluding to the fact that a K can actually occur in two places for this pair--and in position 2 is not distinguishing. By the end of the experiment, she was both clear and correct on an initial K or T distinguishing 011 from 010, but also knew how to choose 010 positively, by D or H.

The pair 100, 101 seems not to have given as much trouble initially, since 100 never has a K and 101 always has one in position 3. Thus, the capture point is clear and distinct, and the reconstructed decision tree is simple and straightforward, though 101 is sometimes confused with 010--possibly because both show at times either a VM or an HK. In 010, however, the pattern is VM- or HK-, whereas in 101, it is -VM or -HK. Thus, again, we see other similarities overriding the subject's sense of position in the stimulus array.

The pair 110, 111 presents a more confusing and confused picture. In the first place, this pair never really develops the choice of one as representative and the other as exception. Instead, between 323 and 353 the subject seems to have gotton the notion that a V in the middle identified 110 and a K in the middle identified 111 (after sm-cap-sm separated the pair from the others). We find that after that point, she consistently identified the stimulus as 110 if bit 1 = 1, and as 111 if bit 1 = 0. Eventually, she seems to straighten out the true distinction, but she has constructed a rule which is unnecessarily complex because she first makes a completely irrelevant distinction based on the middle letter. As she expresses it at the end, "I've resolved one thing on the 7 and 8 (110 and 111) squares. It's 8 (111) if the middle is a bright capital K, except if one of the small letters is a d, which makes it a 7 (110). It's similar for a bright V in the middle which would normally be a 7 (110) but the conditioning letter makes it an 8 (111)." The "conditioning letter" is apparently a K, which she "once thought conditioned everything" and seems to have been the source of much confusion, since it appears at some time or other in four of the eight stimulus classes and slows response times when it appears. The top row letter patterns for this pair are:

110	111
hKd	tKk
dKd	kKk
hVd	tVk
dVd	kVk

Obviously, V is no more "normally" a 110 than it is normally a 111, but the subject has fixed on this idea and cannot seem to shake it off. This conclusion is supported by the correlation plots, which show generally broader, shorter peaks than in the first experiment and by the $\bar{g}$ and f plots, which show very little, if any, gestalt. This is the kind of picture we might expect if the subject for some reason breaks down the entire problem into sub- and sub-sub- cases, for improvement will then be in one bit at a time and evidence to confirm or deny such narrowly defined hypotheses will be slow in coming.

c. Verbal Report

In describing her rules, the subject seems to have improved a great deal in verbal clarity between the middle and the end of the experiment. Her movement, however, was always toward clarifying the rules, rather than simplifying them, and she seems never to have collapsed the decision structure but only refined and elaborated it. Since this was the last experiment for her, the experimenter offered to answer her questions at the end of the formal verbal report. She immediately asked what was the "real" rule. When she was told that any number of rules were equally good, she expressed dissatisfaction and wanted to know if some were not better--simpler--than others. She was then told that, for example, certain single letters could have determined the entire rule. She denied that this was possible. She was told that, in fact, the letter in position 3, which she had been using, was sufficient to determine the rule. She vigorously denied this possibility. Finally, the experimenter turned the experiment back on for her and asked her to look. After about eight or ten stimulus presentations, she let out a loud and wistful "OH", and was immediately able to see the rule based only on position 3. Upon checking her test trials, it was found that she was already able to operate with only position 3, but the knowledge that she could had definitely not crept into her verbalization.

In the formal verbal report, she expressed the feeling that she was often angry with the machine, because--contrary to its behavior in the first experiment--it did not seem to want to cooperate by giving helpful cases to test her hypotheses. In the light of our reconstruction of her two strategies, the reason for this difference seems perfectly clear--with the fine resolution of the second experiment, much more explicit (and thus much less probable) events would be needed to seem helpful. She also expressed that the second experiment seemed "three times as long and twelve times as complex."

The subject said that she sometimes "had the feeling that there were too many test trials, so it got so I didn't care what I answered on tests, at times." If, however, we examine the test trials, there is no evidence whatsoever of any deterioration of performance on them--other than what can be explicitly accounted for because of the deleted positions. Furthermore, there is not even any indication of either a speeding up or a slowing down during the test trials. She also said that she "wasn't as upset at the beginning of this experiment, but as the difficulty mounted, I got much more discouraged." Here, too, where seems to be no obvious behavioral correlate.

B. Subject 2

1. First Experiment (Experiment 1)

The first striking feature of this experiment is how different two subjects can be. The second is that this subject could not complete this task and after 1500 trials, asked if he could give it up. In many experimental situations, the cases where the subject just gives up are discarded--perhaps because they cannot sensibly be lumped with the statistics from successful experiments. In our experiments, however, we can apply our analysis techniques just as well to failures, so that every experiment is meaningful. Moreover, failures in problem solving--because of their rarity--may shed more light on the critical factors in the process.

Of course, part of the difficulty the subject experienced can be attributed to this being the first experiment he performed. Indeed, he was a very suspicious subject, and it took him a long time to really get down to work on what we might consider the obvious task. At the end of 495 trials, he asked for a pause. At that time, he said, "At first I wanted to disregard the patterns altogether and just look for a sequence of rights and wrongs regardless of what was showing." He also commented that he was looking for sequences conditioned on his previous responses. The one thing he had noticed was the sequence of deletions in the test trials. After a few words of reassurance from the experimenter that there were no tricks to this, he went back to work and shortly thereafter began to show signs of learning within the confines of the experimenter's conception of the experiment.

Actually, to say that he failed is only partially correct, as he eventually mastered bits A and B. Essentially, his failure was to isolate and master the third component of the problem, and our analysis should be directed toward understanding this failure.

a. Decision Structure

Since part of the experiment was mastered, it should be possible to reconstruct as much of a decision structure as existed for that part. Although the data listing says that responses 000, 100, 010, and 011 were captured, even a casual examination reveals that their "captures" amounted to no more than fortuitous successions of responses which, since two of the three bits had already been mastered, easily met the criteria for capture. We can, however, look for the capture of these response pairs, which are precisely those captured by subject 1 in her second experiment, 000-001, 010-011, 100-101, and 110-111. For these, we get the following table:

Pair	Capture	Leading Member
000-001	782	000
010-011	1116	no preference
100-101	880	100
110-111	1116	111, then 110

As with our previous example, this subject shows a marked preference for identifying the pair with one member of the pair, seemingly on a very similar basis--perpendicularity and closeness. The 110-111 case is unusual in that the subject switches his preference about 200 trials from the end. These preferences, incidentally, seem to account for most of the slight correlation with bit C that appears late in the experiment. Over the period when that correlation occurred, the preferred values predominated over the others by a ratio of 53/47. Thus, simply by exercising this response bias, the subject would have gotten a 53 percent correlation with bit C, compared with the 56 percent he did get over that interval. Inasmuch as he never learned C, we cannot know whether this response bias would have been associated with the decision structure as it was with our previous subject.

Attempts to reconstruct the partial decision structure are first frustrated by two features of the data--an amazing inconsistency in responding to the same stimulus and seemingly very little influence of the test trials on behavior. On more careful examination of the test trials, however, we get our first clue. Over the last four test trials--those after pair capture had taken place--position 5 seems to have had the most influence on the subject's responses. Out of 12 times 5 was deleted, 7 led to errors involving bit A. Furthermore, those seven stimuli all had bit A = 1, while the other five had bit A = 0. Clearly, position 5 is used to determine bit A when it is a 1, but not when it is a 0. When we examine the response times in these two cases, we find that the average is 2.20 when bit A = 0 and 2.55 when it is 1. This slight difference may not be significant, but it leads us to ask whether the other test deletions, though they do not show the clear error structure of 5, show some systematic effect on response time. The following table shows what was found:

Test Position	A = 0	A = 1	B = 0	B = 1	A	B
1	2.50	2.35	2.25	2.50	.20	-.05
2	3.25	2.40	2.95	2.75	.90	.40
3	2.20	2.50	1.90	2.95	-.25	-.85
4	2.30	2.50	2.60	2.15	-.15	.65
5	2.20	2.55	2.55	2.20	-.30	.55
6	2.70	1.80	2.50	2.30	.95	.40
7	1.80	1.80	1.75	2.00	.05	-.05
8	2.25	2.20	1.75	2.65	.10	-.70

The numbers in the columns A and B are the time differences, adjusted for the fact that the overall mean times differ--(bit A = 0, 2.10; bit A = 1, 2.15; bit B = 0, 2.00; bit B = 1, 2.20). If these times be taken to indicate relative disturbance of the subject's decision structure, we can definitely conclude that:

1. Position 2 contributes to the determination of A = 0 cases.
2. Position 3 contributes to the determination of B = 1 cases.
3. Position 4 contributes to the determination of B = 0 cases.
4. Position 6 contributes to the determination of A = 0 cases.
5. Position 8 contributes to the determination of B = 1 cases.

In addition, it is probable that:

6. Neither 1 nor 7 contribute to any cases.
7. 2, 5, and 6 make some contribution to B = 0 cases.

If we look at the structure table of this experiment, one hypothesis fits these conclusions in a striking way, namely, that the subject is determining responses according to the pattern of bright letters he sees in each stimulus. The only evidence to suggest that other information is being used is our conclusion 7, which might indicate he is taking some notice of letters--in 2 and 6--and, more probably, some notice of the capitalization of position 5. At first thought it seems strange that the value under A in position 5 is only -.30, but this fact is quite understandable when we consider that if the subject is only looking at the bright letters, he probably considers a deleted letter as not different from a dim one.

We can carry this examination one level deeper by seeing whether or not there is any linkage between bits A and B for any position. The only result of significance is that for 5, the 100-101 pair has an average time of 3.05 and one bit B error out of five cases while the 110-111 pair has an average time of 2.20 and no bit B errors out of eight cases. Adjusting for the difference in average times for these pairs, we get a difference in time of 1.05 seconds. Although based on only 13 trials, it seems quite probable that this difference is significant.

Further investigation of the data reveals no more clues as to the decision structure. What this may well indicate is that the subject does not--except for pair 100-101--have a single way of determining each pair, for this tactic would make him relatively impervious to the effects of test trials. For instance, our evidence is consistent with a decision procedure couched in the vague terms of Figure 4-4. Such a decision procedure would explain all of the positive and negative facts available, such as:

1. The key role of position 5.
2. The order of mastery of the pairs, including 01x and 11x simultaneously.
3. The general lack of errors in tests and the specific confusions connected with 5.
4. The lack of sharp, clear points of mastery of certain aspects.
5. The inability of the subject to give any better verbal description of his rule other than that he uses "general patterns of brightness."

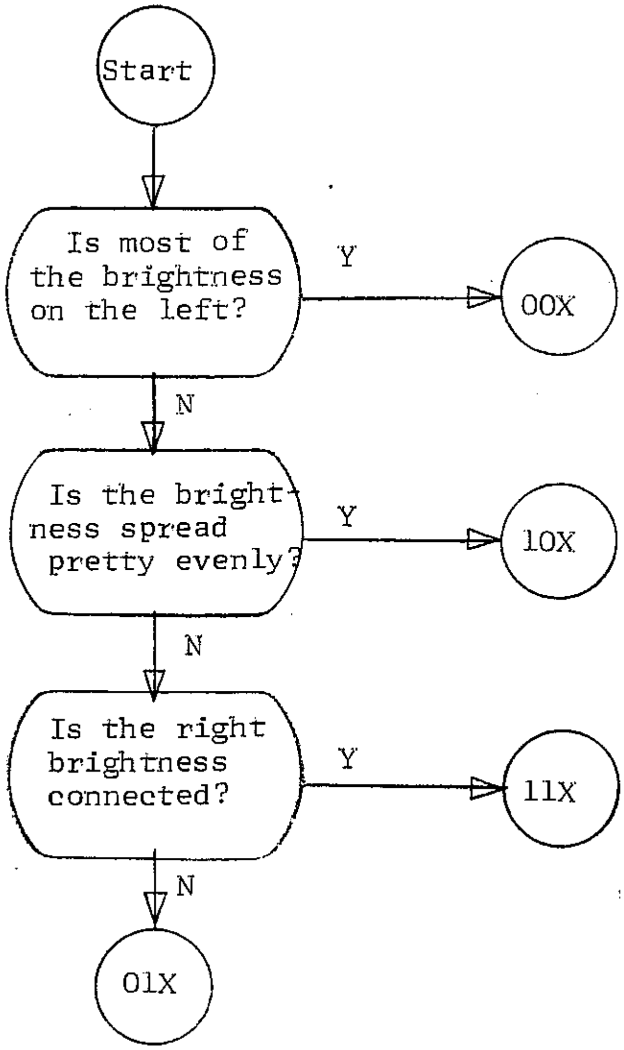

Start
Is most of the brightness on the left?
Y
00X
N
Is the bright-ness spread pretty evenly?
Y
10X
N
Is the right brightness connected?
Y
11X
N
01X

Figure 4-4

b. Strategy

After the initial 500 trials--which seemed to have served to relieve the subject of some of his suspicions--his strategy may be said to have begun. The significant aspects of that strategy were the following:

1. The simplification of stimuli by concentrating on brightness patterns.
2. The grouping of stimuli into recognized classes by these brightness patterns and then associating them with responses.
3. The inability to use information other than what was in the bright letters.

When we look at the structure of this experiment, we see why this strategy gave the subject so much difficulty. First of all, unless the three bits are represented in the brightness patterns themselves, bits not in the pattern must be picked up in two different places, at least. This immediately complicates any potential rule and also lessens the frequency with which relevant cases will be seen. Furthermore, when one of the brightness bits is a noise bit (in this case, bit 3 in position 7), even the initial classification of responses into pairs is at least doubly complicated.

There is some evidence that he tried to discover the third bit--possibly with even slight success--but the correlations clearly show that he was repeatedly correlating it with the brightness of some letter or another, rather than with information even within the bright positions. Although he may have tried--as he said--using information other than brightness, the task may have been too difficult for him. He seems never to have considered using another master strategy. Finally, he just gave up.

In general, this subject does not seem to have been able to use local features of an easily identifiable kind. One exception to this is the capture point of the 100-101 pair. Here we find the highly conspicuous and unlikely structure of four 100's in a row, after the last of which he had mastered 100-101. Of course, it is difficult to imagine just which features will be conspicuous or helpful because the picture of the decision structure is so fuzzy.

One other point which is striking is the area immediately surrounding trial 722. It is at this point that the subject seems to have "captured" the division between "top" and "bottom" stimuli on the one hand and "side" stimuli, on the other--or, in other words, learned bit B. Before this point, he seems to have been performing at about 60 percent on bit B. Immediately after this point, he attains a string of 21 consecutive correct bit B's, and thenceforth maintains at least 90 percent correct B's. In fact, the only bit B errors that occur after that point have bit 3 = 1, meaning that position 7 is bright, which is in agreement with our general reconstruction of the basis for the decision structure.

In a similar manner--although not marked by such a sharp onset--the subject seems to have eventually cleared up his confusion about the involvement of 7 in the left-right discrimination.

Here, when bit 3 = 0, he tends to respond 01x to 11x, and when bit 3 = 1, the mistakes are in the other direction. Position 7 being dim tends to "disconnect" the pattern, while 7 being bright tends to "connect" it, thus this confusion pattern further supports our analysis.

With respect to bits A and B, then, the subject's strategy does not present quite the confused picture it once appeared. His strategy for bit C, however, seems impenetrably locked into the data. If any clue at all exists, it is in the verbal report where he talks about "sequences" again, perhaps reverting to his early idea that bit C, at least, did not depend on the pattern but occurred in certain time patterns or in relation to whether he was right or wrong on the previous trial.

c. Verbal Report

From our reconstruction of the subject's decision structure and strategy, we might predict that he has a fairly accurate picture of the response space--divided up and down or left and right. We might further predict that his conception of the structure of the stimulus space was quite poor, at least with respect to exact positions of things. Both these predictions are borne out in his ability to discuss what he is doing. He can easily describe the pairing of responses and distinguish one pair from another, but he can give no clear picture at all of how he identifies the stimuli for each pair.

He does, on the other hand, have a remarkably clear idea of how long he has been working and how well he has been doing.

For instance, he estimated that he had used between 1200 and 1500 trials, the actual number being 1561. At the time his performance on bit C was about 56 percent, he stated: "I'm probably doing a little better than fifty percent now." Earlier, he had said, "Right now it's getting to be fifty-fifty," and that was just at the point he had mastered bit B and was just about to master bit A.

2. Second Experiment (Experiment 4)

The original plan for this subject was to follow experiment 1 with experiment 2, as had been done for the first subject. In view of the discouragement expressed by the subject upon giving up on experiment 1, it was thought advisable to let him try as easy an experiment as possible. He had said that he was only using the pattern of bright letters, and a quick look at experiment 2 revealed that--like experiment 1--one of its value bits was not in the brightness pattern. Thus, the sequence was changed, and he was given experiment 4--which has each value bit in the brightness pattern at least twice--under the assumption that this would give quick results with the use of the same strategy.

Indeed, there was a marked difference between this and the first experiment. Quite quickly he was performing much better than chance--and, furthermore, he was getting whole responses correct, not just breaking them down into pairs. After his initial success, however, he slowed down quite a bit and seemed to be having a great deal of trouble mastering the remaining parts of the task. Finally, after 817 trials, he finished--feeling much better than he had when he began.

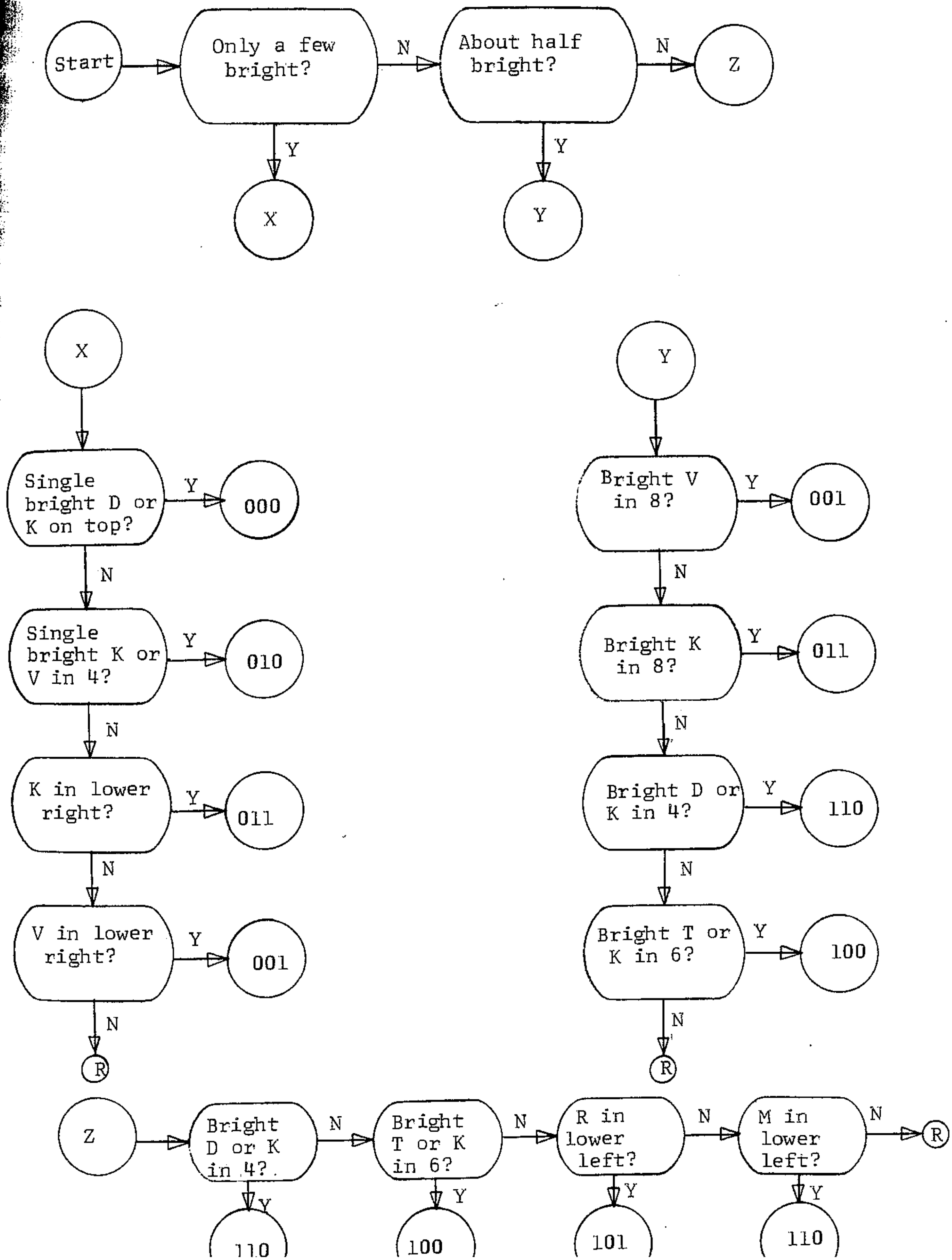
a. Decision Structure

Since the subject had been known to be using only brightness in the first experiment, it was expected that he would find one of the number of simple combinations of bright and dim letters which would identify all responses--just as subject 1 had done with capitalization in experiment 1. Unfortunately, this did not turn out to be the case, and the analysis of the decision structure turned out to be perhaps the most difficult and ambiguous so far. The difficulties of analysis seem to be related to the following facts:

1. The distinction between two of the responses (101 and 111) turns out never to have been learned at all.
2. Certain responses, particularly 100, seemed to have been used to express the response: "I don't recognize this one."
3. Analysis of the critical positions revealed only that the subject seemed to be using different positions for almost every response.
4. The rather too rapid ending limited the amount of test data available.
5. Several of the responses seemed to be recognized in at least two different ways.

Eventually, by analyzing each response in its development, a clearer picture emerged. Very early, after no more than 70 or 80 trials, 000 and 010 were recognized as belonging to a distinct group of stimuli. Soon after, 011 seems to have been added to this group; and, finally, after about 250 or 300 trials, 001 was added as well. Actually, only some cases of 011 and 001 were added, the other cases becoming part of one half of a second grouping which then split in two. Roughly, what was found was that some cases of 001 and 011 were being confused with some cases of 000 and 010; other cases were being confused with some cases of 100 and 110. On the other hand, cases of 100 and 110 were being confused with some cases of 101 and 111. Certain cases of 000 and 010, and of 101 and 111, were not confused except with one another. When we look at the distinctions between the cases, the best explanation seems to lie in a decision structure such as is shown in Figure 4-5, which has two distinct levels, which may or may not be separated as shown.

The main distinction between the divisions seems to be in the value of noise bit 2, which determines the brightness of position 8, and thus throws marginal patterns from one group to the other. Generally, the vague descriptions in the diagram seem consistent with the method of division as shown in the data, both in logic and in lack of rigor. The first part of the structure labeled X seems to have been evolved or evolving before the division of the rest of the stimuli into Y and Z developed, but the second two decisions were added after the first two of Y were formed. Later, as the remaining cases of 001 and 011 were being worked out, they were confused with 000 and 010, in just the way we would expect if the subject had momentarily forgotten which branch, X or Y, he was currently working on--for the V and K in position 8 only identify


Start
Only a few bright?
N
About half bright?
N
Z
Y
X
Y
Y
X
Single bright D or K on top?
Y
000
N
Single bright K or V in 4?
Y
010
N
K in lower right?
Y
011
N
V in lower right?
Y
001
N
R
Y
Bright V in 8?
Y
001
N
Bright K in 8?
Y
011
N
Bright D or K in 4?
Y
110
N
Bright T or K in 6?
Y
100
N
R
Z
Bright D or K in 4?
N
Bright T or K in 6?
N
R in lower left?
N
M in lower left?
N
R
Y
110
Y
100
Y
101
Y
110

001 and 011 if 000 and 010 have already been excluded. This situation is quite similar to that of the second experiment for subject 1.

This diagram seems to account for all the confusing facts. The "distinction" between 101 and 111 is no distinction at all, it leads to incorrect choices exactly half of the time. Response 100 was apparently used to identify classes Y and Z while they were still lumped together. At various times, the subject uses each position in determining the major division; and in the minor division, he uses positions 2, 4, 5, 6, and 8, at least. The rapid ending is explained by the coincidental occurrence of just those 101's that would be identified as 101's and just those 111's that would be identified as 111's by the given rule. One further confusion seems to have resulted from the experimenter's expectation that only bright letters would be used, but the evidence indicates that 001, 011, and 110 were at least partly determined by dim letters.

b. Strategy

The subject seems to have started out looking at brightness patterns, as he had in the previous experiment. In that experiment, however, there were never any patterns in which only a single letter was bright. In this one there were, and the two responses for which these occurred were the first for which he had any success, namely, 000 and 010. Having been able (in contrast to his first experiment) to consistently identify at least some individual responses--rather than pairs--he seems to have set upon a response isolation strategy.

Each response followed a roughly similar course on the way to mastery, though, naturally the different stages occurred at different times. 011 is a typical specimen. By about trial 80, it is at least partly recognized as distinct from some other patterns, for it is responded to in a consistent manner. By trial 188, the subject is able to identify those cases of 011 that have a bright K in position 8, and perhaps a bright T in position 1 as well. In learning to apply this rule, he apparently first uses the letters without regard to their exact position, for he also starts responding 011 to other stimuli that show both a bright T and K. After a few such mistakes, he learns to narrow his criterion by discriminating on the basis of the position of the critical letters. His next improvement comes at about 300, where he starts to be able to identify cases in which the K is dim, as well. Here, he may be using the presence of a D or an R, along with the T, for as soon as he begins to show improved identification, he starts making the mistake of responding 011 to cases of 001 which show a D or an R. Then he seems to have dropped this criterion because he stops making responses to 001 and also reverts to making errors when 011 appears with a dim K in the corner. Then, at about 390, he starts improving again, this time apparently on the basis of a K in the corner, for he also starts responding 011 to 010, which had been mastered long before. After a few examples of this kind, he seems to have cleared up the distinction and have mastered 011, except for two reversions when he is concentrating on 100 later on.

When we go through the details of the learning of each response, we can often identify clear points at which some change of behavior occurred. Of the 23 points so identified in this experiment, 11 of them had the P or N structure, and several others had a single intervening irrelevant stimulus. At least in these cases, then, the subject seems to be making use of memory aids available to him.

Another interesting feature we have not previously observed is found when examining the response times to 000. 000 appears in two forms, from this subject's point of view: a single K or D in position 2, or the K or D accompanied by a V in position 8. At first, when 000 has been successfully identified, the double letter forms are systematically responded to faster than are the single letter forms. Then, as other stimuli are sorted out from 000, the double letter forms start to become distinctly slower. Both effects are of the order of .5 second. Finally, by the end of the experiment, the times become essentially the same.

c. Verbal Report

Certain parts of the verbal report confirm our reconstruction of the experiment. The subject verifies that he is using letters and that sometimes he did, indeed, use dim letters. His rule for 011 he describes thusly: "Always when there were several lit up and the K in the bottom right corner was lit up or not lit up--either one--it pointed to the square on the right (011)." Of course, he need not have even referred to brightness of the K, unless he had at one time distinguished the two cases. His ability to identify precisely the location of the K is also in agreement with our reconstruction, especially as in other cases his description often went like this:

S: Then when two bright letters occurred, I had a ready pattern for that (010).

E: Where were the two bright letters?

S: Where were the two bright letters?

E: Anywhere?

S: No. They were usually a...I don't know how to describe the positions of them.

Sometimes his verbal description is just wrong, as when he talks about identifying "some patterns" using a "z or v" in position 8. In no stimulus does a z ever appear in position 8. Or, again, he says that "the top (000) was always lit by a single letter d." If by this he means that a single d always represents 000, he is correct--and that may be the way he first noticed it. But if he means this to be his only rule, he would be wrong three-fourths of the time 000 appeared.

His description of his "rule" for 101 and 111 explains why they were confused. He knew that "when there were two r's lit, one upper left and one lower left," it was a 101. Otherwise, apparently, he chose 111. This rule turns out to be completely equivalent to the rule we deduced; that is, with the evidence we had (since position 1 was never deleted in a test of 101) we could not have distinguished between them. He was perfectly aware, incidentally, that he was not actually able to distinguish 101 and 111. Another place where our analysis was probably wrong was in his rule for telling 100 from 110. He said, "If the d or r (in position 4) wasn't lit...but the T in the left--lower left-- corner was lit, it was the bottom one. This rule only works in half the cases, but if he looks for a bright T anywhere, he will always be right on 100. This would also explain his confusing behavior when seemingly irrelevant positions were deleted.

He seems to have felt that K was a "fundamental letter." Perhaps this is what led him to start noticing dim k's. He also had the feeling that there were fewer patterns than in the first experiment (which, of course, is not true) and that "seemed to make it easier to figure out which...to associate the patterns with the squares."

3. Third Experiment (Experiment 3)

In many ways, this experiment seems similar to the first experiment the subject performed. In particular, he seems to have gotten half of the problem quite readily and then to have had increasing difficulty getting the remaining parts. Although this experiment terminated automatically, analysis reveals that the subject was quite far from mastering the rule, just as he was in the first experiment. Also, whereas the half right in in the first experiment consisted in being able to get each response right half of the time, in this experiment, the subject could get half of the responses right all of the time and the other half were indistinguishable from one another.

a. Decision Structure

The four responses which were mastered early--000, 010, 100, and 110--were distinguished according to the diagram which we saw in Figure 3-19. The decision structure for the remaining responses varied over the course of the experiment; but at the end it had the structure shown in Figure 4-6. This structure seems quite rigid--that is, it accounts for all of the subject's behavior over a fairly long span--but it contains a number of errors. First, the choice of 011 whenever position 1 = H utterly fails to distinguish anything that has not already been decided farther up in the tree and ensures that half of the 001's will be incorrect. Of the remaining 011's, half are correctly identified by the two bright capital M's, but the other half are always chosen as 011's. The use of Z and T to "distinguish" 101 from 111 is completely meaningless and results in each being wrong half of the time.

On the average, this structure will yield the right decision 25/32--or 78 percent--of the time. As it happened, the subject terminated the run by getting a sequence of 20 stimuli which this rule decided correctly.

b. Strategy

In a very short time, the subject succeeded in partitioning the stimuli into XX0's and XX1's. He then proceeded to

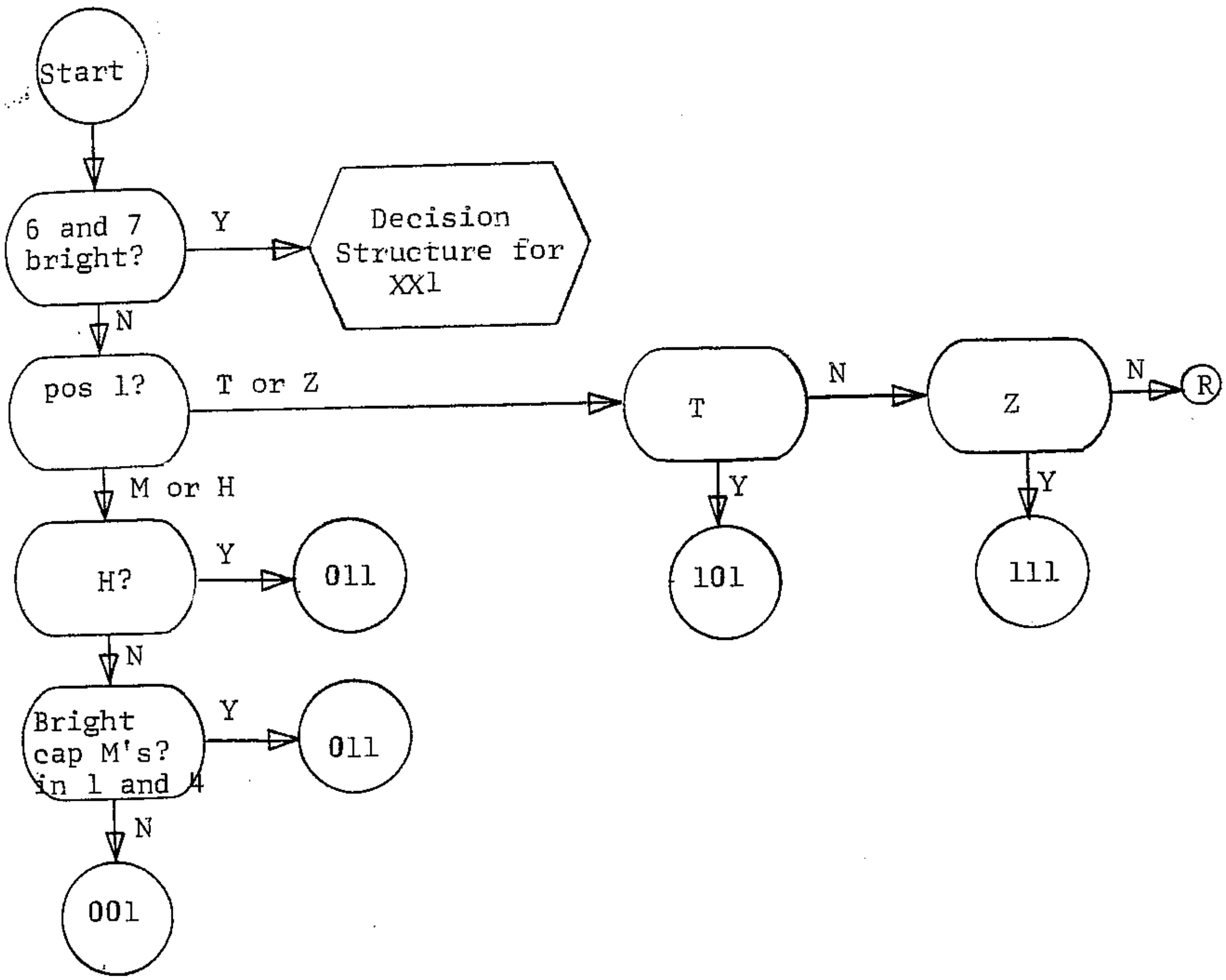

Start
6 and 7 bright?
Y
Decision Structure for XX1
N
pos 1?
T or Z
T
N
Z
N
R
Y
Y
101
111
M or H
H?
Y
011
N
Bright cap M's? in 1 and 4
Y
011
N
001

Figure 4-6

distinguish among the XXO's, showing the same processes he exhibited in the previous experiment--making his hypotheses wider and narrower until settling upon one of just the right size. After mastering the XXO group, he seems to be trying the same strategy on the others, but experiences much more difficult. Eventually he is able to decompose the set into OX1 and 1X1 sets (based on position 1, primarily) and seems to have been in the process of further decomposition when he accidentally terminated. The rate at which he was able to change the current decision structure seems to have been slowing down as the structure got larger.

One specific relationship between this and the previous experiment may have aided him in getting started--the intensity of position 8 was determined by a noise bit in both cases. In the previous experiment, he had learned to use the letter in position 8 even though it was dim, and he carried over this procedure into the new decision structure. Nevertheless, he still retains the use of brightness pattern for making the major division of the stimuli, leading him into the same sort of complex decision structure he has created for the previous experiments.

c. Verbal Report

One thing that does seem to have developed during the previous experiments is the subject's ability (or willingness) to verbalize about both his decision structures and his strategies. His concept of the spatial structure of the stimulus seems much improved--though he still is confused about the position and role of the plus sign. The lower right corner seems particularly well fixed in his mind, and several times he uses it as the origin of a system of referencing other positions.

At the time of the first pause, he expressed his rule for 110 and 010 in terms of two positive criteria--R or M for 110 and Z or V for 010. By the next pause, however, about 700 trials later, he says, "...it is either an R or an M for the left (110) and I don't care what it is for the other." Since the second test is indeed redundant, this may be a case of eliminating redundancy by dropping remote branches off the decision structure--perhaps as a way to conserve on memory requirements.

He again expresses the concept of a "key character," though this time he believes it to be an M. He is not as certain about it as he was in the previous experiment, however, and he says, "That's the trouble...I'm not really keying in on something." He was particularly adamant about having to keep dealing with the XXO patterns after they were learned, for he seemed to feel that they were interfering with the "keying in" process. "It probably would be helpful to get the four I can't get by cutting off the four I know already. There is no sense showing them any more."

4. Fourth Experiment (Repetition of Experiment 4)

The main interest of this experiment is that it is a repetition of an experiment the subject had previously done--with the intervening time filled by the third experiment and dinner, a total of perhaps four hours. Such a repetition can give us information about what sort of memory is involved in the decision structures, and how the various parts of the structure are retained.

The subject obviously recognized the experiment in the first few trials. Nevertheless, he did not immediately meet the terminating criterion, though he soon cleared up his difficulties and finished in 205 trials. He seemed quite at ease throughout the experiment, even though he had by this time completed over 4000 trials in one day.

a. Decision Structure

Our immediate question about the decision is whether or not it is exactly the same as it was at the end of the original running of experiment 4 (Figure 4-5). On the whole, the same overall structure seems to have been maintained, but there are some interesting differences in detail. First of all, the subject actually was now able to distinguish 101 from 111. He had definitely discarded his old rule entirely and was now recognizing 101 by a bright K or V in position 7. What he did retain was the position of this decision in the structure and the determination of 111 as what was left over after 101 was selected out. We have no evidence of positive identification of 111, though it might exist. Other differences--at least by the end of the experiment--were slight, or at least hard to identify with the small amount of data. If anything, there seems to be some consolidation of the critical positions away from the right side (5 and 8) and to the left and center (1, 4, 6, and 7). This trend might have ultimately led to a simpler decision structure, but with our terminating criterion, we could not go on long enough to see it.

It is particularly interesting to notice what the subject had forgotten of the previous decision structure at the beginning of the experiment. First of all, there is no trace of his old method of "discriminating" 101 from 111. Secondly, he forgot that 001 and 011 were identified by dim letters in position 8 as well as bright ones; and thirdly, he forgot that 110 was distinguished by a D as well as an R in position 4. We can be quite sure of these facts for they are verified by the actual errors made as well as by the response time behavior after the structure is recovered. For instance, response 011 is an average of 1.15 seconds slower for dim K's than for bright ones; 001 is .65 seconds slower on the dim V's; and 110 is .45 seconds faster on R than on D. Another less certain fact is that he was confused about 100 for a while, but here there was no splitting of cases so we have no response time evidence other than a general slowness.

The important thing to notice about these decisions is that they are the ones on the tail ends of the branches on the decision structure. They are also the last ones learned and thus the ones that have had the least reinforcement--or confirmatory evidence. This method of forgetting, of course, can be quite useful--as we see in the case of 101-111--but it also may make it difficult to discard erroneous higher level decision structures.

b. Strategy

The subject starts out using 100 as the "don't know" response; but starts to deviate from them as soon as the third trial, when he sees 101 for the second time. After he has seen a few of the better learned patterns--such as 000 and 010--the "don't know" strategy quickly dissolves and he proceeds to clear up individual cases as if he were near the end of a normal experiment. 000, it is interesting to note, was never missed in any way throughout the entire experiment, whereas other responses met with varying degrees of difficulty.

As he redevelops the lost rules, we again see the characteristic pattern of isolating a critical feature, overapplying it, and then discriminating cases--a good sample of which is the confusion with unknown cases showing first bright, then dim, V's as the V criteria for identifying 001 were relearned. We might summarize the entire strategy for this experiment by saying that it represents a recompletion of another experiment. From the subject's point of view, there seems to have been no difficulty picking up--not where he left off, but a little before he left off.

c. Verbal Report

The verbal report essentially confirms our analysis. The subject says that "single letters were a give-away because I remembered them," and that "as soon as I got a couple of things cleared up it was straightforward." In fact, he had very little to say about the experiment, and expressed no frustration or other emotion about it.

In describing one of his recovered rules, he says, "with a small k in the middle left and a large K on the bottom middle..." when he obviously must mean "large K on bottom right." There are three things of interest in this simple statement:

1. It is the first time he has referred to capitalization.
2. It may show a movement of attention to the left side.
3. It may show one way he is overcoming his lack of a clear spatial picture--by relying on redundant cues, as are given by the cases of the letters.

5. Fifth Experiment (Experiment 6)

This was the last experiment the subject performed; and although he was visibly fatigued by the long work on the previous experiments, this was by far his easiest and best performance. He not only finished much faster--365 trials--but for the first time he seems to have finished by mastering the complete rule. In fact, the last 100 trials or so seem to be merely an attempt to bring actual performance up to the theoretical level possible with his rule, for his few scattered errors show no discernible pattern.

a. Decision Structure

Although the experiment was over too fast to allow a great amount of data to be obtained through test trials, it is still possible to reconstruct most of the decision structure with confidence, as shown in Figure 4-7. We have enough evidence to place the final decisions for 001, 100, and 101 in their proper places on the structure, but we cannot tell what were the keys to these decisions. We do have information, but it only tells us a number of things which are not the keys.

This decision structure is quite similar to that for experiment 4, but the evidence here indicates a much stronger connection between the parts of the structure. In other words, the subject never seems to be confused about which branch he is working on. When we examine the structure of the two experiments, we can readily account for the resemblance and the differences, for the two have exactly the same distribution of value and noise bits in the intensity row. A appears three times; C appears twice; B and NB each appear once; and one noise bit appears. Thus, the total brightness of each value in the two experiments is the same: 000 and 010 have one or two bright letters; 001 and 011 have three or four; 100 and 110 have four or five; and 101 and 111 have six or seven. However, the spatial distribution of the brightness is different in the two experiments, so the problem of distinguishing the cases would not be exactly the same. Again, as in experiment 4, the subject could have completed the identification using only the brightness patterns, but he chooses to use letters as well.

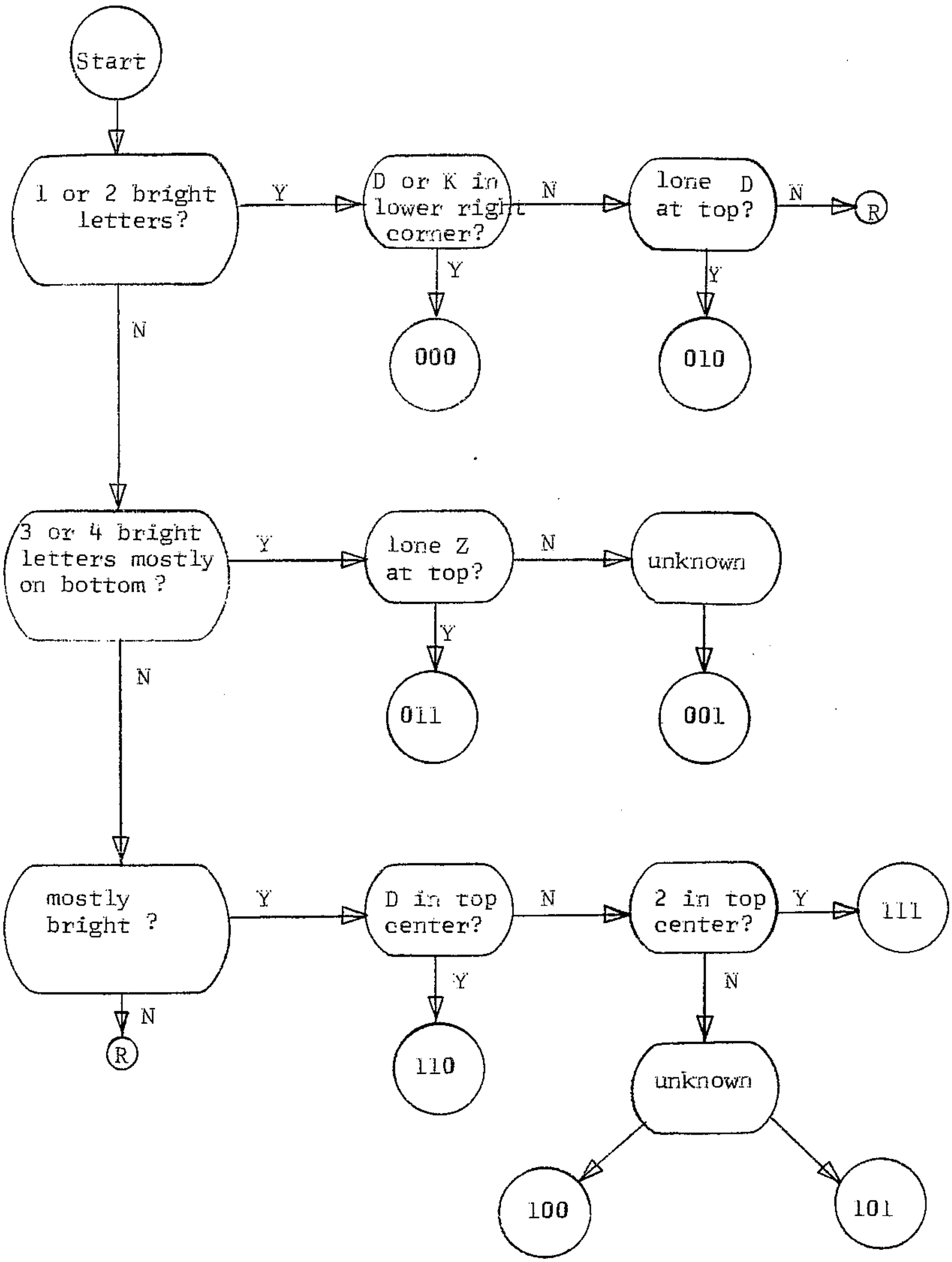

Start
1 or 2 bright letters?
Y
D or K in lower right corner?
N
lone D at top?
N
R
Y
000
Y
010
N
3 or 4 bright letters mostly on bottom ?
Y
lone Z at top?
N
unknown
Y
011
001
N
mostly bright ?
Y
D in top center?
N
2 in top center?
Y
111
N
R
Y
110
N
unknown
100
101

Figure 4-7

b. Strategy

The similarity of brightness patterns seemed to have made it quite easy for the subject to adapt his ideas from the previous experiment. Within the first 70 trials he had clearly and permanently distinguished the OXX cases from the 1XX cases, a fact which shows quite strikingly on the correlation plot. Even within that period we can detect the structure we would expect, namely, that 000 and 010 were distinguished first, and that 001 and 011 were never confused with 101 and 111. He seems to have made the critical distinction between the ambiguous cases on the basis of the location of most of the brightness. Once this was done--and was established much more firmly than it had been in experiment 4--he seems to have taken one response at a time and worked out the precise discriminating rule. His final difficulty seems to have been with the set (111, 100, and 101), but this cleared up quickly once he had separated 111 from the others.

Another part of the strategy which may have been carried over from previous experiments was the reliance on position 8 and possibly position 2. On the other hand, we cannot be sure that he is not using these positions for other reasons. Position 2 is especially interesting because it is the only case in all these experiments where only two different letters appear in one position. The subject would thus see these two letters much more frequently than other letters, so perhaps he chose them as "key" letters.

c. Verbal Report

The most striking thing about the verbal report in this experiment is the evidence of development of precise language for describing the experimental situation. The confusions about positions of characters in the stimulus have all disappeared. A complete system of identifying the response squares has been developed--"top," "bottom," "left," "right," "top left," "top right," "bottom left," and "bottom right." Sometimes the terms "top middle" and "bottom middle" are used as well, though strictly speaking, they are redundant. Even more interesting is the development of a second terminology for describing the response space, namely, the division into "right" and "left." In this categorization, "left" includes not only the squares designated as "left" in the other scheme, but also 100; whereas "right" includes 000. Since 000 and 100 are neither right nor left relative to one another, this choice of grouping seems to indicate that the subject has developed a name for the concept corresponding to bit A, which divides the response array in half along a tilted vertical line into "left" and "right." This concept has been used in this experiment and in experiment 4 to make the major division in the decision structure--which seems much more clearly made in this experiment. The other set of names seems directly related to the perceived structure in experiment 3, or even experiment 1.

One consequence of this improved ability to describe the response space may be the subject's increased ability to make specific tests of hypotheses about particular responses. He was quite aware that this ability had improved, and spoke of "sorting out the cases on the left." What he seems to have been doing is a serial response isolation on this left half, something he could not seem to accomplish previously.

C. Subject 3

1. First Experiment (Experiment 1)

Before the experiments even started, it was clear that this subject was going to cause trouble. He had been instructed to get a good night's rest, but had been up until 6 a.m. grading examinations. When we started the experiments at about 1 p.m., he had been awake about an hour and was not very alert. He was much slower in his performance than any other subject, and didn't seem too troubled about going over the time limit on a trial. Furthermore, he would switch the pen from one hand to the other every ten to thirty trials. Later, he explained that he was ambidextrous; but it was not likely that any consistent information would be obtained from the response time data. Then, he had trouble operating the light pen--forgetting to press the button, pointing it carelessly, or holding the button too long. In a way, it was fortunate to have such a subject, for he represents about the worst conditions under which an analysis might have to be made.

He worked on the first experiment for about three hours, with two pauses. At the end of 1653 trials, he expressed a wish to try something else. He was not discouraged, but he was losing interest, so we switched to the next experiment. At the termination he was performing at about 70 or 75 percent correct responses, the result of an almost linear climb in performance since the beginning. A look at the tabulations reveals that there was not a single response that he could perform perfectly, though some were markedly better than others. His response times to most stimuli were extremely erratic, and there were only 10 out of 64 possible errors that he had not made at some time in the experiment.

a. Decision Structure

It would have been easy to dismiss this entire experiment by saying that there was no decision structure, but this would leave the high performance level unexplained. As we delve into the test trials and data listings, however, a pattern gradually begins to emerge. We first recognize that, for the most part, the subject is using different positions for each response--which is reminiscent of the brightness strategy of subject 2. This time, however, the choice of positions does not seem to correlate with brightness. We find that there are at least five responses which can be definitely associated with stimulus positions: 000 with 2, 001 with 3, 010 with 5, 110 with 4, and 111 with 1. In addition, 100 seems associated with 4 and 5 (in some cases) and 101 with 4 and 7 (in some cases). If we arrange this information in a table, we get:

	1	2	3	4	5	6	7	8
000		X						
001			X					
010					X			
011								
100				X	X			
101				X			X	
110				X				
111	X							

Very little sense can be made of this table, until the inspiration comes to rearrange the columns so that they read in clockwise order around the stimulus, the way the responses are numbered. We then get the following table:

	2	3	5	8	7	6	4	1
000	X							
001		X						
010			X					
011								
100			X		M		X	
101					X	M	X	
110							X	
111								X

with the M's marking the centers of symmetry of the double entries. The diagonal nature of this matrix is clearly not due to chance--it must be a manifestation of a very strong bias toward compatability between stimulus position and response position. In other words, something in the stimulus must "point" to the correct response.

Once we know that the concept of pointing is being used, our analysis gains momentum. To find out what other concepts are being favored, we make a table of the responses, their "corresponding" positions, and what appears in those positions when the responses are to be made:

R	S	INTENS	CASE	LETTERS	RIGHT	PCT RIGHT	BRT-PT	SML-PT
000	2	br	sm	z and t	110	.64	.52	.84
001	3	dim	noise	h	59	.32	.35	.96
010	5	dim	cap	z and v	12	.20	.43	.94
011	8	br	sm	k and v	91	.51	.96	.98
100	7	noise	sm	r	90	.73	.70	.92
101	6	dim	cap	t and h	54	.44	.89	.89
110	4	dim	sm	z and m	71	.54	.98	.94
111	1	br	sm	k and t	137	.72	.46	.85

This table indicates a bias toward rules using small letters to "point." The last column in the table makes the strength of this bias apparent. Each stimulus value "points" with its small letters to half of the other responses, on the average. The last column, SML-PT, gives the percentage of erroneous recognitions in which one of those pointed responses was chosen. The next to the last column, BRT-PT, shows the same calculation for the responses "pointed" by the bright letters. The high figures in this column occur only in places where bright and small point essentially to the same things--they might indicate a slight bias toward brightness as well as small letters, but the evidence is not very substantial.

The two responses, 100 and 101, which are at least partly determined by two positions, give us an additional clue to the subject's behavior. When we examine the cases of 100 which are learned, we see that they are the ones with noise bit 4 = 0 and noise bit 2 = 0. In these cases, positions 4 and 5 both contain a small bright m's, which symmetrically bracket the small r in 7. 101 was first and best learned for the cases when noise bit 4 = 0, and in these cases, positions 4 and 7 each contain a small v. These two are symmetrically placed around the small h in 6, which "points" to 101.

Working with such clues, we can reconstruct not so much a decision structure as a set of decision rules which are applied in no particular order. In addition to the rules already seen for 100 and 101, the subject seems to be using rules such as the following:

1. Two small r's in 5 and 7; pick the lower one (100).
2. Small t in 2, pick it (000) unless (and this is a later modification) there is a small h in 3, in which case, pick it (001).
3. Small t in 1, pick it (111) and (later) small k in 1, pick 111.
4. Small m in 4, pick it (110), but this conflicts with the previous rule for 100 with the two m's, and the data show the resulting confusion.
5. Small v or k in 8, pick it (011), but this causes 010 to be consistently identified as 011 also.

b. Strategy

As we unraveled the decision rules, an important part of the subject's strategy was revealed, too. He seems to have started with his idea about small letters "pointing" to correct responses, and after finding quick partial success with that idea (for responses 000 and 111), he seems to have been unable to give it up. As a result, he accumulated a large number of rules and exceptions which not only taxed his memory but which did not include some cases and were contradictory on others. After a while, he seems to have added the idea of double letters being significant; but unless they were in some clear pointing relation to the associated response, he could not seem to remember which response to make.

The evidence from the correlation plots indicates that he was able to get rather firm hypotheses and discard them within a reasonable time if they did not work out. Thus, he made steady but labored progress throughout the entire experiment, until he finally got tired of trying to remember everything.

c. Verbal Report

The verbal report has many interesting features when compared with our reconstruction of the experiment. First of all, the subject says, "I tried matching mostly small letters first." But "that didn't seem to work," so he "tried something else." It is apparent to us that he retained that idea throughout the experiment, whatever he thought he was doing. Additional evidence of how this bias affected his perception is given in his discussion of the role of the letter t in the experiment. "The small lower case (notice the redundancy) bright t is always golden," he says, by which he means that it always indicates the right response when it appears, with certain exceptions which he mentions as "context effects." Later, he says, "I think the t only occurred near the top of the thing...like in the upper row in the middle or left hand corner." Now, a T also occurred in 6 and was also "golden"; but it was not a small t, and he never noticed it.

He did, apparently, attempt to use capital letters when he was forced into it by his "pointing" strategy. He said, "Let's see, one other capital letter that seemed to work...there was the combination of k's that had...one k in (2) and another in (8), which usually required that the k in the lower right hand corner be chosen...except if there was a capital V in the middle of the right hand row, then it appeared that you wanted to capital V, but I'm not sure I had that one to the point where it was thoroughly checked out..." Seemingly, he did not; for though this explicit and correct hypothesis was given at one of the intermediate pauses, he was not able to follow the rule implied by that hypothesis when he resumed the experiment.

It is also interesting to notice the language he uses in describing the choice of responses. He consistently uses such phrases as, "you wanted the capital V," "pick the letter down in that corner," "pick the k," and "pick the h." As long as he is unable to distinguish between the stimulus and response arrays, it is unlikely that he would ever be able to make responses not physically related to some feature of the stimulus.

His report about specific letters used is often vague or wrong. He says, "...I forget what the two letters were, a couple of r's perhaps in the middle of the left hand row and the middle of the bottom row indicated that you were supposed to pick the letter down in that corner." This statement apparently represents a confusion with rule a. The letter in the corner, he does not even have an idea about, while the pair of letters are v's, not r's; but he is able to pick this response correctly every time. He can do this because there are no other cases where such a formation of identical small letters appear in this experiment. Where there was some sort of conflict, he was forced to notice specific letters and had a great deal of difficulty. For instance, he said, "I think the m in the middle of the left hand row was something that I got nailed on about three or four times, and then finally I decided I was going to watch for that in particular, because it appeared--it occurred--along with something else.... that was normally...normally suggested some other alternative." Presumably he is here trying to speak about his two conflicting rules involving m's, which are in constant conflict for his attention. He describes the situation by saying, "...you make a mistake on them about twice and then you make a mental note that it's time to look out for this particular situation." Indeed, we find exactly the kind of alternation we would expect from using this strategy with an inconsistent rule.

2. Second Experiment (Experiment 4)

Although this experiment shows about twice as rapid a rise as the first experiment to the 70 or 75 percent level, the subject expressed the same kind of boredom with the process at that point. When somebody came in to use the computer, he expressed no particular interest in continuing, so we terminated. Evidently, he felt that he had worked out all the cases, and that any errors he was now making were ascribable to the haste with which he had to make his decisions. Under the circumstances, he felt that the remaining task would only involve learning to be faster and more accurate in using the rules he already knew.

a. Decision Structure

By making a tabulation of errors, we discover that the subject still has a strong "pointing" bias. We also find that the bias toward small letters is retained, but that in some cases, it is overridden by a definite brightness bias. When we examine these exceptional cases, we discover that they are just those two cases (000 and 010) in which single bright letters occur and which were so easy for subject 2. In the case of 000, it happens that the bright letter "points" to the correct response--so that the pointing does not result in errors. We find that the subject performed quite well on 000, but had trouble getting started on 010, for he was always pointing to the two responses indicated by its one or two bright letters (110 and 011). When bit 2 = 0 (a single letter was lit), 110 was invariably chosen; but when bit 2 = 1 (letters in 4 and 8 lit), the subject alternated between choosing 110 and 011.

When we investigate the other bias we found in the first experiment--towards pairs of identical letters "pointing" to responses, we find an even more striking correspondence. In this experiment, such pairs which point correctly are not as frequent as they are in experiment 1; in fact, there are exactly four cases--all of which are mastered in the first 50 trials and never forgotten. In addition, for the first 300 trials, the cases in which a pair of identical letters "points" incorrectly are inevitably chosen erroneously as the "pointed" response.

We can even tell that the subject looks for these pairs before he looks at other features--the brightness pattern, for example--because in those cases where an erroneous pair exists, its indicated response is chosen over the one indicated by the other method. This applies even in the case of 000, in which the brightness pattern points correctly. For instance, when a pair of small d's appears in approximately the locations they appear for 100-XX11, 100 is almost always chosen instead of 000.

Eventually, he begins to add other items to his decision structure, which, as closely as we can determine from the data, was as shown in Figure 4-8 at the time we terminated the experiment. If we trace the effect of each of these decisions made in the environment of experiment 4, we get the following results:

1. Gets 1/2 of 101's, but makes 1/2 of 111's wrong.
2. Gets 1/4 of 110's.
3. Gets 1/2 of 101's, but sometimes gets 1/4 of 011's and 1/2 of 100's which have very similar structures.

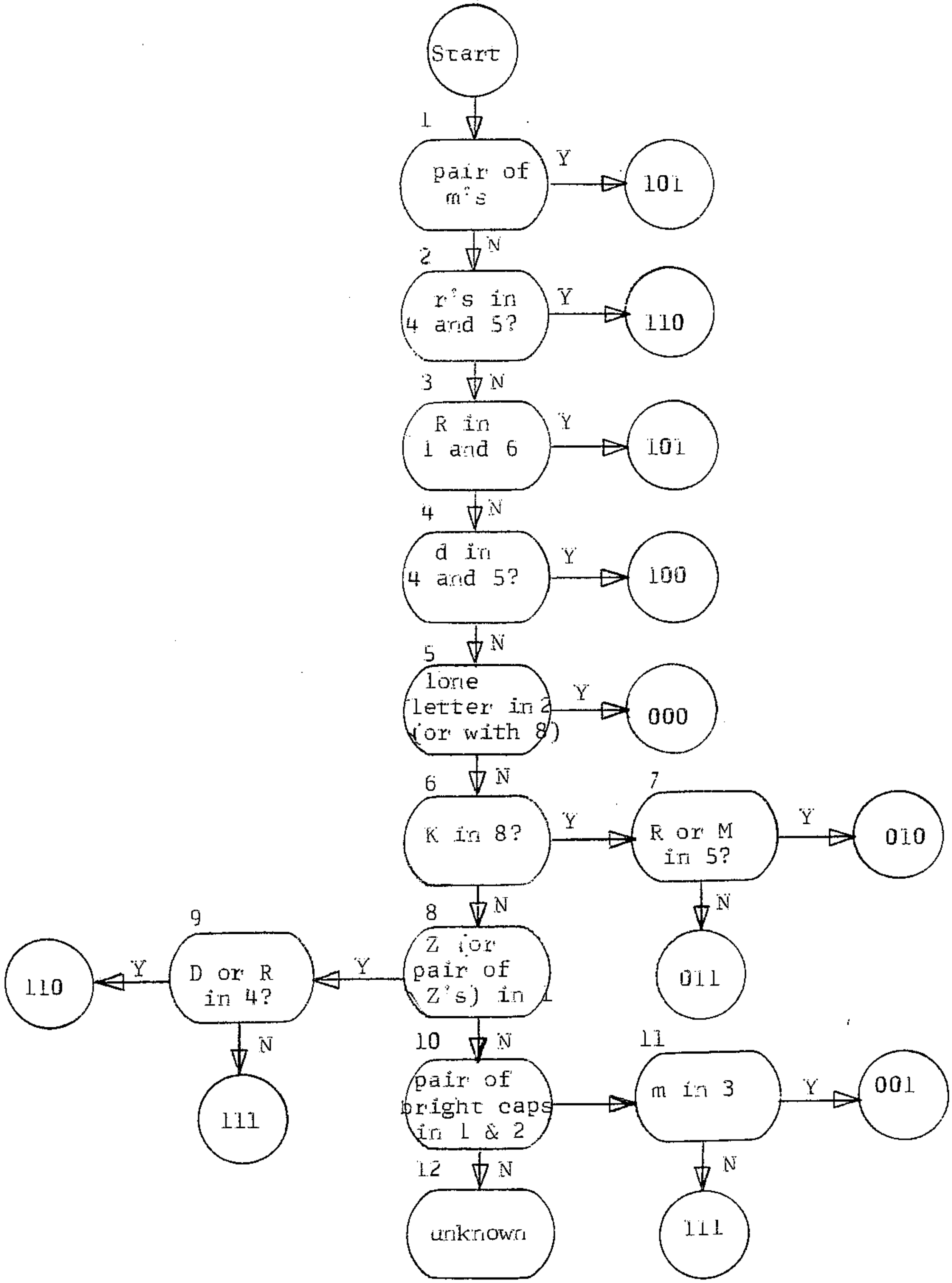

Start
1
pair of m's
Y
101
N
2
r's in 4 and 5?
Y
110
N
3
R in 1 and 6
Y
101
N
4
d in 4 and 5?
Y
100
N
5
lone letter in 2 (or with 8)
Y
000
N
6
K in 8?
Y
7
R or M in 5?
Y
010
N
011
N
8
Z (or pair of Z's) in 1
Y
9
D or R in 4?
Y
110
N
111
N
10
pair of bright caps in 1 & 2
11
m in 3
Y
001
N
111
N
12
unknown

Figure 4-8

4. Gets 1/4 of 100's but tends to miss 1/4 of 000's.
5. Gets remainder of 000's, but also seems to get some of the 111's which are the complementary pattern, dim on bright.

6 and 7. Gets all of remaining 010 and 011, except that the subject seems able to keep only one of the exception letters in operation at a time.

8 and 9. Gets remainder of 111's and 110's, but also seems to alternate in using the exception letters.

10 and 11. Gets 1/4 of the 001's, but ensures that 1/4 of them will be wrong.

Altogether, these rules result in about five out of eight responses being correct, two out of eight being incorrect, and fail to account for the remainder, which seem to be right about 1/2 of the time. It is interesting to notice how step 2, though redundant with the addition of step 9, remains in the structure. The only reason we can find out that the two coexist is that the subject is not yet at the point where he can consistently perform step 9, so we can see that although he misses the other cases during certain periods, he never misses the case covered by step 2. If the subject's response time behavior were more consistent, we could make such inferences from that, but, probably because of his switching hands, we cannot seem to draw a consistent response time picture.

b. Strategy

The subject apparently started out on this experiment with the same set of basic concepts he used in the previous one. When the concept of "pointing pairs," after some early success, did not get him too far, he seems to have given up the "pair" part of the concept, but not the "pointing" part. He then apparently was able to draw upon a concept of "pointing brightness," at least to get 000 and to make consistent mistakes in 010. It may very well be that step 6 in the decision structure was originally added as an exception to a more general version of step 5, such as, "single bright letter (or with bright 8), point to it." Then, because the letter in 8--when it, too, was bright and the subject had to decide between two pointers--was always a K, he might have finally been able to modify the structure so that the presence of the K was a major decision criterion. The sequence of data definitely supports this picture.

This subject shows quite a definite ability to vary his hypothesis systematically--though he may only be able to hold one at a time, as evidenced by his ability to apply only one-half of the rule at a time in cases such as step 7 or step 9. He also displays a certain amount of ability to allow performance to decrease for a short time, following which it climbs to a new level. This is the sort of performance we would expect to see when a partially successful hypothesis--such as, steps 10 and 11--was dropped.

An interesting confirmation of the "pointing" strategy being used is found in a note in the experimental log. While watching the subject perform, the experimenter noticed: "There is one pattern where a single bright D appears in position 2, which "corresponds" (points) to 000. On many occasions, the subject first pointed the pen at D itself, before correcting himself by pointing to the response square."

c. Verbal Report

The subject reported that he did, indeed, start by looking for "pairs of letters in places where they'd indicate something." However, he soon found that he was "trying to look for pairs...to the extent that I'd miss single letters."

At this point, the subject (who is a computer programmer) "came up with a sequential technique of about 6 or 8 steps which I think will separate these things into as many categories as are necessary. There are about two or three that don't fit the scheme, and I can recognize them by eye." In view of this subject's familiarity with exact specification of logical procedures, it is especially interesting to relate his description of the "algorithm," as he calls it:

> "The first thing I check for would be a k in the lower right hand corner. If there's a k in the lower right hand corner, you then check to see if the middle of the right hand side is either an r or an n (he means "m"). If it is, you pick the r or n. If not, you take the k. Okay, assume that

> the thing is not a k. Next thing you look at is the upper left hand corner, to see if that is a z. If that's a z, we check the middle of the left hand side to see if it's...I believe a d (or an r). Anyway, there's a key letter that needs to be picked over the z. Now if the key letter isn't there, then we pick the z. Now let's see, where do we go from there. The next thing I check for is sort of a...what I really looked for was the overall brightness of the pattern. If there was a letter...if the middle of the top...uh...there wasn't much else lit up...the lower right hand letter, maybe...then I picked that one. Near the top...that is, in the center of the top. Now if the one towards the left and the one in the center on top were both large letters, then I took the one on the left, provided that the one all the way on the right hand side of the top was not an n, in which case I choose that...."

Later, he describes some of the double letter rules, which he says he discovered originally. He is vague about the place of these rules in his algorithm; but when pressed he claims that he checks the k and z rules first--contrary to the evidence from his behavior.

He displays a remarkable verbal insight into the process he was using and the troubles with it:

"You seem to sit there staring at the thing and... uh...trying sort of random patterns for a while, and then everything starts to fall into place, except for maybe a third of the things that don't fit that pattern. (One trouble is that he constantly overestimates how well he is doing.) So after a while you revise the system that you're using. But it seems that you never revise the system so thoroughly that you throw away everything that you started with. You know. There's always some little remnant of what you originally decided to haphazardly try to use--even if it's no darn good."

D. <u>The Experimenter as Subject</u>

In order to provide a standard of performance against which the naive subjects' performance could be measured, the experimenter performed three experiments as a subject. In order to eliminate--insofar as possible--prior information about the experiments, these experiments were chosen at random from a number of experiments which had not been used frequently on other subjects (experiments 2, 5, and 7, in that order). The distribution of the value bits in 5 and 7 was extremely biased, which could have made them much more difficult than any experiments the subjects performed; but this was more than compensated for by the experimenter's knowledge of the common structure of all the experiments.

1. First Experiment (Experiment 2)

On the first trial, as I ran into the room after starting the program, I noticed the structure consisting of three M's in a column down the right side of the stimulus and immediately associated it with 010 which was lit next to it on the right side. The next trial showed a different structure, so I made a different response; but the third trial showed three M's again, so I responded 010, only to find that 000 was the reinforcement.

The only difference I had noticed between the two cases was that the M in position 3 had been small the first time and was now capital like the other two. I immediately reasoned that 5 and 8 could not be completely discriminating, since the same values had indicated two different responses. (I was wrong about this, because the brightness had changed--and 8 could have been used if I took brightness into account.) I thus decided to watch 3. On the next trial, capital M came up again in 3, so I chose 000 and was correct.

The fifth trial displayed a Z in 3, so I made a different response. I was particularly watching to see whether or not the M was a unique determinant of 000, for I knew that if I found a position containing such a determinant, I could construct a rule involving only that position. Thus, when the next trial showed a new letter, I made a new response and noted that it was neither 000 nor 001, which I had associated with M and Z respectively. The next trial was another Z, and I responded 001 and was right;

but the next trial was a small m and I responded 000 incorrectly because I was concentrating on letters after I confirmed the Z. I reminded myself of the capitalization, which I was able to reinforce on the next two trials. They were a Z and a z, and I got them both right, reasoning that the capitalization must choose the B bit, since it discriminated 000 and 010.

When a d showed up in 3 on the next trial, I knew it was not any 0XX or 10X response, so I chose 111. It was a 110, and so was the next trial, which I therefore got right. By this time, I was already convinced that 3 would be sufficient to determine the rule, and after a few errors because of faulty translation of bit B into response positions, I mastered the rule completely on trial 35. My conscious rule was the following:

a. If 3 is an m, choose 000.

b. If 3 is a z, choose 001.

c. If 3 is a d, choose 100.

d. If 3 is a k, choose 101.

e. After choosing, check the capitalization. If lower case, rotate 90 degrees clockwise before responding.

It was only in the physical translation of the 90 degree rotation that I had any difficulty after the fifteenth trial.

The test trials at the end clearly show that position 3 was the only position being used, for the only four errors were made the four times 3 was deleted. Further confirmation of the rule is obtained by comparing the average response times after mastery of the rule:

000	001	010	011	100	101	110	111
1.51	1.56	1.90	2.02	1.60	1.88	2.34	2.27

The systematic time differences are accounted for by the time required to move the hand around 90 degrees after the initial selection.

2. Second Experiment (Experiment 5)

Although I applied the same strategy to this experiment, finding the rule was not quite so simple. This time when a position indicated some regularity, I would work with it for a while until I found that it did not lead to the unambiguous kind of rule I was seeking. One after another, positions were discarded, either because they had the same character indicating different responses (implying insufficiency) or because they had more than one letter associated with a single response (implying the presence of a noise bit). I wanted to avoid insufficiency because I wanted a one character rule if one was available. I wanted to avoid noise bits because I wanted to avoid disjunctive rules (X or Y in position Z).

In the course of testing different positions (3, 1, 6, 7, 8, and 4, at least once each), I was often able to get partial rules, but I avoided the temptation to master part at a time and thus abandoned each one when it proved unsatisfactory. In the course of this effort, I began to notice that bits B and C were quite easy to find, so I suspected that bit A was not present in too many places. (Actually, it was only present in two places.) I decided to look for it in one of the binary variables--case or intensity--and soon noticed that position 2 was capital or small, depending on whether the proper value was "left" or "right." Again, about twenty trials were needed to commit the letter sequence to memory: m, h, v, k selected 000, 001, 010, and 011, in sequence; then I moved 180 degrees across the screen if the letter were capitalized. The entire experiment took 285 trials, about 80 of which came after position 2 was first noticed.

The test trials confirm the exclusive use of position 2; the average response times confirm the sequential aspects of the rule:

000	001	010	011	100	101	110	111
1.52	1.61	1.71	1.78	1.87	1.99	2.25	2.16

3. Third Experiment (Experiment 7)

I had decided to make one small modification to my strategy. Rather than depend upon conspicuous patterns to lead me to positions for testing, I would test the positions systematically. In this way, I would not have to tax my memory with a list of the positions already tried and discarded, yet I would not risk overlooking any position. I decided to start at the top right hand corner (position 3) and work my way from right to left across the rows. (Why I did not go in English reading order, I cannot say, except that I might have been influenced by previous successes on the top row and to the right.)

Position 3 was discarded in about ten or fifteen trials. I then proceeded to 2 and was greeted with immediate success. Again there was a brief period of committing the sequence to memory, followed by perfect performance to the finish, in 91 trials. This time the rule was as follows: on k, v, h, and m, choose 000, 001, 010, and 011, in order; if the letter were capitalized, I moved 180 degrees across the screen before responding. The test trials once again confirmed the use of position 2; but the response times do not show quite as unambiguous a picture:

000	001	010	011	100	101	110	111
1.63	1.50	1.67	1.62	2.15	1.75	1.77	1.78

The resemblance between this rule and the previous one is remarkable. I was conscious of the similarity in the sense of both requiring the memorization of a sequence of four letters for the first four responses and then using capitalization to get the four 180 degrees opposite. I was not, however, conscious of the fact that I was using the same position or the same set of four letters reversed.

CHAPTER V

GENERAL RESULTS

In this chapter we shall present those experimental findings which were common to all the subjects. Some of these results will be generalizations of material presented in Chapter IV, while others will be presented here for the first time.

Before examining the more specific results, we should note the significance of the subjects being able to learn at all. Nobody had ever performed quantifiable experiments of this complexity before, and more than one experienced psychologist expressed doubts that subjects could learn to perform these tasks. Indeed, there were individual experiments on which particular subjects gave up--which only serves to indicate that we are, indeed, operating in a problem region of interest.

A. Decision Structures

1. Representation of decision structures

One of the questions we set out to answer was whether or not human decision structures could be adequately represented as Boolean expressions or computer flow diagrams. The analysis of our experiments indicates that the answer to the question is "not completely." Let us consider the two representations in turn.

Boolean expressions, of course, are logically capable of representing the effective final result of any deterministic decision structure. As we have seen, final decision structures, at least, are as deterministic as any human performance can be. Nevertheless, Boolean expressions fail to represent the sequential nature of much of the decision structure; and thus they throw no light on either the evolution of the structures themselves or their relationship to response time data. For instance, when a subject makes a certain response when a letter is D _or_ K, it is logically irrelevant but psychologically important whether these conditions are tested in the order D then K, K then D, simultaneously, or on completely different branches of a more complex structure. In other words, each Boolean expression represents a large _class_ of decision structures, each of which chooses the same responses to the same stimuli. Therefore, the Boolean expression can tell us nothing about evolution or timing; and it also can tell us nothing about how the subject would respond to a stimulus taken from outside the set Y, such as we get when the test trials are run.

Computer flow diagrams are essentially contingent sequences of Boolean decisions. Consequently, they are somewhat superior to single Boolean expressions as psychological models of fixed decision structures. Several pieces of evidence from our experiments, however, indicate that they, too, have shortcomings. We have seen (Figures 4-3 and 4-5) that certain behaviors cannot be explained in terms of a fixed flow diagram, but might better be represented by a disjointed diagram whose parts are probablistically connected.

A second effect we have uncovered is the spreading of the criteria for decision--which, though perhaps unconscious--enables the subject to perform better on mutilated stimuli than single sequence decision structures would imply. In other words, much of our evidence indicates that the actual decision structure would be more accurately modeled as a number of decision structures or partial decision structures evolving in parallel--with one of those structures being in "consciousness" at any time.

In spite of these shortcomings, flow diagrams--or flow diagram-like constructions--are useful in talking about decision structures. In many cases, they seem to represent the subject's actual behavior quite accurately--particularly when the structure is not too large and the stimulus does not present any peculiar problems.

2. Choice of Elements

Although each subject chooses the stimulus elements used in the decision structure in his own way, there are a number of general features to this choosing process which all subjects seem to have in common. The general process can best be described as having two stages: elimination of _irrelevant_ elements, and selection of _conspicuous_ elements from the remainder. Of course, what is irrelevant and what is conspicuous is different for each subject; but for a given subject they represent a consistent combination of _a priori_ ideas and ideas carried over from earlier experiments. Irrelevant, then, means "excluded from consideration in advance"; but the basis for this exclusion may be quite different in different subjects. We have seen, in our small sample, one subject who started the experiments with the idea that capitalization was the only relevant factor and another who had the same idea about brightness. That one idea led to success and one to failure was only a complete coincidence in the structure of the experiment.

As the subject gains experience, he may change his mind about what is relevant. In our experiments, he broadens his ideas; but perhaps that is because everything _is_ relevant--or potentially relevant--in our set of experiments. Nevertheless, there is a remarkably strong tendency to preserve the original idea of relevancy as the cornerstone--even though it may lead to extremely complex, even unmanageable, decision structure. We see no evidence in any of our subjects that the _a priori_ idea of what is _most_ relevant ever changed, even over thousands of trials. Apparently, we would have to work much longer with a subject to affect this idea.

The concept of what is conspicuous seems to be made up of a number of interacting concepts, several of which are shared by all of the subjects. One such concept is that of the "key" or "conditioning" letter. The key letter alerts the subject when it appears and is often used as a reference point for determining where to look for other cues. What particular letter is chosen as key seems not to be predetermined, but may sometimes be carried from experiment to experiment. In most cases, however, the key seems to be a letter that appears with more than average frequency in the stimuli of the current experiment.

Another concept of conspicuousness shared by all subjects is that of a compatible spatial relation between the stimulus array and the response array. Though the reliance on this concept varies, each subject shows a definite and strong preconception that the stimulus should somehow "point" to the response. This effect may be augmented by the short reinforcement time and the method of reinforcement, both of which contribute to the value of looking at clues close to the response, so that no large eye movement is required.

Inasmuch as the decision structures are primarily sequential, the placement of the different responses in them will have important consequences for their ultimate simplicity. All the subjects seem to share certain preconceived ideas about the response space. In particular, the top center position is almost invariably the first one learned; positions 010 and 100 are almost always the "don't know" positions; and, in fact, the vertical and horizontal positions (XX0) are much favored over the "corner" positions (XX1). In the stimulus array, on the other hand, if there is any bias at all, it is in the "corner" positions. Biases in the choice of element position in the stimulus array are not critical, because of the redundancy; but any biases about the response array must be overcome in order to solve each experiment. Subjects get much better at communicating verbally about the response array than about the stimulus array.

3. Growth and Form

Certain forms of decision structure are definitely favored in these experiments. We can understand these forms in terms of the processes by which they are created, and the constraints imposed upon them in that process.

The basic structure which seems to be preferred by all subjects is a linear sequence of positive identifications of responses, which is perfectly exemplified by Figure 4-1a, except for the last decision, which should be followed by the decision "top cap?" which would identify 101. This structure is the natural result of a perfect serial response isolation procedure, and is built up one link at a time, always adding at the single open end. This form and building procedure gives a number of advantages to the subject as he tries to solve the experimental problem:

a. It is simple and systematic, and thus conserves memory.

b. It requires a maximum of eight remembered decisions.

c. It ensures that each decision already learned will be tested on each trial, thus preventing a wrong decision from lasting long.

d. Positive identification protects against misclassifying previously unseen stimuli and checks against changing rules.

e. It may permit all decisions to be based on the same items, thus providing a possible check for completeness.

f. The strict ordering may permit simpler decisions, because certain cases will have been positively eliminated by the time a particular decision is reached.

g. Once started correctly, it ensures that enough information is available for a complete solution.

There are, inevitably, disadvantages to this form as well:

a. It may take too long to make all the decisions for the last few responses in the chain.

b. Only one response can be worked on at a time, so the subject is at the mercy of the chance sequence of cases.

c. There can never be fewer than eight decisions in the chain.

d. Each decision must be different from the others, so there must be eight <u>different</u> decisions.

e. It must be started with one perfectly identified response.

As we have seen, a subject who knew the structure of the experiments could readily simplify this serial decision tree and thus overcome most of the disadvantages. Naive subjects, however, deviate from this structure at their peril. Figure 4-4 is a specimen of one of the pitfalls--if the first step does not uniquely identify a single response, the underlying decision criterion (overall brightness) may not be sufficient to make a complete solution. Another trap for the unwary is starting with a unique identification but with too complex an underlying decision. The individual decisions may take so long that there will be no time to respond to the ones on the tips. In general, then, we see that a firm root is needed to make this procedure work. If a subject is anxious to start quickly being partly right, he will probably be led into difficult variations.

Figure 4-5 is a typical illustration of what happens when such premature starting is done. Each branch of the original tree becomes the root for a new sequence, which may then grow according to the basic pattern. Splitting the problem in this manner could have several advantages:

a. The number of decisions actually made on any path is smaller than for the full string, thus response time requirements may be more easily met.

b. The same criteria may be used several times in different branches ("bright D or R in 4?").

c. It may lead to perception of the deeper structure of the environment.

d. There are several open ends so that more than one response can be worked on simultaneously.

On the other hand, the disadvantages are quite serious:

a. Many more individual decisions have to be remembered.

b. There is a possibility of forgetting which branch is being used, especially when the same decision appears more than once.

c. Even when one branch starts with a unique identification of a response, there is no assurance that the others will be able to evolve into a complete solution.

d. Most of the testing provided by the linear sequence may be lost, although positive identifications are still possible.

In our experiments, those in which the naive subjects could stay close to the serial decision structure were the most successful experiments. The major difficulty in the other cases seemed to be the inability to handle the extreme size of the structure required, for the difficulty seems to increase exponentially with the number of decision elements in the structure. Of course, if the subjects had been able to abandon an unwieldy structure and start over, they might have been more successful; but the only place they seemed able to change the decision structure was at the tips of the branches. Once a poorly chosen decision was buried more than one level in the branch, it seemed unassailable. For the serial structure--with its confirmatory properties--this tactic makes sense, but for these derived structures it may be fatal.

B. Strategies

1. The Stages of a Problem Solution

There seem to be three major and fundamental parts to the strategies of each of the subjects in these experiments. The first stage involves finding a distinguishable stimulus, that is, one which is rememberable as "the same" as previously seen stimuli--same within the narrowed field which the subject has brought to the experiment. The second stage is associating that stimulus (which is actually a stimulus class) with some single response. This is the stage when most of the "thinking" takes place. Finally, the subject must memorize the association, cementing it in place with all the other associations he has mastered.

The first stage seems to be the only one where any significant amount of parallel processing of data can occur. The subjects seem to be able to hold a number of different descriptions in mind at one time; but when they see one for the second or third time, they drop the others until stages two and three have been completed. Subject 3, because he looked only for cues which would identify the response to which they "pointed," never really had to go through a separate second stage. He paid for this simplification, however, by having more complex rules to remember and by being unable to remember those few associations he tried to make where the "pointing" was not direct. The memorization of an association became more difficult the more associations already memorized; thus, the difficulty of completing an experiment went up exponentially with the complexity of the derived rule. Subjects complained, in these cases, of being unable to remember the ones they knew before.

Our evidence for the existence of the first stage comes from the behavior of the correlation curves. The behavior of the peaks and valleys indicates a rapid acquisition, testing, and discarding of temporary hypotheses in this stage. Furthermore, if we compare the behavior in the experiments with different distributions of the value bits, we see the sort of picture we should expect. Suppose we say that the average length of stage one for one of the value bits is determined by choosing the point at which its proper correlation reaches the three sigma level and remains there for the remainder of the experiment. Using this criterion, we find that in the 17-9-6 experiments, the average numbers of trials in stage one are 262, 335, and 637, respectively. In only one experiment (subject 1, experiment 2) did the order of acquisition deviate from 17-9-6--and in that one, the bit with 9 representing dimensions had reached the three sigma level only 48 trials ahead of the 17 dimension bit. In the 10-11-10 experiments, the average numbers of trials were 221, 272, and 182 for the A, B, and C bits, much more balance than in the 17-9-6 experiments. Evidently some kind of search procedure is being carried out in this stage.

2. Tactics

Just because the subject had no direct influence on the sequence of stimuli presented to him did not mean that he was unable to use different tactics. One tactic was waiting for a specific stimulus to show up so that some hypothesis about it could be tested. The basic difficulty in using this tactic was the temptation to try to do something with those stimuli coming up while waiting--with the result that no progress was made at all. A more successful tactic was holding an association for just one or two trials, so it could be tested if a related case came up immediately after. This not only conserved on memory but on patience as well.

There were two ways, from the experimenter's point of view, that a subject could arrive at mastery of a single response--widening and narrowing. From the subject's point of view, however, there is probably only a single way--narrowing. When the subject has identified a certain stimulus (class), he starts responding with what he thinks is the correct response. If he is never wrong, he commits the association to memory and moves on to the next. If he has done this prematurely, because he happened to see a succession (usually two or three was sufficient) of favorable cases, he does not discard the association already learned. Instead, he continues making the same response to that stimulus, but begins looking for some variation within or close to his field of observation which correlates with the times he is wrong. If his field of observation is too narrow, he may not be successful in finding such a variation; but only rarely, if ever, do subjects drop learned associations at such a time. Instead, they continue to make that response to all the stimuli in the class and move on to another three stage cycle.

If a variation is found, it is tacked onto the existing association as an exception: "Choose the k unless there is a capital V above it"; "Choose 1 unless there is a Z on the right." Now the rule with its exception may be subjected to the same process. If it still leads to some wrong responses, an additional exception may be tacked on. Sometimes the next appendage turns out to include the previous one, but they are both retained, as in, "Choose 1 unless there is a pair of Z's or if there is a Z on the right." Sometimes the rule is not completely right even with its appendage, but no further variation can be found. In such cases, the process is stopped, and the rule is retained as is. If the exceptions are narrow enough, the subject may be performing at 75 or 85 percent on that response; in which case he is likely to begin ignoring exceptions and may never perfect his rule.

If this is the basic process the subject uses, how does <u>widening</u> ever come about? In the first place, most widening seems to be merely a result of arriving at two or more part rules at different times and by different paths. Subjects do not seem too interested in information about what the <u>right</u> response is when they are wrong. Either it confirms their hypothesis or refutes it. Only when they are in the first stage do subjects extract information on what the specific response was supposed to be. Consequently, if a subject picks up a part hypothesis--such as (in Experiment 1), "capital R in position 1 implies 000"--he is likely to pick up the other part as a completely independent hypothesis, such as, "capital M in position 7." Even when the second hypothesis includes the first, the first will be retained, apparently because it is further back in the decision tree. Thus, widening is not generally something the subject is conscious of; though if the two hypotheses are similar enough and close enough together in time of acquisition, they may be consolidated as a single rule, such as, "capital R or capital M in position 1." In these cases, the two are acquired so close together that it does not have the appearance of widening; and, in fact, the two hypotheses may never have had separate existences in the subject's mind.

A second way in which widening may come about is as a by-product of the narrowing of another hypothesis. Sometimes--particularly when the two confused responses differ only in bit C and thus are adjacent--the subject captures the second response in terms of an exception to the first. When this happens, each successive restriction added to the first response appears as a widening of the second. In these cases, the separate parts seem to consolidate. It is as if the rule is built up as follows:

a. If A, then X.

b. If A, then X; unless B, then Y.

c. If A, then X; unless B, then Y; or unless C, then Y.

d. If A, then X; unless B or C, then Y.

In contrast to this reduction, subjects seem unable to make the reduction from:

"If A, then X; unless B and C, then Y; or unless B alone, then Y."

to the form:

"If A, then X; unless B, then Y."

but only to the form:

"If A, then X; unless B and C or B alone, then Y."

This last reduction, of course, is the same as the one from c to d.

In our reconstructions of experiments, we can often identify when the subject is in which stage for which response. When we compare these stages with the annotated data listings, we find that the P type events are often found in stage one correlated with V's. This seems reasonable, since seeing the same stimulus twice in a row would seem to be the easiest way to remember it. On the other hand, the N events seem to convey little information until stage two is reached. When response Y is in stage two, events of the form:

X Y N
Y

are found to be highly correlated with the beginning of a new narrowing step.

3. The Role of Initial Assumptions

Although the subjects were very similar in their strategies and tactics, their results were quite strikingly different. The major reason for differences seems to lie in the higher level concepts adopted without question and held throughout the experiments. These assumptions were the following:

Subject 1:

a. The same stimulus positions would hold the rules for all responses.

b. Three positions were about enough to work with.

c. The three positions should be in a column or row.

d. Capitalization of letters held the primary clues.

Subject 2:

a. The bright letters contained all the needed information.

Subject 3:

a. Each response was determined by the stimulus element adjacent to it.

b. Small letters held the primary clues.

c. Pairs of letters, particularly in some symmetrical arrangement, were also significant.

In addition to these assumptions, all three subjects shared the following two assumptions, though in varying degrees of strength:

a. There was some sort of compatability between the stimulus and response.

b. There was some "key" or "critical" letter for each experiment.

Of course, even deeper than these lie certain assumptions which are not even stated, such as, the assumption that there <u>are</u> rules, that they are relatively stable over time, and that the necessary information for the rules lies somewhere within the display on the screen.

Each of these assumptions was questioned verbally at least once, but subject's often expressed that the problem would be "too difficult" if their basic assumptions did not hold. Only the most specific of the assumptions, however, were actually varied at all in the experiments--assumptions about capitalization, for example. Of course, assumptions more specific than that were also made--such as, "a k in 3 determines response 100"--but these seemed not to cross the boundaries between different experiments. In a way, then, the subjects shared another assumption: that the logical relations among the characteristics of the stimuli were the variable items from experiment to experiment, so that hypotheses about them were to be of lowest rank among the entire set of assumptions.

For example, if we observe that sometimes k in 3 implies 100 and sometimes it implies 110, we can take this as evidence to contradict the hypothesis about k determining 100, or to support a hypothesis about the rules being changed periodically. What seems to happen is that subjects will not change assumptions which have given them a certain amount of perceived success in making positive identifications. Thus, if a lower level assumption seems to work, all of the higher level assumptions supporting it are validated. When subject 3 finds four or five symmetrical double small letter patterns which point to the correct responses, he is reinforcing his assumptions about pointing, about small letters, about double letters, and about symmetry. When he finds four or five other double letter patterns that do not work, he only regards these as accidents--the positive cases have already shown the worth of the higher level assumptions.

In a way, this attitude towards higher level assumptions implies another assumption of a very high level--namely, that there is only one "real" rule. Because of this assumption, anything which gives performance above a chance level cannot be due to chance and must be part of the whole rule. Consequently, once something works, it must be retained--and with it, all its basic assumptions.

The experimenter's behavior is in marked contrast to that of the subjects. A look at some of his assumptions tells us why:

a. The relationship between each stimulus class and each response is fixed throughout each experiment.

b. There are many ways to characterize this relationship.

c. There is a fixed relationship between the position in the stimulus and the information it carries.

d. There is no fixed relationship between letter values, capitalization, or brightness and the response set.

e. There is noise in the stimulus.

f. There is rarely more than one noise bit per position and per characteristic.

g. There is probably a single position which will be adequate to determine all responses.

h. The response space has a fixed and known structure.

In contrast to the subject's assumptions, these assumptions are all correct--and furthermore, they are known to be correct and not subject to verification. Of course, they still do not uniquely determine the experimenter's behavior on any given experiment. There are other assumptions which are more like personal preferences, or acknowledgments of personal limitations.

In theory, the experimenter, knowing all the correct assumptions about the experiments, should be able to solve an experiment in as few as three trials. To see how this might be done, consider that the three stimuli seen are the following (with underline meaning bright):

Trial 1: r $\underline{v}$ $\underline{m}$
T + Z = 010
$\underline{V}$ H $\underline{k}$

Trial 2: $\underline{h}$ R $\underline{V}$
m + $\underline{M}$ = 110
D $\underline{d}$ $\underline{Z}$

Trial 3: z $\underline{v}$ $\underline{H}$
K + T = 011
$\underline{r}$ $\underline{T}$ $\underline{v}$

Given the knowledge of assumptions, using trials 1 and 2, we can deduce the following facts about capitalization rules:

a. 4 is either NA or noise.

b. 3 is either A or noise.

c. 1 is either NB or C or noise.

From trials 1 and 3 we can deduce that:

d. 3 is either C or noise.

e. 6 is either NC or noise.

f. 1 is either A or NB or noise

From (b) and (d), we can deduce:

g. 3 is noise.

and since we assume there is no more than one noise bit per characteristic, from (a), (c), (e) and (f), we can deduce:

h. 1 is NB.

i. 4 is NA.

j. 6 is NC.

This, of course, is exactly the rule that subject 1 <u>induced</u> in more than 500 trials. Other equivalent rules could be deduced from the same three trials.

The best actual solutions--given the full set of assumptions--took about ten times the theoretical minimum. The difference can be attributed to three factors:

a. The inability to carry out the requisite number of simultaneous deductions in the required time.

b. Inability to remember whole stimuli.

c. Inability to remember derived rules perfectly without some practice.

The second of these limitations forces the use of a strategy which works serially on selected stimulus parts. The first requires that not more than a few responses be worked on at once; and the third stretches out the mastery time for many trials.

Each of these limitations, naturally, also handicaps the naive subjects; but they have the additional burden of unproved assumptions. Subjects 1 and 2 have major assumptions which deal with limitation (b); but subject 1 comes much closer to the simplest correct assumption and thus has a generally easier time. Subject 3 has no assumption which reduces the number of stimulus locations scanned; perhaps this accounts for his average response times being a full second slower than the other two subjects'. Indeed, his "pointing" assumption requires that he look at every one of the eight positions. Eventually, he tries to overcome this difficulty by examining the locations in a systematic order and does succeed in improving his performance somewhat.

Limitation (a) forces each of the subjects into a primarily serial strategy. Limitation (c) causes particular trouble because of the contradictions between these derived rules and the compatibility assumption. Subject 2, after several experiments, had induced certain properties of the response space, and this helped him shorten the memorization time. Subject 3 had no trouble at all remembering rules in which the pointing was direct; but could never adequately remember those in which the pointing was ambiguous or indirect.

In summary, then, we can indeed account for most of the subjects' deviation from optimal human performance (as approximated by the experimenter's performance) merely by showing how their assumptions deviated from the narrowest correct assumptions. We might say that they had a more difficult time because they were working on a larger class of problems. In a way, the assumptions the subject makes about the difficulty of the problem are of the nature of a self-fulfilling prophesy: the more hypotheses his assumptions allow, the more hypotheses he must eliminate; thus, the harder the problem.

C. The Relationship of Gestalt Measures to Performance

If our model of the subjects' performance is correct, there are certain predictions which we can make about the performance and g-factor curves. First, assume that the subject is using a pure serial response isolation strategy. In this case, both his performance curve and his g-curve should be flat until he captures the first response, then increase monotonically until the problem is solved. The shape of the monotone increase will depend on how the difficulty of isolating each new response changes with the number of responses already isolated. If there is no effect, the rise will be linear; but if the difficulty increases, the second derivative of the rise will be negative, leading to a convex upward curve.

When the subject deviates from the serial response isolation strategy, the g-curve should be sensitive to the change, though the performance curve might not change at all. Suppose, for instance, that the subject first gradually learned bit A, then bit B, and finally, bit C. As the performance on bit A rose from .50 to 1.00, the overall performance would rise from .125 to .25, but g would remain zero. As bit B was mastered, the performance would rise from .25 to .50; then from .50 to 1.00 as bit C performance increased. If the bits were of equal difficulty, the performance curve would be concave upward; but if the difficulty increased with the amount learned, the rise in performance might appear linear. In any case, increases in performance without corresponding increases in g indicate a deviation from the serial response isolation strategy.

We can establish a rule of thumb which will tell us early in an experiment whether or not a subject has adopted a serial response isolation strategy. We observe that in such a strategy, a, b, and c should each be 1/2. In that case,

$$PABC = 7/8\ g + 1/8$$

so that if (PABC - 7/8 g) is much greater than 1/8, there is a definite deviation. Now, if we classify all twelve of our experiments as "easy" or "difficult," based on the length, presence or absence of a solution, and verbal report (all of which give the same result), we find that we can use this rule of thumb to make the same division--but before the experiment is over.

Those experiments for which PABC reaches .5 before g reaches .4 are just those experiments which are classified as "difficult"; while for the "easy" ones, g reaches .4 at or before the point that PABC = .5. Since this level is reached long before the end of the experiment, the rule becomes a predictor of impending difficulty. Before PABC reaches the .5 level, any corresponding rule of thumb is not quite as powerful. If PABC reaches .25 before g reaches, say, .15, we can predict that the experiment will be difficult; but if the condition is not fulfilled, we cannot say with certainty that the experiment will be easy. Nevertheless, we have a measure which promises to be useful in future experiments where the computer's behavior would be changed in the course of the experiment in order to help the subject overcome the difficulty he does not yet know he will experience.

As the performance level gets higher, the subjects seem less capable of changing their basic hypotheses--the "roots" of their decision tree. By the time the .5 level is reached--which represents, in a pure serial strategy, about three to four responses isolated--there seems to be no turning back. Therefore, this level provides a good prediction of future trouble, for, as we have seen, deviation from the serial strategy produces splitting of the decision structure, and splitting of the decision structure produces increasing difficulties.

For those experiments in which serial strategies were used, the linear shape of the PABC curves indicates that the difficulty does not increase as a result of the increasing size of the decision structure. Figure 5-1 is the performance curve for the most serial of our experiments, and its linearity is quite remarkable. The more the strategy leads to a larger decision structure, on the other hand, the more the curve bends downward at the end. Yet in spite of the linearity of the performance curves, the f and g-curves for the "easy" cases consistently show a marked dip about halfway through the experiment--at about .55 or .60 performance level. Figure 3-14 shows these curves for the experiment whose performance curve is Figure 5-1. The dip at 252 is less marked in the g-curve than it is in the other "easy" experiments, but is quite easy to see in the f-curve. In fact, these dips are so characteristic of the "easy" experiments that three different people, when asked to divide the f and g graphs into two groups according to their "similarity," independently isolated the "easy" ones from the others.

The repeated experiment (experiment 4 for subject 2) was excluded from these classifications. When one of the classifiers saw it, she spontaneously remarked, "Why, that looks like the last half of one of those other ("easy") curves." Indeed, it does have that appearance, a fact which not only supports our analysis of the experiments concerned but which also gives us a clue to the meaning of the dip in the f and g-curves. Consider our model for g. If g truly represents the percentage of the stimuli recognized correctly

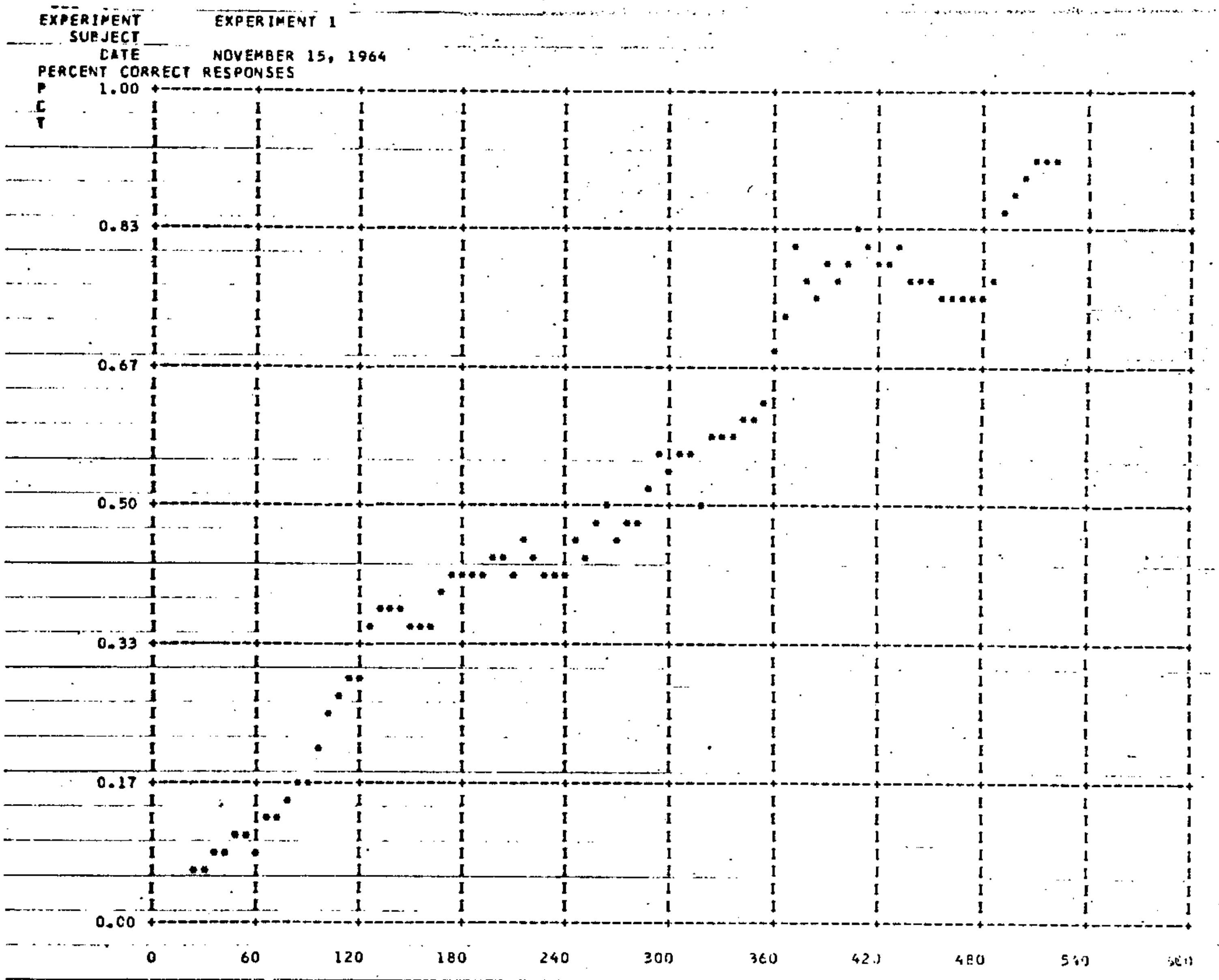

EXPERIMENT
SUBJECT
DATE
EXPERIMENT 1
NOVEMBER 15, 1964
PERCENT CORRECT RESPONSES
P
C
T
1.00
0.83
0.67
0.50
0.33
0.17
0.00
0
60
120
180
240
300
360
480

Figure 5-1

as wholes, it should be monotonically increasing over the entire experiment--unless the subject momentarily forgets or is unable to apply what he already knows. We may thus conjecture that the dip in the curves represents a period during which the subject's attention is devoted more to mastering the unknown stimuli than to performing well on those he has already mastered.

D. Relationship of Response Times to Other Measures

The dip in the f and g-curves on the "easy" experiments corresponds to a minimum in the overall response time curve. This minimum results from the addition of two different kinds of curves:

1. Response times to already captured stimuli which have been decreasing steadily since capture but which now stop decreasing or even increase; and
2. Response times to unlearned stimuli which have been relatively steady since the beginning of the experiment and which now begin to rise quite sharply.

These observations confirm our conjecture about the dip representing a transition point between the capture of the first few stimuli and the determined effort to concentrate on the remainder.

Once we have seen how well response time behavior correlates with one measure, we are moved to ask about other correlations. In particular, we are interested to see the relationship between response time to a stimulus and the history leading to the capture of that stimulus. Figure 5-2 shows a typical response time curve for a stimulus which was captured in stages. Its history is reconstructed as follows:

0. Up to point 1, no consistency in behavior to 001 can be detected. This portion of the response time curve precisely matches the overall average curve.
1. At this point, trial 188, the subject begins to recognize 001-00XX as a unique stimulus.
2. At this point, trial 258, all cases of 001 begin to be recognized--for they begin to be consistently identified as 010.
3. At trial 324, the confusion with 010 disappears and the subject tries three different responses to 001.
4. At trial 353, one-half of the cases (001-X1XX) are mastered.
5. At trial 400, one-half of the remaining cases (001-00XX) are mastered.
6. The remaining case (001-10XX) is cleared up at 516, but with the result that certain cases of 000 begin to be identified as 001.
7. The 000 cease at 555 and the subject makes no more errors involving 001.

Figure 5-3 shows a response time curve for subject 3, our ambidextrous subject. Though his curves are not as smooth as those of the other subjects, this one shows a typical curve for a response

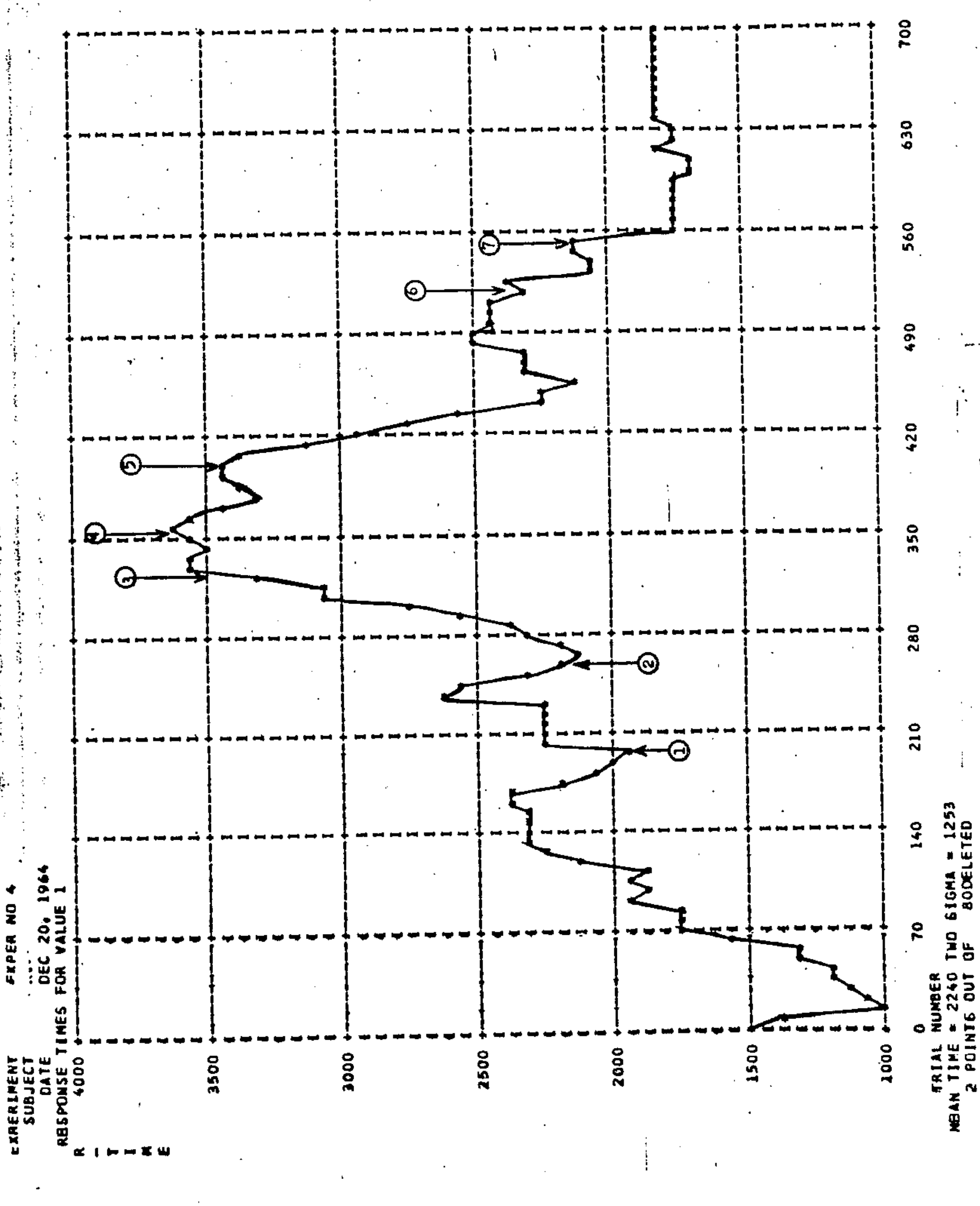

EXPERIMENT EXPER NO 4
SUBJECT
DATE DEC 20, 1964
RESPONSE TIMES FOR VALUE 1
R
T
I
M
E
4000
3500
3000
2500
2000
1500
1000
0
70
140
210
280
350
420
490
560
630
700
TRIAL NUMBER
MEAN TIME = 2240 TWO SIGMA = 1253
2 POINTS OUT OF 80DELETED
1
2
3
4
5
6
7

Figure 5-2

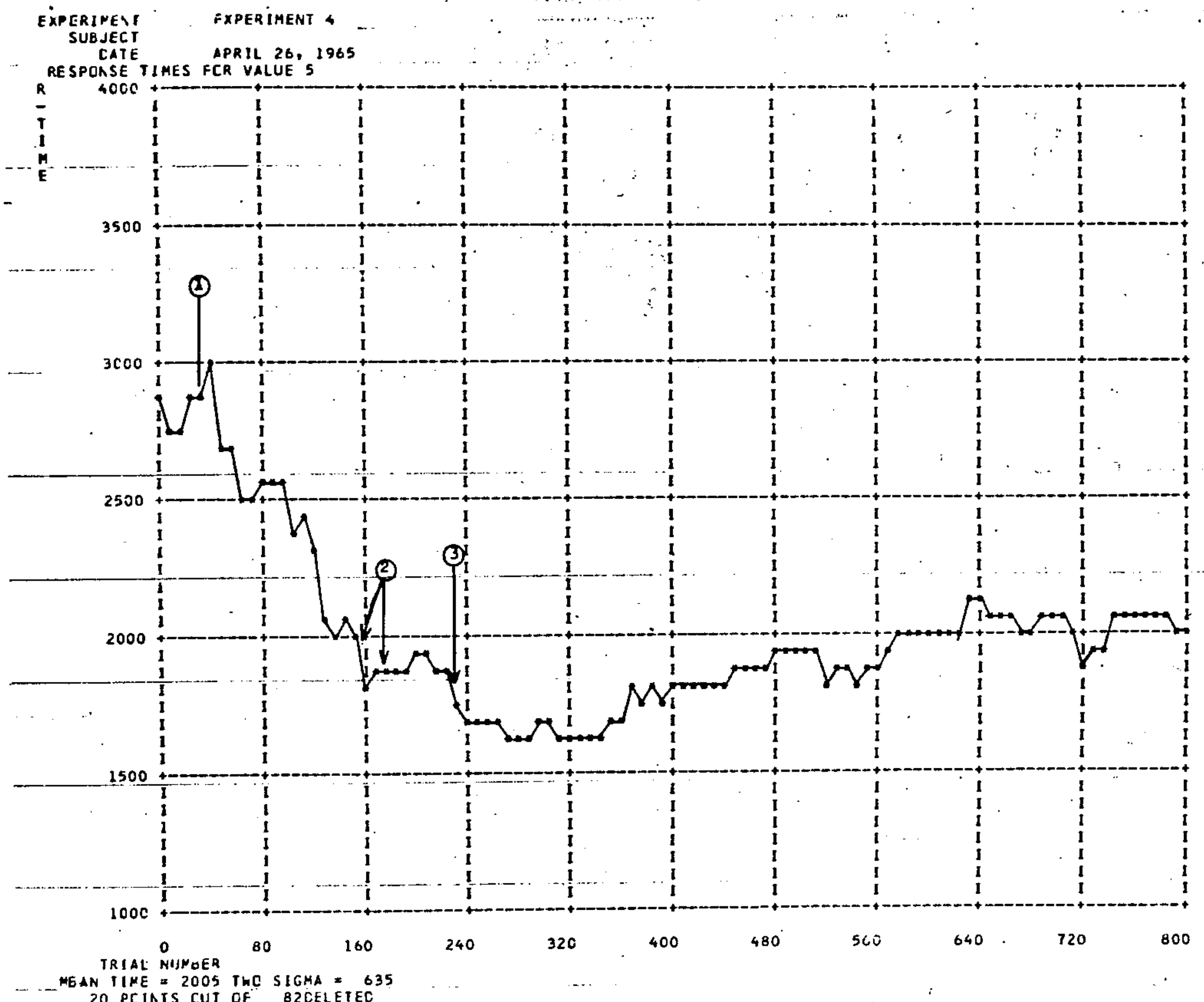

EXPERIMENT EXPERIMENT 4
SUBJECT
DATE APRIL 26, 1965
RESPONSE TIMES FOR VALUE 5
R-TIME
4000
3500
3000
2500
2000
1500
1000
0
80
160
240
320
400
480
560
640
720
800
1
2
3
TRIAL NUMBER
MEAN TIME = 2005 TWO SIGMA = 635
20 POINTS OUT OF 82DELETED

Figure 5-3

that was captured early and well. The curve is particularly striking in view of the overall response time curve over the same period, which varies between 2400 and 3400 milliseconds. The history of this response is quite simple:

1. At trial 30, 101 is mastered.
2. At trials 156 and 176, two cases of 101 are missed, after perfect performance previously.
3. At trial 229, one more error--the last--is made in 101.

These two curves--Figures 5-2 and 5-3--are typical of the multitude of response time curves from both "easy" and "difficult" experiments. Before a response is isolated, the time follows the overall time curve, as we would expect if the subject is unable to distinguish these stimuli from the others. As soon as it begins to be recognized as a unique stimulus, the response time slows--often as much as 2.5 seconds--until the first correct associations begin to be made. Then, the response time begins a rapid decline until it is checked by a few errors. Then it begins a slower decline which may also be checked or even reversed by errors. Eventually, the curve reaches a minimum, following which it remains steady or increases slightly--as in Figure 5-3. In other words, the subject seems to be adjusting his response time to the fastest he can without making errors in already captured stimuli.

As we have seen, the average time to give a particular response is not a reliable measure. On the other hand, if we look at the response time to the stimulus _following_ the giving of a particular response, we get indications of an interesting result:

> In the period when the stimulus class has been recognized but not yet properly associated with the response (e.g., between 1. and 3. in Figure 5-2), whenever the correct response is given--to any stimulus--the following response time tends to be over its average by about 300 milliseconds.

Although this result is not as unequivocally supported as our other results, it does seem reasonable in the light of the other findings. This slowdown indicates that the subject is also considering the response during this period--that after he makes it, he must steal a certain amount of time from the next trial in order to consider what happened when he made it. We do not, on the other hand, find such a slowdown after one of the stimuli in this class has been shown--unless it has been correctly identified.

These two results support our other results which show that the subjects prefer to use the N type events to gain information in this stage of their solution. The stimuli, being more complex, can best be dealt with when they are actually present; while the responses, being a much simpler class, can be kept in mind over several trials, or at least into the beginning of the next trial. This fact also explains the value of increasing the frequency of the response during this period; since such an increase makes N type events or chance hits on the right response more likely, thereby increasing the rate of acquisition of usable information.

Up to this point, we have used response time information in two different ways. First we have used the very large variations (of the order of a second or more) on individual trials to identify precise points of extreme difficulty--in test trials--or sudden transition--in ordinary trials. Once we have filtered out these impulses, we can investigate smoothed curves which show us the more gradual changes in behavior--changes which not only range over more than one trial, but which have a time scale more on the order of hundreds of milliseconds. Only when the shorter, sharper variations have been eliminated can we hope to make observations on this scale.

Can we carry this process of elimination one step further? If, in some way, we can filter out the 100 millisecond variations, we could observe effects on the order of 10 milliseconds--about the limit of measurement of our equipment. When we consider the physical aspects of our experiments, we see that because of variations in arm movements necessary to make responses, we are unlikely to be able to find any 10 millisecond data unless:

1. The subject generally uses a "homing" position before and after each response, and
2. The particular response being measured is well practiced.

In at least one experiment, we do approach these conditions. Subject 2 did generally use a homing position. Furthermore, on experiment 3, he captured half of the responses very early in the experiment but was unable to capture the others. Therefore, he was able to practice each captured response for several hundred trials--enough to engender highly consistent response times.

Since we have been able precisely to specify the placement of certain decisions in the decision structure for these four captured responses (Figure 3-19), differences in their response times could measure the time required to make the differentiating decision. Indeed, we do find differences in the expected direction. The mean time to make response 000 exceeds the mean time to make response 100 by .018 seconds, while the mean time for 010 exceeds that for 110 by .012 seconds. The standard deviations are large, however, and a difference of means test gives only a .20 level of significance to each of the differences. Nevertheless, this result points the way to a new set of experiments which could produce reliable measures of times for cerebral processes. Our general finding that responses deeper in the structure take longer indicates that there are, indeed, measurable times for the underlying cerebral processes--though we cannot entirely discount practice effects as a source of much of this time difference.

E. The Effects of Test Trials

Test trials were supposed to provide a probe into the subject's decision structure without disturbing that structure. When we examine the data, however, we find a number of evidences of disturbance. In the first place, response time seems to slow down a bit after test trials--or at the very least, the decline of response times for captured responses is temporarily arrested.

The effect of test trials on performance does not show such a clear picture. Sometimes we see clear evidence that performance on one response is hurt by the test trials--the subject forgets a discrimination he was just beginning to make correctly. At times, on the other hand, performance improves after the test trials, especially when the subject has been using an erroneous--but consistent--rule. In spite of appearances, though, these two different effects are actually the same.

We have seen that the decision structures are formed by building out at the tips of the branches and that forgetting also takes place at these locations. The test trials, being trials without reinforcement, give the subject an opportunity to forget part of his current decision structure--and that forgetting also takes place at the extremities. Since the subject has not yet completely solved the problem, these extremities are not necessarily right. They are, in fact, the hypotheses currently under test and have been the least verified or disproved of all the rules. Clearly the subject cannot yet tell which of these hypotheses are wrong--if he could, they would have been discarded. Thus, when he forgets one of them, he sometimes forgets a right one and sometimes forgets a wrong one--leading to worse performance in one case and better performance in the other.

F. Other Observations

In the course of testing and demonstrating the experimental apparatus, a number of people chanced to observe a "subject" at work on an experiment. Standing behind a subject at work, it is possible to get a perfectly clear view of everything the subject sees and does. Nevertheless, no observer was ever able to extract the information from an experiment that a subject could--even when he watched the subject from start to finish. Although this result can by no means be considered proved by such uncontrolled observations, it suggests an independent verification of one of our observations--namely, that each response the subject makes is considered by him to be a test of a hypothesis currently held. If an observer prefers to take information only insofar as the reinforcement confirms or denies his current hypothesis--and not from what it tells about alternative hypotheses--observers will be able to learn effectively only when the subject's current hypothesis leads to the same responses as the observer's would. As we have seen, it is quite unlikely that the two would hold the same hypothesis; thus the observer would be severely handicapped in trying to solve the problem. In other words, the subject has--from his point of view--some measure of control over the sequence of trials while the observer does not.

Although none of the subjects managed to alter his major assumption over the course of several experiments, each displayed some degree of attempts to carry learned concepts from one experiment to the next. Subject 2 in particular was able to improve his perception of the structure of the stimulus space--with a resulting improvement in performance over his five experiments. In addition, each subject at some time picked up ideas about the importance of some "key" letter or position. This idea was often carried across experiments, with the result that the letter or position was more likely to appear in the decision structure of the next experiment--or at least in the early hypotheses.

Finally, it is interesting to note, though we cannot quantify this observation, that all subjects were quickly able to see simpler rules that were pointed out to them at the end of their last experiment. In spite of this ability, they were all loath to observe that there were a number of alternative rules available to them--that somehow their solution was not the "true" one.

CHAPTER VI

SIGNIFICANCE OF THE RESULTS FOR OTHER MODELS

Concept learning, inductive learning, or problem solving--whatever term is used--shares with cosmology the property of having more theories than data. We have tried to avoid adding to this unfortunate condition. Rather than try to build precise models on the basis of twelve experiments on four subjects (though they may be considered in many ways equivalent to 96 experiments of the ordinary kind), we have kept our analysis on the pretheoretical level so admirably used by Bruner, Goodnow, and Austin (1956). In this form, it may be difficult to relate our results to the more prominent theories of learning in complex situations. Therefore, we undertake this task in this, a separate chapter. In doing so, we may further clarify our own picture.

A. Basic Learning Theory

Stimulus-response theory seems to be the doormat on which each new entry to the arena of concept learning must wipe its feet. Although a vast number of different theories and sub-theories all vie for the title of "basic learning theory" (Hilgard, 1956), the generally accepted view (outside of basic learning theorists) is that the continuous nature of conditioned learning makes it unable to stand as an explanation for the seemingly discrete events of concept learning. Rather than using continuously growing habit strength (Luce, 1959), "all-or-none" models (Estes, 1950; Bower, 1962) visualize the concept learning as something that either happens or does not happen on a single trial. Once the learning has taken place, it is irreversible. Before it happens, the subject performs randomly; afterwards, he is consistently right. This model has been shown to give striking fits to the data from concept learning experiments (Bower, 1962) and, at first glance, seems very much in accord with our results.

We have seen how "capture points" can be defined and isolated quite precisely. These are the points postulated by the all-or-none theory. Even if we sometimes find them for only parts of a stimulus class, their existence still supports the theory--which says nothing about the division of the stimulus class. Indeed, the existence of split stimulus classes may have tended to obscure the all-or-none nature of concept learning in other experiments. In our experiments, we had to define such ideas as "pair-capture" in order to make the idea of capture points work; but when we did, much of the behavior was accounted for. If we had not been able to reconstruct the subject's behavior in such detail, we should not have been able to identify such things as pair capture points.

In further support of the all-or-none theories is our observation of the frequency with which certain structures (type P and N) are found at the capture points. It may be that the probability of capture at any given point (the basic parameter of all-or-none theories) can be partly accounted for by the probability of such "helpful" events. Thus, for instance, the learning rate constant, c, might be expressible as the sum of a constant term, c', and a term, k, which depends on the likelihood of helpful events. Since k could vary in a known way, its existence could be tested in a controlled series of experiments in which the number of helpful events was the independent variable.

Once we are over the initial impression of how well the all-or-none model describes certain aspects of our results, we are moved to reconsider the basis for that success. Suppose that, in fact, the habit strength of a particular stimulus-response connection were changing continuously--or at least gradually--with each new paired presentation. Since the response is competing with other responses at all times, it must reach a given strength before it emerges above the general level. If the emerging response is the correct response, it is unlikely ever to sink below the levels of the other responses. Thus, a gradual rise in habit strength might easily be seen as a sudden appearance of the correct response. As an analogy, consider a submarine rising to the surface. When it breaks the surface, we say that it "suddenly" appears, even though it may have been rising slowly for hours. Carrying the analogy further, if the day is calm, the submarine is likely to remain in view from the moment its periscope breaks the surface. If there are waves, on the other hand, it may disappear from time to time until it is fully emerged.

Do we have any evidence that a similar phenomenon might be operating in the learning of the individual concepts? In the first place, certain responses do not show sharp capture points--but wavy ones as might be expected in the continuous model. We can further argue that in the concept learning experiments cited as fitting the all-or-none model by Bower, only two choices of responses were available to the subject. Since the conditioning to one response is complementary to the conditioning to the other, we would expect the emergence of the "right" response to be sharper--just as we would expect the emergence of a large submarine from a small pond to be sharper because the water level must fall as the submarine emerges.

Our most striking support for a continuous learning model, however, is found in the response time curves. In psychology, it is always valid to ask, "what is being learned?," when someone is talking about learning. In almost all of the experiments reported in the literature, concept learning is taken to involve learning certain deterministic rules relating stimuli and responses. Thus, in a way, the experiments exclude the possibility of continuous learning a priori. Why not, as in other learned behavior, consider the speed of response as part of what is learned? In fact, why not consider the response in general to be a vector--some components of which are measured in any particular experiment?

If we do consider the speed of response, our results clearly show that all learning does not take place at one "capture point." After the capture point, the subject steadily improves his performance; for though he cannot do better than 100 percent right, he can increase the number of right responses per unit of time. In any reasonable model of an adaptive system, the rate of payoff is a critical parameter. Even more significant is the slowing down of response time before the capture point. If the subject had learned nothing before this point, he could hardly slow down his responses to one stimulus class and not the others. (In a two choice experiment, this effect would be hard to detect, even if the experiments chose to measure response times, because of the delaying effect critical cases have on following cases.)

In spite of the support our data give to the role of stimulus-response models in concept learning, there remain a number of facts for which simple conditioning cannot account. A number of these criticisms have been leveled on the basis of other concept learning experiments (Hunt, 1962) and need not be reviewed here. In our experiments, two new pieces of evidence arise that simple conditioning would have a hard time explaining. The first problem is created by the serial nature of the solutions to our problems. A pure conditioning theory would require that stimulus-response pairs which were equally reinforced should reach the same level of mastery at about the same time. Since all pairs were equally reinforced, the S-R theorist would have to explain why all eight concepts were not learned simultaneously. Granted that, with the addition of such artifacts as "mediating responses" (Kendler and Kendler, 1962; Goss, 1961), such behavior can probably be explained. But in S-R theory, mediating responses--like God--probably explain too much.

No doubt mediating responses could be used to explain away the other major difficulty our experiments raise for S-R theory--namely, the picking of stimulus cues which were not as well reinforced as others. For instance, if a noise bit is present in a letter position, that letter will have two different values for the same response. If, as Bourne and Restle (1959) suggest, the most reinforced cues will be used when alternative cues are present, split letters would never be used as cues when single letters conveying the same information were present. Our results, in which each subject at some time or another uses such split letters in the presence of single letters, clearly contradict this prediction, for each of the split letters has received only half as much reinforcement as the single letter. In fact, we get some cases where cues with only 1/4 as much reinforcement (the double letters of subject 3) are chosen over single letters. This effect cannot be explained away by arguing that the subject prefers a certain modality, because the information is available in the same modality elsewhere--sometimes even the same letter as one of the split letters, so that a priori letter preferences cannot account for it. Similarly it cannot be explained by saying that the other cue was not in the subject's stimulus field, because we have cases where the other cue is used for other responses in the same experiment. Clearly, S-R theory is seriously challenged by these results--if it attempts to provide a complete explanation of concept learning.

B. Perception Theories

For a learning theorist, the stimulus and response are primitive undefined terms--fixed throughout the region of interest. What he is trying to do is discover the laws concerning the development of the transformations between fixed stimulus and response. Thus, if the stimulus or response is changing, his theories cannot be expected to give an adequate description of behavior. One of the most fundamental ideas in psychology is that conditioning takes place between the stimulus trace--the subject's internal representation of the stimulus--and the response trace--the subject's internal representation of the response--(Hull, 1943; Spence, 1957; and Logan, 1960). (Although "response trace" has not been given the same attention as stimulus trace by students of perception, the idea is important to the interpretation of our results.) What the idea of stimulus trace accomplishes is the interposition of another stage of possible learning--learning about the mapping between object and stimulus--which allows the possibility of more complex learning behavior.

At this point, it would be appropriate to mention the theories of stimulus sampling (Estes, 1950; Bush and Mosteller, 1955), which represent a sort of half step between pure S-R theory and the more elaborate theories of perception. Stimulus sampling theories, rather than taking the stimulus as an indivisible primitive, consider the stimulus as being composed of a number of primitive stimuli. The learning behavior displayed by a subject can thus be modeled as the summation of the conditioning of a response to each of the molecular stimuli. A fundamental assumption of most stimulus sampling theories is that not all stimulus parts are conditioned on any one trial (otherwise there would be no need to decompose the stimulus). Furthermore, those parts which are to be conditioned are chosen by random sampling among all possible parts. Although stimulus sampling theory does not say just what the parts of the stimulus are, it seems clear that they would not be subdivided by the character boundaries in our stimuli. That is, no primitive element would lie partly in one character and partly in another. Under this interpretation, our data shows quite clearly that random sampling does not take place. As a matter of fact, the subject's biases in selecting parts of the stimulus to attend to are a crucial determinant of his behavior.

In a sense, the Gestalt psychologists (Wertheimer, 1959) are at exactly the opposite position from the stimulus samplers, for they maintain that analysis of the stimulus into parts destroys any possible understanding of it. Does our refutation of the stimulus sampling position thereby support the Gestalt position? Actually, in our experiments we find cases which support and cases which refute the Gestalt hypothesis. Subject 1, for example, on experiment 2 showed an ability to fill in for missing stimulus parts--completing the whole in good Gestalt fashion. Yet the same subject on experiment 1 made precisely predictable errors when individual parts of the stimulus were deleted. In all this, of course, we do not even consider the question of the non-Gestalt behavior involved in limiting perception of the stimulus to precisely selected portions--or the slight evidence we have of a subsequent spreading of perception outside of these portions.

Neither stimulus sampling theory nor Gestalt theory can adequately account for the kinds of transformations our subjects use between stimulus and stimulus trace. One of the difficulties they share is that they both assume that the entire stimulus is used and that it is used in the same way throughout the experiment. Our analysis, however, shows a quite different picture. Indeed, the first--and longest--part of the learning of any one of the responses is finding a part of the stimulus which is relevant to it. Once the relevant part is found, a different phase of behavior is inaugurated--a phase in which the stimulus does remain constant, a subset of the entire stimulus.

According to our description of what is happening in the first phase, the subject is searching for the "proper" characterization of the stimulus. At least two factors would contribute to the length of this search: the probability of finding a "proper" set of dimensions and the difficulty of verifying its propriety. The difficulty of verification should remain relatively constant over all of our experiments, but the probability of finding a relevant dimension varies with the redundancy with which each bit of the response is represented. Suppose h = the mean number of trials necessary to test a set of dimensions and establish or discard it.

Let p = the percentage of the dimensions which are appropriate to a particular bit. Then the mean number of trials, t, to select an appropriate stimulus would be

$$t = h/p$$

Of course, using the mean number amounts to assuming the subject chooses new trial dimensions in a random fashion with respect to the manner in which the dimensions were assigned. If the dimensions are assigned at random, this assumption will be valid regardless of the subject's strategy. This model also assumes that the subject does not retry dimensions once they have failed. This assumption is probably not valid, especially as the number of dimensions increases. Thus we would expect actual behavior to be worse than this model predicts, since any repetition of a previously discarded dimension has zero probability of success.

Although our data suggested this model, they alone are not sufficient to test it. In the literature, however, we find much simpler concept formation experiments which provide adequate data. Bourne and Haygood (1959) performed a series of experiments specifically designed to test the effects of stimulus redundancy. The subjects had to choose between two responses based on a stimulus with two to eight dimensions, some of which were relevant and redundant and some of which were irrelevant and redundant. Bourne and Restle (1959) attempted to fit Bourne and Haygood's data using a conditioning model with two free parameters. Their fit to this data is shown in Figure 6-1a.

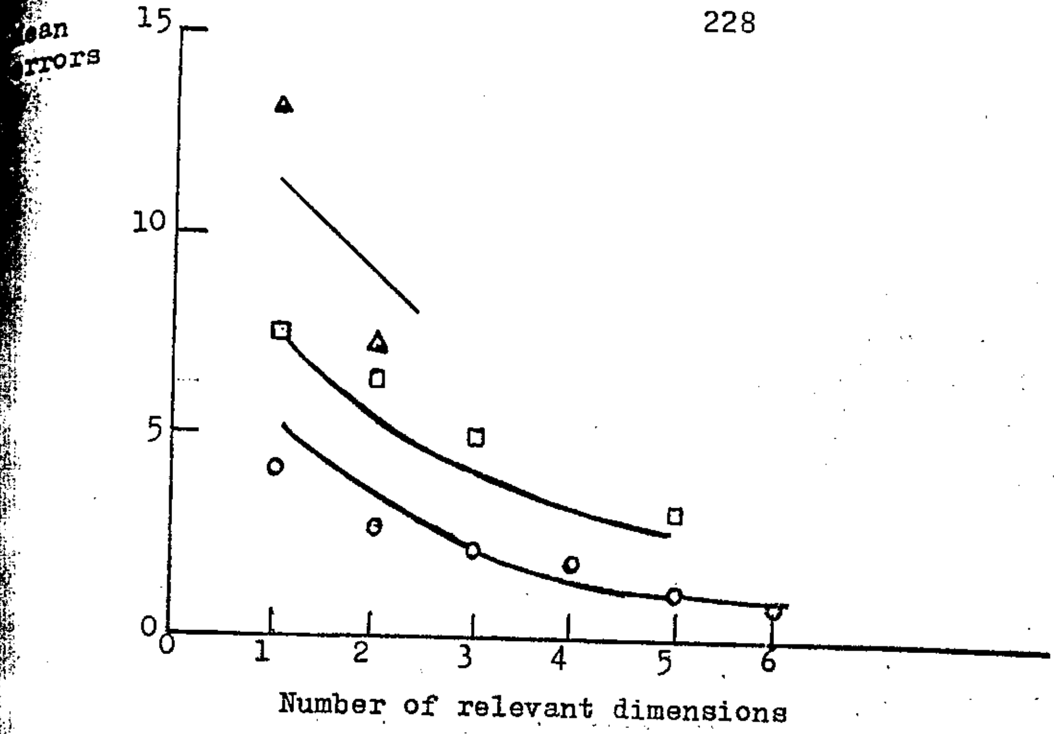


Figure 6–1a

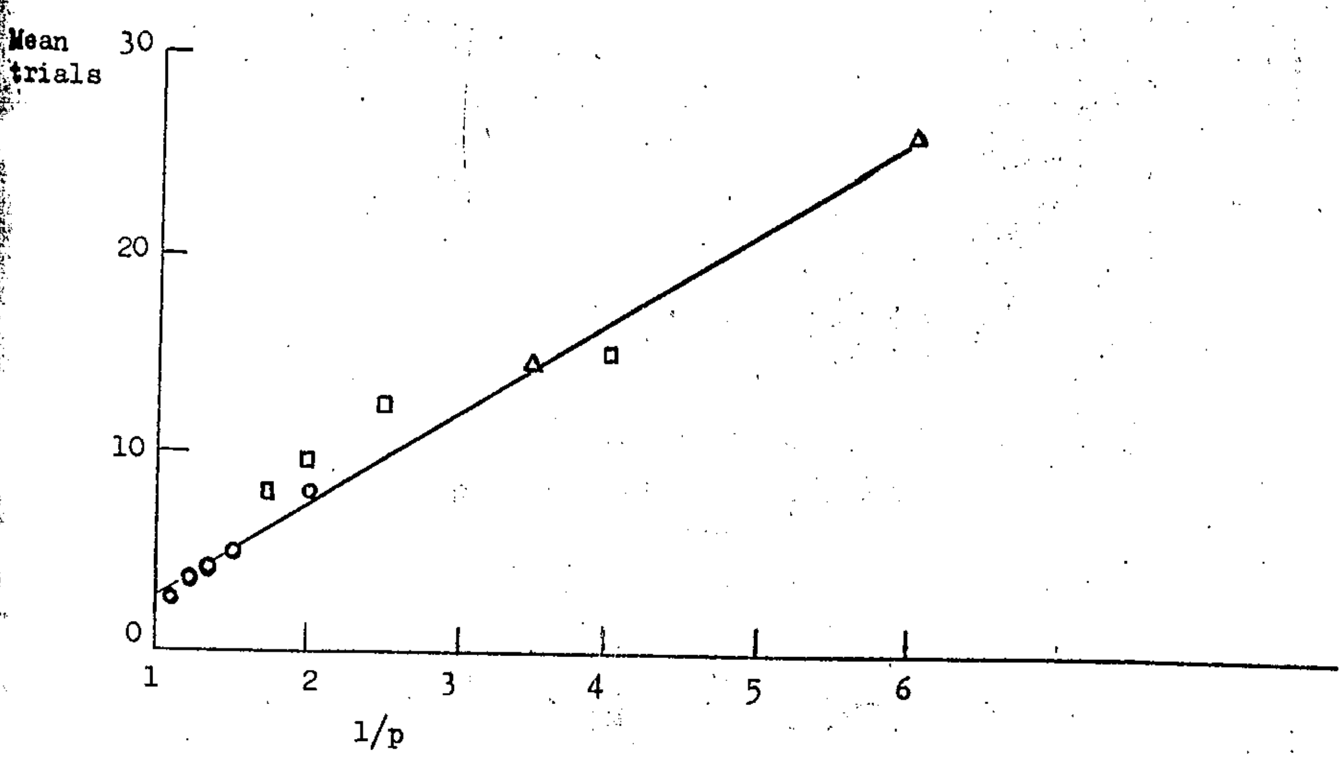


Figure 6–1b

In order to try our model on the data, we have to transform their "number of errors" measure into a "number of trials" measure. Since their subject has only a binary choice, all-or-none conditioning would indicate that the number of trials is just twice the number of errors. Using this transformation, we replot these data in Figure 6-1b, using 1/p as the ordinate so the slope of the line will estimate h. Clearly, this crudest of models--using only one parameter and consolidating the three curves into one (triangles represent five irrelevant dimensions; squares, three; and circles, one)--fits the data better than Bourne and Restle's model. The line in Figure 6-1b is fitted by eye. It gives a slope of h = 1.6, which seems reasonable inasmuch as the minimum number of trials to test a dimension is one, unless parallel testing is done.

If we average our data on the number of trials to reach and maintain the three sigma level for each bit, we find that we get estimates of h ranging from 7.5 to 11.0 for the experiments of 17-9-6 distribution and from 5.0 to 6.9 for the 11-10-11 experiments. The 17-9-6 figures are probably high because they include all of the first experiments. If we allow for an "adjustment" period for those experiments of 250 trials, we get estimates of h = 9.0, 6.5, and 9.0. We cannot readily compare these estimates with the h's for Bourne and Haygood's data because our three sigma level does not correspond to their time of last error. In their experiments, the end of the search period would also be the end of the experiment, for as soon as the subject found a dimension that worked, he was finished. In our experiments, the subject could never be so sure that he had found the right dimension; nevertheless, the three sigma level probably approximates the average point of selection for the eight different responses which use each bit.

The fact that our estimates of h are all less than 8.0--the average number of trials to see any particular response--might be an indication that the subjects have learned something of the structure of the response space or--more likely--that they have a preset tendency to try the same dimension on more than one response. A possible explanation for this tendency lies in the presence of the letters in our experiments which may be two, four or eight-valued dimensions. Use of letters as dimensions would thus tend to tie the acquisition of one stimulus set to another, thus reducing h. Reduction of h in this manner would lend support to the Gestalt position.

C. Information Processing and Coding

One of the more sweeping papers in recent psychology is Miller's "The Magic Number Seven, Plus or Minus Two" (1956). In this paper, Miller considers the ways humans recode information from several dimensions into single dimensions. In general, the ability to recode stimuli or decision structures into more manageable forms would enhance a subject's ability to handle problems of concept learning. To what extent does recoding show itself in our experiments?

Perhaps the most important observation is that our subjects seem to prefer binary dimensions for use in working out their concepts. Indeed, even when they use the dimensions with four or eight values, they seem to "unchunk" them by considering only the presence or absence of one of the possible letters. It is as if the subject perceives the stimulus so that only binary decisions are added to the decision structure--perhaps because these additions are only tentative and subject to testing. It seems as if any recoding that might take place would occur after the binary decision was firmly attached to the decision structure.

Another prediction of the information processing idea of human coding capacity is that dimensions which carry more than seven (plus or minus two) variations either have to be recoded by decomposition or only discriminated approximately. Our response space is right at the upper limit of this region, and we do indeed see that the two or three less prominent response squares tend to be confused. Furthermore, as the subjects proceed, they show evidence of restructuring the response space so that fewer discriminations (upper left, top middle, and so forth) are required. As they do this, confusions among the responses disappear. Furthermore, to the extent that this recoding fits the underlying structure of the response space (left and right, for instance), the task of learning one concept becomes more tied to the others.

Recoding within the decision structure is a matter of some interest. Much of the literature on concept learning has been devoted to the question of conjunctive versus disjunctive concepts. Certain types of strategies would seem to favor one type of concept or another, and experiments (Hunt and Hovland, 1960; Wells, 1962) to test for strategy preferences have shown that conjunctive concepts are generally chosen when a choice is available. In our experiments, we are able to get beneath the simple classification of conjunctive versus disjunctive and throw more light on the issue. One clear finding is that many of the disjunctive concepts are actually the composition of two conjunctive concepts separately learned. Piaget (1957) calls this type of concept "pseudodisjunctive," and regards it as more primitive than the "true" disjunctive concept. Our data also suggest that there may be a developmental relationship between pseudodisjunctive and disjunction concepts. When two different concepts for the same response develop as successive decisions on the same branch of a decision structure, they may merge and become a single disjunctive decision; but when the two decisions are separated by one or more intervening decisions or are on different branches, they do not merge. These effects are shown in the verbal reports and are verified by response time data and logical consistency.

One of the favorite forms for our subjects has not been reported in the literature, perhaps because few experiments have used more than two responses. This form is "A and not B" and, as we have seen, it develops from a decision which is too inclusive being narrowed by the negation condition. Other forms and transformations of those forms have been discussed in Chapter V. No analysis of this type has been possible with earlier experimental techniques, so that little in the way of theory has been developed. The existence of favored and disfavored transformations on logical relations, however, is suggestive of Chomsky's (1953) work on transformational grammars. Cataloging these transformations might be a useful line for future research, particularly in relating concept learning to verbal learning.

D. Statistical Models vs. Algorithmic Models

When we examine the various theories of concept learning and the experimental results rallied in their support, we are struck by what might be termed the "university effect" in the measures used to evaluate performance. University professors spend a goodly portion of their time evaluating the performance of students. Much of this evaluation is based on the student's ability to "solve problems," and since large masses of students usually have to be evaluated, mass measures have to be applied in order to reduce the task to manageable size. Essentially, the student is graded according to the number of problems he "solved" or the number of "errors" he made; no classifications exist between these extremes.

When the professor leaves his classroom and moves into the laboratory, perhaps he carries a certain prejudice. Although he starts out to discover how people solve problems, he quickly transforms his research into an evaluation of how well people solve problems. Consider the measures used in concept formation experiments: number of subjects solving problem, average number of errors before solution, number of trials before solution, number of instances of given types before solution. Each of these measures insists upon a solution before the behavior can even be considered. If, for example, one subject fails to solve a problem, the average number of errors to solution for the group must be infinite. If we insist on solution-based measures, we must either throw such subjects out of the data or design the experiment so that it is so simple we can be sure in advance that every subject will "solve" it.

An example of this orientation and what it leads to is the method advocated by Hovland (1952) to avoid "problems" caused by the subject and the experimenter having different ideas about the structure of the stimulus. He states that the experimenter and subject must agree on a definition of dimensions and values before the experiment begins. In this way we are assured that each subject is solving the "right" problem--but we are also assured that most of the "problem" has been eliminated. One of the strongest results from our experiments is that the major source of variance in problem solving behavior is a difference in preconception about what the problem is. Once we are given the subject's basic concepts which he brings to the experiment, we can predict quite well what kind of difficulty he will experience.

But does this argument not prove that Hovland's admonition is right and proper? Not at all, for Hovland himself must have been aware that he was wishing rather than admonishing. There is no way to be certain that all subjects share the experimenter's view of what the problem is, and the more complex the problem we are dealing with, the less chance there is of anything being shared among the subjects. If we want to study really complex behavior, then, why not give up the futile attempts to force the subjects into a Procrastean bed and design experiments which measure their differences, not conceal them?

To take this approach, we must abandon the idea of measuring "success," at least in any fixed way, because each of the subjects--having a different view of what the task is--is working toward a different measure of success. Consider, for example, subject 2 working on experiment 1. During the first few hundred trials, he was testing various major hypotheses about the structure of the experiment. In particular, he was entertaining hypotheses about sequences, rather than about stimulus-response associations. Among other things, this led him to discover that in the test trials the characters were deleted in a specified sequence. Before we dismiss his behavior as irrelevant, we might consider that the major task in solving real problems is discovering what is relevant and what is not relevant. Nobody told Galileo that his task was to find out how many moons Jupiter had; if he had been told, his discovery would have been a trivial matter.

Viewed from the proper vantage point, subject 2's discovery would be a "solution," and the other subjects would have failed if they failed to notice and mention to the experimenter that the test trials occurred in sequence. If we allow the experimenter's definition of "solution" to dominate our analyses, we shall never even see the most interesting behaviors. Is there no difference between a subject's applying a theorem to a particular set of data to see whether it holds, proving a given theorem and conjecturing an entirely new theorem where none existed before?

Another good example of the pitfalls of looking for statistical right and wrong measures is given in a study by Hunt and Hovland (1960). In this study, as in ours, the experiments were explicitly defined so that more than one rule would correctly categorize the stimuli. Hunt says, "As an interesting sidelight, most subjects appeared amazed to find that there was more than one answer." We observed this same phenomenon, but we did not dismiss it as an "interesting sidelight." If we can establish that the concept of "the right answer" is essentially universal, we will have made a major contribution to understand problem solving and its difficulties. Observe how much easier the task on experiment 5 was for the experimenter because he was not afraid to abandon partial success to search for a simpler solution. Any "real" problem, in mathematics or in any other field, has many solutions. Any problem of adaptation has many local maxima on which a system might be trapped if it did not have the ability to look for maxima in more than one way. Yet a simple error tabulation gives us no power to distinguish between the subject who makes a quick solution because he has freed himself from the concept of "the right solution" and the subject whose more rigid concepts happen to match those which the experimenter had in mind.

Just as the implicit model of the student in the classroom influenced the classical direction of concept formation studies and models, the existence of the computer as a potential problem solver has given studies a new direction. In working with computers, the program becomes the primary focus of our interest and understanding. It is not surprising, then, to find that the new class of concept formation or problem solving models are cast in the form of answers to the question, "What sort of program or algorithm would exhibit the kind of behavior we see in concept learning experiments?" Once this question has been taken as a goal, a new way of doing experiments and analyzing data becomes imperative. If we want to construct a deterministic program, no element of behavior can be dismissed or hidden in the statistics.

The algorithmic model must be distinguished from other models which are essentially statistical but which can be programmed and simulated on computers. Although, as computer models, these too are deterministic, they do not attempt to identify individual steps in the process of learning a concept (or recognizing a pattern, which is the same thing) with the verbal steps which problem solvers report in their protocols. In fact, there is nothing essential in these models which prevents them from performing their "computations" in parallel, even though, to be simulated on a computer, they might have to be programmed in a serial interpretation.

The possibility of parallelism makes these models attractive as interpretations of some of our results which the serial algorithmic models would be hard pressed to explain. One whole class of these models (Samuel, 1959; Selfridge, 1959, Kochen, 1961; Uhr and

Vossler, 1961) is based on carrying out a weighted summation for each possible response, the largest sum for a stimulus determining the response given. Obviously, such computations could be carried out in parallel and present possible neurophysiological interpretations.

On the other hand, these models tend to produce an all-or-none learning of concepts or patterns, and thus are weakened by our finding that concept learning only <u>appears</u> to be all-or-none. Still, they do admit of a rather simple modification which would produce the type of response time behavior we have found. Suppose the criterion for choosing a response was not the final result of the summation, but rather the result of an intermediate test applied after one or more terms had been added. Suppose, for instance, that the criterion was to choose a response as soon as it had a partial sum which was a certain amount, d, above the others. If one response had an <u>a priori</u> weighting which made it much larger than the others, it would tend to be the "don't know" response and be made rather quickly in the beginning. As the weights for some other response began to grow, it would take more and more terms to create a difference of d between the two sums. For a certain period the time would be maximum, and the probability of giving one or the other response would be about the same. Finally, as the weights for the correct response continued to build, the response time would grow shorter at the same time that the probability of giving the correct response became almost a certainty.

To illustrate this effect, consider the following two-choice, deterministic model. The test value for each response is

$$y_j = \sum_{i=1}^{x} a_{ij} \text{ , } j = 1, 2$$

A response is chosen whenever

$$\left| y_1 - y_2 \right| \geq d$$

Now, in the simplest case, if the a_{ij} are fixed and the a_{ij} grow linearly with the number of trials, t, the decision criterion may be rewritten as

$$\left| ax - xkt \right| \geq d$$

Since x is the number of terms in the sum, it is always positive and may be taken outside the absolute value. Thus, the criterion expressed in terms of the number of steps in the summation is

$$x \geq d/\left| a - kt \right|$$

The response time to any given stimulus should contain a constant term plus a term proportional to the number of summation steps, x, so this simplified model predicts a response time of

$$T = T_0 + xT_1 = T_0 + dT_1/\left| a - kt \right|$$

which has precisely the type of behavior we find in our experiments.

Algorithmic models, on the other hand, could not use such a simple modification to incorporate response time behavior. In such models, one would have to postulate changes in the basic decision times to account for the rising and falling of response time. Such a procedure is not as parsimonious as the one outlined above.

E. Algorithmic Models and Verbal Reports

Algorithmic models of learning allow themselves many more degrees of freedom than does the typical statistical model; but in return, they promise to account for behavior in far more detail. In particular, they hold out the hope of explaining not only each individual trial of an experiment, but of explaining it in terms of what the subject says he is doing. In other words, they promise to relate the major steps in their algorithms to the kinds of procedures people say are involved in thinking--generalization, testing hypotheses, searching for clues, getting hunches, and so forth. In this sense, such models are in the spirit of the data we have taken and the verbal picture we have constructed from them.

Quite likely the greatest danger in building such algorithmic models is the reliance on erroneous verbal reports. Newell, Shaw, and Simon (1959), for example, constructed their General Problem Solver on the basis of protocols taken from a number of subjects. From our experiments, the dangers in such protocols are obvious: First, the verbal reports are often simply wrong when they try to tell what the subject is doing; and second, the verbal reports seem to follow, rather than lead, the process of learning. For instance, we have seen how subjects develop language for speaking of the structure of, say, the response space. Once developed, this new language may be used in speaking of procedures carried out before it was developed, which throws them into an entirely erroneous light. Of course, we have also seen how terminology might be used to trace an accurate protocol because of the tendency to use terminology related to the genesis of certain concepts. Nonetheless, it is doubtful whether such subtle information could be used properly were our objective analysis not available.

Whatever they have learned, subjects can be counted on to give some verbal report. Until we know more about the transformations which subjects tend to make in speaking of their hypotheses, we can use such specific information only as suggestive material. Yet even if all verbal reports were _ex post facto_, we would remain interested in relating them to the underlying processes which are identified by different terms when they force their way into consciousness. For instance, some theories (Watanabe, 1960; Kochen, 1961) postulate the existence of some measure of confidence that a subject has or ought to have in his hypothesis. Polya (1954) has also emphasized the role of such an "index of plausibility" in his descriptions of problem solving, though he has given no quantitative form to it. Since it is known from discrimination studies that response time can be an indication of _confidence_, we might interpret our response time curves as measuring the subjective index of plausibility.

Looking at the response time curves in this light only accounts for the decline in response time after the response is captured. What can be the meaning of the rise before the capture? Rapoport (1964) suggests that:

> "The acquisition of insight is an expansion of one's ability to encompass situations directly, bypassing

> analysis. Concomitant to an insight is an increase in receptivity to certain stimuli..."

The selective retardation we have observed is just what we might expect as an indication of "an increase in receptivity to certain stimuli." Rapoport goes on to say:

> "To gain insight into the nature of a problem is not necessarily to solve the problem....With insight one is led to ask the right <u>questions</u> about the problem. These questions may lead to an answer, but they may also lead to a new insight, for example, that the problem is unsolvable and must be reformulated."

We might rephrase this argument in terms of our experiments:

> The process of capturing a response begins with an increased receptivity to certain stimuli. This process may be identified with "insight." Once receptivity has increased to a certain point, the insight (or hypothesis) may be explicitly tested. This testing leads either to an increase or a decrease in the plausibility of the insight. If its plausibility is sufficiently increased, it becomes a firmly established part of the solution.

In a similar manner, we have earlier identified what subjects might be talking about when they speak of the computer being "helpful," or what they mean by a "conspicuous" stimulus or a "key" letter. Ultimately, using information like this, we might be able to design algorithmic models which not only solve problems but which "talk" about what they are doing as they solve them.

F. Algorithmic Models and Objective Behavior

The most obvious question to ask about the various algorithmic models (such as, Newell, Shaw, and Simon, 1959; Hunt, 1962; Kochen, 1961) is whether they could "solve" our experiments, (Naturally, we assume that suitable input-output adaptations or interpretations would be made.) A cursory examination indicates that the answer is "yes," though the only way really to tell would be actually to execute a simulation. In fact, however, "yes" is probably a bad answer for a program to give if it hopes to be regarded as a simulation of human performance, for we observed that not all subjects could solve all experiments. Why would these models fail to match the human behavior?

In the first place, we have evidence that current models may be too rigid. For example, Hunt's model produces decision <u>trees</u> as a result of its learning algorithm. Although our data indicates that decision structures are often in the form of trees, we also have evidence that a "forest" (Riordan, 1958) is a more appropriate description of others. Another expression of this rigidity is the way in which some of these models never waver from a particular substrategy until it is worked out to its end result--positive or negative (Newell, Shar, and Simon, 1959). Our subjects, on the other hand, often are unable or unwilling to stick with a certain line of reasoning to its conclusion, perhaps because some conspicuous stimulus has diverted their attention.

At the same time that they are too rigid, many algorithmic models are too flexible. For instance, these programs always discard a refuted hypothesis immediately. Models such as these could not account for the rigidity which the decision structures obtain once they have passed preliminary screening. Of course, the models could never obtain a wrong element in a decision structure at all, but by the same token could never solve a probabalistic problem. In a way, then, their flexibility is a form of rigidity, for their changes in behavior are rigidly controlled. Samuel's (1959) type of problem solving model, on the other hand, seems more reasonable in that it really does not "solve" a problem--it just keeps trying to improve what it is doing.

Perhaps the most important difference between Samuel's model and the algorithmic is that Samuel's has much less idea of what is the "right" way to operate. That is, the problem it is working on is _much_ too big for it, not just a little bit too big. The other models would behave on our problem more in the way the experimenter performed than in the way the subjects managed to do. Since assumptions are so important to the subject's behavior, we should like to see models in which the assumptions were made more explicit than in Hunt's model and less flexible than in the GPS, where the program is willing to change assumptions all the way up the decision structure at any time.

None of these criticisms should be taken as a refutation of the algorithmic approach. On the contrary, criticism is easy only because this approach comes much closer to describing the kind of process our experiments allow us to see in human problem solving. For example, one aspect of our experiments which can be predicted quite easily from Newell, Shaw, and Simon's analysis of a problem solving heuristic is the way a subject fails to improve on a particular response if he happens to get a 3/4 rule. One of the critical points in their analysis is the test for difference between present state and desired state. (This test also assumes a critical role in Miller, Galanter, and Pribram's (1960) descriptive theory of human behavior.) It seems reasonable to postulate a testing mechanism which operates on a running average performance, in which case there would be a certain percentage of success which it could not discriminate from perfect performance. In the same way, this test would have to obtain a certain rise above random performance before it could signal "success" to the other parts of the process. Viewed in this way, the model does exactly what a good model should do--suggest future experiments which could either refute it or refine its predictions.

Another place in which our results give support to the flavor of the simulation models is in our finding that learning takes place only at the tips of the branches of the decision structure. (We mean, of course, learning in the error sense and not in the response time sense. Response time improvements occur throughout the structure, though most markedly while the appropriate decision is at the tip.)

One of the persistent problems of algorithmic models is that of hypothesis development. Kochen (1961) has described an entire system which develops hypotheses or guesses in concept learning problems. However, like the hypothesis formation mechanisms in the more complete problem solving systems, Kochen's mechanisms do not ring true as descriptions of the human processes involved. (Kochen, in his defense, made no claim that his system was supposed to simulate human behavior.) From our experiments, we get the impression that hypotheses are not so much formed as they are found. The subjects behave as if they had a very limited and unreliable "active" memory in which "hypotheses" were stored. Much of the strategy employed seems to be directed to the task of reducing the reliance on that memory over periods longer than two or three trials (up to 15 seconds). The subjects utilize conspicuous patterns and fortuitous sequences of patterns to keep the quantity of material stored and length of storage to a minimum. Once a definite association between a particular stimulus and response has been formed, however, the subjects seem to transfer that association to a somewhat larger and more permanent memory--the place where the decision structure are kept. Once there, the association seems much more difficult to remove--but also much less likely to be forgotten if it is correct.

The lack of capacity and veracity on the part of this short-term hypothesis memory seems to be the dominant factor in performance on the tasks of these experiments. On the simple concept learning experiments performed in the past, many subjects had been able to execute strategies which were as optimal as an infinite memory computer could have done (Bruner, Goodnow, and Austin, 1956). Models which would adequately simulate that behavior, however, would not necessarily simulate our subjects' behavior adequately; for they would have had no particular reason to include such limitations on the model's capabilities. (Model builders--even if their stated goal is simulation--are human enough to want their programs to do well. Nobody wants his child to be the stupid one in the class, and nobody, to my knowledge, has built a simulator which behaved like the subjects who could not solve the simple problems.) Thus, we get simulations which, for this class of problems, are too good.

G. Combination Models

As a matter of fact, it may be unreasonable to assume that the behavior displayed in these experiments is generated by the same algorithm used in simpler situations. From our subjects' verbal reports and early behavior on the initial experiment, we might conjecture that humans have a large repertoire of problem solving algorithms--or parts of algorithms--each capable of adjusting in some way for the discrepancy between the abstract requirements of the problem and the real limitations of the problem solver. The early behavior in a new situation, then, can be interpreted as a series of tests of higher level hypotheses about the type of environment being confronted.

Viewed in this light, it is easy to understand why each type of model and each specific model has some ring of truth and some data for which it cannot account. We can refrain from retelling the fable of the blind men and the elephant, but its moral is clearly applicable. If we sufficiently restrict the scope of our investigations, only certain facets of a complex phenomenon will be evident to us. Our models will be simpler, to be sure, but we will be at a loss to explain how they are transformed views of the same thing. Only by trying to see the whole elephant at once can we hope to understand that the tail goes on the back and the trunk on the front--all supported by the four stout legs.

CHAPTER VII

FUTURE EXPERIMENTS

In a piece of work which intends to be exploratory, the appropriate way to conclude is by indicating the new paths it has opened, rather than the gates it has shut. When the explorations have been made possible by the creation of a new tool or technique, these new paths will take the form of improvements to the technique and new experiments which will utilize the power the technique has made available. In this concluding chapter, then, we shall discuss both of these alternatives.

A. Improvements in the Apparatus

Although the apparatus as it exists is usable for most of the experiments we shall outline, certain modifications would bring the physical mechanism into better agreement with our stated objectives. The most desirable changes would be to the response mechanism. Without doubt, reliability could be improved by building some type of push button mechanism, physically distinct from the CRT screen. For eight responses, perhaps the ideal solution would be a button for the thumb and first three fingers of each hand. In this way, each response would be a fixed distance from the "rest" position--making response times more consistent and easier to interpret.

For other experiments, we might wish to have more than eight responses available, not necessarily all of the same type. Even more interesting would be the addition of new response measures. For instance, a measurement of pressure on the response button could give us an indication of confidence which was independent of response time. Another measure which might prove useful is the subject's pupil size--a measure which has recently come into use as an indication of heightened attention. At the same time pupil size was being measured, we could also measure where the subject was looking at the screen as he performed various stages of his solution. Finally, certain physiological measures, such as, EEG, EKG, and pulse rate might enable us to further refine our discrimination of the changes in the subject's behavior as he progressed. It might prove difficult, however, to implement such measures without unduly disturbing the subject.

One other facet which might be improved is the retention time of the CRT screen. A sharper discrimination between one stimulus and the next would seem desirable if we are to refine our timing measures.

B. Improvements in the Program

The major area in which the on-line programs could be improved is in the presentation of the test trials. For instance, spreading the test trials out more evenly among the other trials might make them less disturbing to the learning process. Furthermore, if we can make use of our knowledge of the processes involved, we can use a running analysis of performance to tailor the test trials more closely to the information we require. The net result of these modifications would be to make the test trials less onerous to the subjects, less interfering with their performance--and thus more consistent and reliable measuring tools--, and more specifically informative on the currently critical questions.

Another modification to the test trials might be of some interest. As the programs now stand, the subject can always be sure when he is in a test trials because of the missing character. Another kind of test trial has been used in more traditional experiments in which the subject might not be so aware that he is being tested. In these tests, entirely new stimuli are presented which--for instance--match one stimulus class in one part and another stimulus class in another. The way in which the subject classifies these stimuli (which, of course, must not be reinforced) can reveal his current decision structure. One disadvantage to this method is that it is not as susceptible to automatic analysis as is our deletion method. Perhaps both methods could be used together in the same experiments, in order to gain from the advantages of each and to verify our reconstructions from independent sources.

Another area where our new knowledge enables us to envision improvements is in the subroutine which tests for the terminating criterion. Actually, a number of different criteria would probably be needed, depending on the purposes of the experiments, but in general they would be more strict than the one used for this series. When the number of trials to termination is used as an important measure, the terminating criterion must be defined with extreme caution. On the other hand, our methods of analysis have shown how unreliable and uninformative such measures are. Thus, we can afford to err on the side of caution and let our experiments run too long, if necessary, in order to be sure of not terminating before we have seen all behavior of interest.

Generally, in moving to a wider class of experiments--each designed to test more specific hypotheses--the various subroutines in the on-line programs will take over much of the analysis now being done off-line. Taking this direction enables us to do more subtle and precise experiments, but at the same time exposes us to the danger of establishing a filter through which really new and unusual data cannot pass. Thus, there will always remain the need for careful perusal of the data and development of completely new analysis techniques.

For the most part, our experiments are what Bruner, Goodnow, and Austin (1956) call "spectator" experiments, in that the subject has no control over the sequence of stimuli he sees--although we have seen how the subject can exert a certain amount of control over what he seems to see by using a suitable response bias. Also, our experiments are "in-the-head" experiments, rather than "on-the-board" experiments, since the subjects see only one stimulus at a time and are allowed to take no notes. Relaxing each of these constraints opens up a whole new area of experimentation, and most of the experiments we will outline can be done in four different ways: "spectator-in-the-head," "spectator-on-the-board," "participant-in-the-head," and "participant-on-the-board." With suitable program modifications, each of these ways can be realized. In fact, many different degrees of participation and memory aid can be built into the program, thus vastly increasing the spectrum of possible experiments--and also vastly compounding the problems of analysis.

C. New Experiments

1. Experiments to Discover Common *A Priori* Concepts

We have seen the dominant role in concept learning played by concepts which the subjects bring to the experimental situation. Our apparatus gives us the opportunity to explore these concepts systematically. Such explorations can be focused either on the existence of specific concepts or on the differences in concepts held by different groups of subjects.

Among the concepts which can be measured are the following:

a. The concept that there is a unique, "right" solution.

b. The concept that there is a physical compatability between stimulus and response.

c. The concept of continuity in the sequence of stimuli.

d. The concept of continuity in the rules over time.

One way in which we could measure the strength and prevalence of such concepts is to design a series of experiments, each of which matched a given set of concepts in its construction. By ordering these experiments according to increasing difficulty, we could rank the underlying concepts as to their prevalence and the strength with which they are held. In other words, we would be designing a set of experiments, each of which would be the "optimal" experiment for a given set of assumptions. The experiment that best matched the most common and strongly held set of assumptions would be easiest; that which violated those assumptions would be most difficult.

Suppose, for example, that we set up an experiment in which the only indication of the correct response was a small, bright t in the "pointing position" for that response. Suppose further that this t was imbedded in a field of bright and dim, large and small letters from the remainder of the set, varied at random to produce 128 different patterns--eight for each response. In such an experiment, the concepts of physical compatability and of "key" letters are built in, whereas in our current experiments, their presence is only accidental and partial. Our findings would lead us to predict that these experiments would be far easier than our have been.

If such a technique of ranking concepts is successful, it could be applied to different groups of subjects. For example, Piaget (1957) has presented us with a picture of concepts characteristically developing at certain ages. Some experiments (Taylor and McNemar, 1955; Berry, 1959) suggest a possible relation between the subject's sex and the concepts used. Of particular theoretical importance is the claim of certain anthropologists (Lee, 1950) and linguists (Whorf, 1956) that different cultures and languages produce fundamental differences in thought processes. The ability of our techniques to get behind the verbalizations would be especially useful in resolving these arguments.

2. Experiments to Determine the Effects of _A Priori_ Concepts

Our explorations have shown how _a priori_ concepts can influence the subject's choice of strategy. Further experiments could be designed to test this effect more directly.

The basic model for this type of experiment is the experiment in which different groups of subjects are given different instructions. However, merely instructing the subject to look at the experiment in a certain way is not generally a dependable way to alter concepts which may have been building up over a lifetime. Thus, rather than using verbal instructions, the subjects could be preconditioned by a series of experiments in which the desired concept was prominently exhibited.

An example of the kind of question which such experiments could explore is that of preference for certain forms of concepts. Bruner, Goodnow, and Austin (1956) are perhaps the clearest in their statements about the preference of human subjects for conjunctive concepts over disjunctive ones, but their experiments may be designed to encourage the acquisition of one kind over the other. We could precondition two groups of subjects with simple, unambiguous concept learning tasks--one group with conjunctive tasks and one with disjunctive. Then we could give both groups the same ambiguous problems and measure the effects of the conditioning.

Of course, any concept can be preconditioned in this way. Using our apparatus, we have the advantage that the subject need not be aware of the transition from one type of experiment to another. Also, instead of measuring the amount of effect of the preconditioning, we could measure the amount of preconditioning needed to give a certain effect, thus arriving at yet another measure of the strength of _a priori_ concepts.

3. Experiments to Relate Verbal Report to Operational Performance

We have been able to demonstrate by a number of examples how performance and verbal report of performance are related. We could continue to build up such evidence by further replications of these same experiments, but we could also focus on this single question with specially tailored experiments.

We are particularly interested in relating certain well-defined events (such as, the rising response time period just before capture or the declining period just after) to certain kinds of verbalization. Inasmuch as these events can be recognized by the computer, we could build an experiment in which the presentation of stimuli was stopped when the event in question was begun. The computer would then display an appropriate question (or a prerecorded question could be read) and the subject's remarks about what he was currently doing would be recorded for later analysis. Naturally, we would like to design the interruption in such a way as to cause minimum interference with the subject's behavior; but even in the worst situation, a new experiment could begin after each interrogation, since we have enough knowledge to predict the meaning of certain behaviors without watching them develop to their natural conclusion. Any systematic interference (or enhancement) caused by such interruptions would be an interesting finding in itself.

4. Experiments to Study the Development of Verbal Behavior

The response to a stimulus class plays the same role as a name in verbal behavior. Indeed, in many concept learning experiments the responses are names, either familiar words of "nonsense" words. This intimate relationship between concept learning and certain aspects of verbal learning enable any concept learning experiments to be used to study the development of verbal behavior.

One of the most intriguing aspects of naming behavior is that the same object can have more than one name. We can see from our experiments that this behavior can arise because the subject responds to one feature of a stimulus when concentrating on one sub-problem and to another at some other time. One of the limitations of our experiments is that only one "right" name exists for each object. One possible class of experiments could be generated in which a number of different responses were available to be associated with each stimulus class. With such an apparatus, we could study "set" phenomena, development and use of class names versus specific names, production of new names for previously unseen stimuli, and a host of other linguistic phenomena.

5. Experiments to Measure Problem Solving Parameters

There may be certain basic problem abilities which a subject brings to an experiment which affect his performance in various ways. Thus, for example, a subject who is able to remember a few more bits of information may be able to make successful use of a strategy unavailable to another subject; or a subject who can make individual decisions somewhat more quickly may be able effectively to utilize a decision structure with longer decision chains. Other basic abilities might include the ability to estimate the present performance level or the random expected performance level. We have seen how these abilities can affect the subject's performance by either giving or filtering information that he should change his behavior.

If any of these abilities are really basic, we should be able to establish their existence and to measure them for each subject. As an example of the power we have to do this, consider the question of an elemental decision time. By proper choice of stimulus classes and by adjusting the sequence of stimulus presentations, we should be able to induce decision structures of various forms. After the form of the decision structure is established, we can allow the subject to have sufficient practice--carefully balanced for each response--so that his response times approach a steady value. Then, sufficient data can be taken to measure the differences in response times for the different branches and to relate them to the different decisions encountered in arriving at each one.

6. Experiments to Measure the Effects of Experimental Parameters

Much of the difficulty in creating adequate theories of concept learning and problem solving arises from seemingly inconsequential differences in experimental arrangements. One set of experiments presents the stimulus for four seconds, another for six; one set allows a constant time after each response for the next stimulus, another spaces the beginnings of each stimulus a constant difference apart. Variations such as these are constantly being used to explain (after the fact) differences in results. If such objective differences are not available, theorists often call upon more subjective measures, such as motivation, to account for performance differences.

Eventually, someone will have to perform a controlled series of experiments which will enable such parameters to be isolated and studied one at a time. As long as only crude statistical measures (such as total errors to solution) were available, the effects of variations in the experimental apparatus were not adequately measurable. Furthermore, an entirely new apparatus might have to be built in order to make the required variation. With our apparatus, on the other hand, we can characterize performance with the necessary detail, and we can vary experimental parameters by merely changing subroutines or program constants.

For example, to change the absolute or relative times of stimulus presentation or reinforcement, one or two words in the computer memory need be changed. This change can be made in advance or dynamically throughout the program's execution. In fact, the conditions could differ for the different responses, thus enabling us to test the several variations simultaneously with the same subject. Suppose, for instance, that we ran a number of subjects on our present system, but used a different, but fixed, reinforcement time for each stimulus class. We could then relate average learning to the different responses to the reinforcement time in such a way as to eliminate once and for all the charges of variation in conditions--since we would be effectively running eight simultaneous experiments under the same conditions.

In a similar manner, we could manipulate the strength of reinforcement by setting up a payoff function for each stimulus class. Or, again, we could study the effects of the parameter, h, the mean number of trials necessary to set up and test a particular dimension, or any other experimental measure we care to define.

7. Experiments to Measure Learning After Mastery

Each of our experiments terminated when (or before) all of the responses were mastered. We noted that certain predicted recodings of stimulus information did not occur during this period. Furthermore, we observed no substantial simplification of the decision structure, even where many obvious ones were possible. Very likely, the experiments did not go on long enough for these effects to take place, particularly as the subject was devoting himself to an immediate learning task during the entire time. Consequently, it would be of no small interest to perform another series of exploratory experiments in which the subject had to continue performing long after he had reached 100 percent. Perhaps in this extended period following "mastery" we would see recoding, restructuring of the decision, "spreading" of the basis for the decision to make performance impervious to tests, and further development of the verbal report.

APPENDIX

CONSTRAINT TABLES AND STIMULUS PATTERNS FOR EACH EXPERIMENT

The first three pages of this appendix give the constraint tables for each of the experiments described in the text. These tables were used to generate the stimulus patterns which are printed on the following 28 pages. In those listings, the 16 patterns for each stimulus class (000, 001, 010, etc.) are listed in one column, four to a page. Each row represents a given pattern of noise bits (0000, 0001, 0010, etc.), so there are 16 rows in all describing each experiment. Because the computer which printed these lists could not print upper and lower cases, small letters are indicated by an asterisk (*) printed after them. Parentheses around a letter indicate that it would appear as a bright letter on the CRT screen.

EXPERIMENT 1 17, 9, 6, 4, 2, 1, 1

STIMULUS PATTERNS FOR

	1	2	3	4	5	6	7	8
INTENS	A	NA	B	NB	A	NA	3	B
CASE	NB	A	1	NA	B	NC	NA	A
LTRA	A	2	C	NA	C	A	B	N1
LTRB	C	NA	A	4	NA	N1	C	B
LTRC	1	NB	NA	NC	N2	NB	NA	A

EXPERIMENT 2 17, 9, 6, 4, 2, 1, 1

STIMULUS PATTERNS FOR

	1	2	3	4	5	6	7	8
INTENS	NB	A	NA	3	B	A	NA	B
CSE	NA	A	NB	A	1	NA	B	NC
LTA	B	N1	A	2	C	NA	C	A
LTB	C	B	C	NA	A	4	NA	N1
LTC	N2	NB	NA	A	1	NB	NA	NC

EXPERIMENT 3 6, 9, 17, 4, 2, 1, 1

STIMULUS PATTERNS FOR

	1	2	3	4	5	6	7	8
INTENS	C	B	C	B	NB	NC	NC	1
CSE	C	C	NC	N1	NA	B	NB	NC
LTA	1	NB	C	NC	NC	N2	A	B
LTB	A	NA	3	NC	NC	C	4	NB
LTC	C	2	B	C	N1	NA	C	A

TIMULUS PATTERNS FOR EXPERIMENT 4 11, 10, 11, 4, 2, 1, 1

	1	2	3	4	5	6	7	8
INTENS	C	NB	A	B	A	A	C	2
CSE	1	NB	A	C	C	NA	B	NA
LTA	NA	NC	NC	3	NB	NC	1	B
LTB	2	1	N2	NA	C	NC	NB	NA
LTC	B	A	NB	C	N4	N1	NC	A

STIMULUS PATTERNS FOR EXPERIMENT 5 2, 10, 20, 4, 2, 1, 1

	1	2	3	4	5	6	7	8
INTENS	C	1	NC	B	C	B	NB	NC
CSE	C	A	NC	1	B	C	NB	NC
LTA	B	C	NC	NB	N1	3	NC	C
LTB	N1	B	C	B	NC	C	N2	C
LTC	NC	NB	4	NA	NC	C	C	2

STIMULUS PATTERNS FOR EXPERIMENT 6 11, 10, 11, 4, 2, 1, 1

	1	2	3	4	5	6	7	8
INTENS	A	B	A	A	C	2	C	NB
CSE	NA	1	NB	A	C	C	NA	B
LTA	3	NC	NB	NC	1	B	NA	NC
LTB	NA	C	NC	NB	NA	B	1	N2
LTC	NB	C	4	N1	NC	A	B	A

STIMULUS PATTERNS FOR EXPERIMENT 7 20, 10, 2, 4, 2, 1, 1

	1	2	3	4	5	6	7	8
INTENS	A	1	NA	B	A	B	NB	NA
CSE	B	A	NB	NA	A	C	NA	1
LTA	4	NC	NA	A	A	2	NA	NB
LTB	NA	NB	N1	3	NA	A	B	A
LTC	A	B	NA	A	N2	A	N1	B

STIMULUS PATTERNS FOR   EXPERIMENT 1   17, 9, 6, 4, 2, 1, 1

```
0000
    000              001            010              011              100              101              110              111
 R  (Z*) M*     V  (Z*) H*     R* (V*)(M*)     V* (V*)(H*)     (D ) M   V*     (K ) M   K*     (D*) R  (V*)     (K*) R  (K*)
(H )     Z*    (C )     T*     H        Z      D        T      (M*)    (M*)     (R*)    (H*)      M*     (M )      R*    (H )
(Z ) M   C*    (Z*) Z   C*     (V ) H  (K*)    (V*) T  (K*)     T   R*  H       T*  V*  H        K   D* (T )      K*  K* (T )

0001
    000              001            010              011              100              101              110              111
 R  (Z*) M*     V  (Z*) H*     R* (V*)(M*)     V* (V*)(H*)     (D ) M   V*     (K ) M   K*     (D*) R  (V*)     (K*) R  (K*)
(T )     Z*    (K )     T*     T        Z      K        T      (Z*)    (M*)     (V*)    (H*)      Z*     (M )      V*    (H )
(Z ) M   C*    (Z*) Z   C*     (V ) H  (K*)    (V*) T  (K*)     T   R*  H       T*  V*  H        K   D* (T )      K*  K* (T )

0010
    000              001            010              011              100              101              110              111
 R  (Z*) M*     V  (Z*) H*     R* (V*)(M*)     V* (V*)(H*)     (D ) M   V*     (K ) M   K*     (D*) R  (V*)     (K*) R  (K*)
(H )     Z*    (C )     T*     H        Z      D        T      (M*)    (M*)     (R*)    (H*)      M*     (M )      R*    (H )
(Z )(M ) C*    (Z*)(Z ) C*     (V )(H )(K*)    (V*)(T )(K*)     T  (R*) H       T* (V*) H        K  (D*)(T )      K* (K*)(T )

0011
    000              001            010              011              100              101              110              111
 R  (Z*) M*     V  (Z*) H*     R* (V*)(M*)     V* (V*)(H*)     (D ) M   V*     (K ) M   K*     (D*) R  (V*)     (K*) R  (K*)
(T )     Z*    (K )     T*     T        Z      K        T      (Z*)    (M*)     (V*)    (H*)      Z*     (M )      V*    (H )
(Z )(M ) C*    (Z*)(Z ) C*     (V )(H )(K*)    (V*)(T )(K*)     T  (R*) H       T* (V*) H        K  (D*)(T )      K* (K*)(T )
```

STIMULUS PATTERNS FOR   EXPERIMENT 1   17, 9, 6, 4, 2, 1, 1

```
0100
   CCC           CO1           C1C             C11             1CC            101            110            111
R  (T*) M*    V  (T*) F*    R* (K*)(M*)    V* (K*)(H*)    (D ) H   V*    (K ) H   K*    (D*) D  (V*)   (K*) D  (K*)
(H )    V*    (C )    K*    H       V      D       K      (M*)    (R*)   (R*)    (D*)   M*      (R )   R*      (D )
(Z ) M  C*    (Z*) Z  C*    (V ) H  (K*)   (V*) T  (K*)   T   R*  H      T*  V*  H      K   D*  (T )   K*  K*  (T )

0101
   CCC           CO1           01C             011             1CC            101            110            111
R  (T*) M*    V  (T*) F*    R* (K*)(M*)    V* (K*)(H*)    (D ) H   V*    (K ) H   K*    (D*) D  (V*)   (K*) D  (K*)
(T )    V*    (K )    K*    T       V      K       K      (Z*)    (R*)   (V*)    (D*)   Z*      (R )   V*      (D )
(Z ) M  C*    (Z*) Z  C*    (V ) H  (K*)   (V*) T  (K*)   T   R*  H      T*  V*  H      K   D*  (T )   K*  K*  (T )

011C
   CCC           CC1           010             011             1CC            101            110            111
R  (T*) M*    V  (T*) F*    R* (K*)(M*)    V* (K*)(H*)    (D ) H   V*    (K ) H   K*    (D*) D  (V*)   (K*) D  (K*)
(H )    V*    (C )    K*    H       V      D       K      (M*)    (R*)   (R*)    (D*)   M*      (R )   R*      (D )
(Z )(M ) C*   (Z*)(Z ) C*   (V )(H )(K*)   (V*)(T )(K*)   T  (R*) H      T* (V*) H      K  (D*)(T )    K* (K*)(T )

0111
   CCC           CO1           01C             C11             1CC            101            110            111
R  (T*) M*    V  (T*) F*    R* (K*)(M*)    V* (K*)(H*)    (D ) H   V*    (K ) H   K*    (D*) D  (V*)   (K*) D  (K*)
(T )    V*    (K )    K*    T       V      K       K      (Z*)    (R*)   (V*)    (D*)   Z*      (R )   V*      (D )
(Z )(M ) C*   (Z*)(Z ) C*   (V )(H )(K*)   (V*)(T )(K*)   T  (R*) H      T* (V*) H      K  (D*)(T )    K* (K*)(T )
```

STIPULUS PATTERNS FCR   EXPERIMENT 1   17, 9, 6, 4, 2, 1, 1.

```
1C00
  CCC           CC1            C1C              C11              1CC           101           110           111
M  (Z*) M      Z  (Z*) F      M* (V*)(M )      Z* (V*)(H )      (H ) M  V     (T ) M  K     (H*) R  (V )  (T*) R  (K )
(H )    Z*     (C )    T*     H        Z       D        T       (M*)  (M*)    (R*)  (H*)    M*     (M )   R*     (H )
(M ) M  R*     (M*) Z  R*     (R ) H  (V*)     (R*) T  (V*)     H  R*  M      H*  V*  M     D   D* (Z )   D*  K* (Z )

1C01
  CCC           C01            010              C11              1C0           101           110           111
M  (Z*) M      Z  (Z*) F      M* (V*)(M )      Z* (V*)(H )      (H ) M  V     (T ) M  K     (H*) R  (V )  (T*) R  (K )
(T )    Z*     (K )    T*     T        Z       K        T       (Z*)  (M*)    (V*)  (H*)    Z*     (M )   V*     (H )
(M ) M  R*     (M*) Z  R*     (R ) H  (V*)     (R*) T  (V*)     H  R*  M      H*  V*  M     D   D* (Z )   D*  K* (Z )

101C
  CCC           CC1            01C              C11              1CC           101           110           111
M  (Z*) M      Z  (Z*) F      M* (V*)(M )      Z* (V*)(F )      (H ) M  V     (T ) M  K     (H*) R  (V )  (T*) R  (K )
(H )    Z*     (C )    T*     H        Z       D        T       (M*)  (M*)    (R*)  (H*)    M*     (M )   R*     (H )
(M )(M ) R*    (M*)(Z ) R*    (R )(H )(V*)     (R*)(T )(V*)     H (R*) M      H* (V*) M     D (D*)(Z )    D* (K*)(Z )

1011
  CC0           C01            01C              011              1CC           101           110           111
M  (Z*) M      Z  (Z*) F      M* (V*)(M )      Z* (V*)(H )      (H ) M  V     (T ) M  K     (H*) R  (V )  (T*) R  (K )
(T )    Z*     (K )    T*     T        Z       K        T       (Z*)  (M*)    (V*)  (H*)    Z*     (M )   V*     (H )
(M )(M ) R*    (M*)(Z ) R*    (R )(H )(V*)     (R*)(T )(V*)     H (R*) M      H* (V*) M     D (D*)(Z )    D* (K*)(Z )
```

STIMULUS PATTERNS FOR EXPERIMENT 1   17, 9, 6, 4, 2, 1, 1

```
 110C
     CCC            C01             01C              011             1CC             101             110             111
M   (T*) P     Z   (T*) H     M*  (K*)(M )     Z*  (K*)(H )     (H ) H    V     (T ) H    K     (H*) D   (V )   (T*) D   (K )
(H )     V*    (C )     K*    H        V       D        K       (P*)     (R*)   (R*)     (D*)   M*       (R )   R*       (D )
(M ) P   R*    (P*) Z   R*    (R ) H   (V*)    (R*) T   (V*)    H    R*  M      H*   V*  M      D    D*  (Z )   D*   K*  (Z )

 1101
     CCC            CC1             C1C              C11             1CC             101             110             111
M   (T*) P     Z   (T*) H     M*  (K*)(M )     Z*  (K*)(H )     (H ) H    V     (T ) H    K     (H*) D   (V )   (T*) D   (K )
(T )     V*    (K )     K*    T        V       K        K       (Z*)     (R*)   (V*)     (D*)   Z*       (R )   V*       (D )
(M ) P   R*    (P*) Z   R*    (R ) H   (V*)    (R*) T   (V*)    H    R*  M      H*   V*  M      D    D*  (Z )   D*   K*  (Z )

 111C
     CCC            CC1             01C              011             1C0             101             110             111
M   (T*) P     Z   (T*) H     M*  (K*)(M )     Z*  (K*)(H )     (H ) H    V     (T ) H    K     (H*) D   (V )   (T*) D   (K )
(H )     V*    (C )     K*    H        V       D        K       (M*)     (R*)   (R*)     (D*)   M*       (R )   R*       (D )
(M )(P ) R*    (P*)(Z ) R*    (R )(H )(V*)     (R*)(T )(V*)     H   (R*) M      H*  (V*) M      D   (D*)(Z )   D*  (K*)(Z )

 1111
     CCC            C01             010              011             1CC             101             110             111
M   (T*) P     Z   (T*) H     M*  (K*)(M )     Z*  (K*)(H )     (H ) H    V     (T ) H    K     (H*) D   (V )   (T*) D   (K )
(T )     V*    (K )     K*    T        V       K        K       (Z*)     (R*)   (V*)     (D*)   Z*       (R )   V*       (D )
(M )(P ) R*    (P*)(Z ) R*    (R )(H )(V*)     (R*)(T )(V*)     H   (R*) M      H*  (V*) M      D   (D*)(Z )   D*  (K*)(Z )
```

```
STIPULUS PATTERNS FOR          EXPERIMENT 2     17, 9, 6, 4, 2, 1, 1

CCCC
    CCC           C01          010          011          1CC           101          110          111
(M ) H* (P )   (Z ) F* (Z )   H  K* (M*)   T  K* (Z*)   (M*)(H ) D   (Z*)(H ) K   H* (K ) D*   T* (K ) K*
V*       R*    V*       C*    V*    (R*)   V*    (C*)   P        V*  P        K*  M     (V*)   M     (K*)
H  (Z*) Z      F  (T*) V*     D  (Z )(Z )  D  (T )(V*)  (M*) R*  T   (M*) D*  K*  (R*) R (T )  (R*) D (K*)

C001
    CCC           CC1          C1C          C11          1CC           101          110          111
(M ) F* (P )   (Z ) F* (Z )   H  K* (M*)   T  K* (Z*)   (M*)(H ) D   (Z*)(H ) K   H* (K ) D*   T* (K ) K*
V*       R*    V*       C*    V*    (R*)   V*    (C*)   P        V*  P        K*  M     (V*)   M     (K*)
T  (Z*) Z      T  (T*) V*     K  (Z )(Z )  K  (T )(V*)  (Z*) R*  T   (Z*) D*  K*  (V*) R (T )  (V*) D (K*)

C01C
    OCC           CC1          01C          C11          1C0           101          110          111
(M ) H* (P )   (Z ) F* (Z )   F  K* (M*)   T  K* (Z*)   (P*)(H ) D   (Z*)(H ) K   H* (K ) D*   T* (K ) K*
(V*)     R*    (V*)     C*    (V*)  (R*)   (V*)  (C*)   (P )     V*  (M )     K*  (M )  (V*)   (M )  (K*)
H  (Z*) Z      F  (T*) V*     D  (Z )(Z )  D  (T )(V*)  (M*) R*  T   (M*) D*  K*  (R*) R (T )  (R*) D (K*)

C011
    CCC           C01          010          011          1C0           101          110          111
(M ) F* (P )   (Z ) F* (Z )   H  K* (M*)   T  K* (Z*)   (M*)(H ) D   (Z*)(H ) K   H* (K ) D*   T* (K ) K*
(V*)     R*    (V*)     C*    (V*)  (R*)   (V*)  (C*)   (M )     V*  (M )     K*  (M )  (V*)   (M )  (K*)
T  (Z*) Z      T  (T*) V*     K  (Z )(Z )  K  (T )(V*)  (Z*) R*  T   (Z*) D*  K*  (V*) R (T )  (V*) D (K*)
```

STIMULUS PATTERNS FOR   EXPERIMENT 2   17, 9, 6, 4, 2, 1, 1

```
0100
   000            001            010            011            100            101            110            111
(R ) F* (M )   (V ) F* (Z )   D  K* (M*)     K  K* (Z*)     (R*)(H ) D     (V*)(H ) K     D* (K ) D*     K* (K ) K*
K*      R*     K*      C*     K*     (R*)    K*     (C*)    H        V*    H        K*    H     (V*)     H     (K*)
H  (Z*) Z      F  (T*) V*     D  (Z )(Z )    D  (T )(V*)    (M*) R*  T     (M*) D*  K*    (R*) R  (T )   (R*) D (K*)

0101
   000            001            010            011            100            101            110            111
(R ) F* (M )   (V ) F* (Z )   D  K* (M*)     K  K* (Z*)     (R*)(H ) D     (V*)(H ) K     D* (K ) D*     K* (K ) K*
K*      R*     K*      C*     K*     (R*)    K*     (D*)    H        V*    H        K*    H     (V*)     H     (K*)
T  (Z*) Z      T  (T*) V*     K  (Z )(Z )    K  (T )(V*)    (Z*) R*  T     (Z*) D*  K*    (V*) R  (T )   (V*) D (K*)

0110
   000            001            010            011            100            101            110            111
(R ) H* (M )   (V ) F* (Z )   D  K* (M*)     K  K* (Z*)     (R*)(H ) D     (V*)(H ) K     D* (K ) D*     K* (K ) K*
(K*)    R*     (K*)    C*     (K*)   (R*)    (K*)   (C*)    (H )     V*    (H )     K*    (H )  (V*)     (H )  (K*)
H  (Z*) Z      F  (T*) V*     D  (Z )(Z )    D  (T )(V*)    (M*) R*  T     (M*) D*  K*    (R*) R  (T )   (R*) D (K*)

0111
   000            001            010            011            100            101            110            111
(R ) F* (M )   (V ) F* (Z )   D  K* (M*)     K  K* (Z*)     (R*)(H ) D     (V*)(H ) K     D* (K ) D*     K* (K ) K*
(K*)    R*     (K*)    C*     (K*)   (R*)    (K*)   (C*)    (H )     V*    (H )     K*    (H )  (V*)     (H )  (K*)
T  (Z*) Z      T  (T*) V*     K  (Z )(Z )    K  (T )(V*)    (Z*) R*  T     (Z*) D*  K*    (V*) R  (T )   (V*) D (K*)
```

STIMULUS PATTERNS FOR EXPERIMENT 2 17, 5, 6, 4, 2, 1, 1.

```
1000
 CCC          CC1          C1C          C11          1CC          101          110          111
(M ) M* (M ) (Z ) M* (Z )  H   V* (M*)  T   V* (Z*) (M*)(M ) D   (Z*)(M ) K   H* (V ) D*   T* (V ) K*
V*       M   V*       H    V*     (M )  V*     (H ) M         Z  M         T  M      (Z )  M      (T )
H   (Z*) M   H   (T*) R*   D  (Z )(M )  D  (T )(R*) (M*) R*   H  (M*) D*   D* (R*) R  (H )  (R*) D  (D*)

1001
 CCC          CC1          C1C          C11          1CC          101          110          111
(M ) M* (M ) (Z ) M* (Z )  H   V* (M*)  T   V* (Z*) (M*)(M ) D   (Z*)(M ) K   H* (V ) D*   T* (V ) K*
V*       M   V*       H    V*     (M )  V*     (H ) M         Z  M         T  M      (Z )  M      (T )
T   (Z*) M   T   (T*) R*   K  (Z )(M )  K  (T )(R*) (Z*) R*   H  (Z*) D*   D* (V*) R  (H )  (V*) D  (D*)

1010
 CCC          CC1          01C          C11          1CC          101          110          111
(M ) M* (M ) (Z ) M* (Z )  H   V* (M*)  T   V* (Z*) (M*)(M ) D   (Z*)(M ) K   H* (V ) D*   T* (V ) K*
(V*)     M   (V*)     H    (V*)   (M )  (V*)   (H ) (M )      Z  (M )      T  (M )   (Z )  (M )   (T )
H   (Z*) M   H   (T*) R*   D  (Z )(M )  D  (T )(R*) (M*) R*   H  (M*) D*   D* (R*) R  (H )  (R*) D  (D*)

1011
 CCC          CC1          01C          011          1CC          101          110          111
(M ) M* (M ) (Z ) M* (Z )  H   V* (M*)  T   V* (Z*) (M*)(M ) D   (Z*)(M ) K   H* (V ) D*   T* (V ) K*
(V*)     M   (V*)     H    (V*)   (M )  (V*)   (H ) (M )      Z  (M )      T  (M )   (Z )  (M )   (T )
T   (Z*) M   T   (T*) R*   K  (Z )(M )  K  (T )(R*) (Z*) R*   H  (Z*) D*   D* (V*) R  (H )  (V*) D  (D*)
```

STIMULUS PATTERNS FOR   EXPERIMENT 2   17, 9, 6, 4, 2, 1, 1

```
 110C
    CCC                CC1            01C              C11            1C0              101              110              111
(R ) M* (M )    (V ) M* (Z )    D    V* (M*)    K    V* (Z*)    (R*)(M ) D    (V*)(M ) K    D* (V ) D*    K* (V ) K*
 K*       M      K*       F     K*      (M )    K*      (H )     H         Z     H         T     H       (Z )    H        (T )
 H   (Z*) M      F  (T*) R*     D   (Z )(M )    D   (T )(R*)    (M*) R*    H    (M*) D*   D*    (R*) R  (H )    (R*) D  (D*)
```

```
 1101
    CCC                CC1            01C              C11            1CC              101              110              111
(R ) M* (M )    (V ) M* (Z )    D    V* (M*)    K    V* (Z*)    (R*)(M ) D    (V*)(M ) K    D* (V ) D*    K* (V ) K*
 K*       M      K*       F     K*      (M )    K*      (H )     H         Z     H         T     H       (Z )    H        (T )
 T   (Z*) M      T  (T*) R*     K   (Z )(M )    K   (T )(R*)    (Z*) R*    H    (Z*) D*   D*    (V*) R  (H )    (V*) D  (D*)
```

```
 111C
    CCC                CC1            01C              011            1C0              101              110              111
(R ) M* (M )    (V ) M* (Z )    D    V* (M*)    K    V* (Z*)    (R*)(M ) D    (V*)(M ) K    D* (V ) D*    K* (V ) K*
(K*)      M     (K*)      F     (K*)    (M )    (K*)    (H )    (H )       Z    (H )      T    (H )    (Z )   (H )     (T )
 H   (Z*) M      F  (T*) R*     D   (Z )(M )    D   (T )(R*)    (M*) R*    H    (M*) D*   D*    (R*) R  (H )    (R*) D  (D*)
```

```
 1111
    CCC                CC1            C1C              C11            1CC              101              110              111
(R ) M* (M )    (V ) M* (Z )    D    V* (M*)    K    V* (Z*)    (R*)(M ) D    (V*)(M ) K    D* (V ) D*    K* (V ) K*
(K*)      M     (K*)      H     (K*)    (M )    (K*)    (H )    (H )       Z    (H )      T    (H )    (Z )   (H )     (T )
 T   (Z*) M      T  (T*) R*     K   (Z )(M )    K   (T )(R*)    (Z*) R*    H    (Z*) D*   D*    (V*) R  (H )    (V*) D  (D*)
```

STIPULUS PATTERNS FCR   EXPERIMENT 3   6, 9, 17, 4, 2, 1, 1

```
CCOC
   CCC           CC1             C1C              C11            1CC          101           110           111
R*  K*  R      (P ) K  (C*)    R* (V*) M     (M )(V )(H*)    V*  D*  R     (Z ) D  (D*)   V* (R*) M     (Z )(R )(H*)
K       (T )    P      (P )   (K )     T     (M )      P     K      (T*)    P      (M*)   (K )     T*    (M )      M*
(H*)(R ) V      T*  P   V*    (H )(R*) D      T    M*  C*    (D*)(D ) Z     K*  H   Z*    (D )(D*) H      K   H*   H*

CC01
   CCC           CC1             C1C              C11            1CC          101           110           111
R*  K*  R      (P ) K  (C*)    R* (V*) M     (M )(V )(H*)    V*  C*  R     (Z ) D  (D*)   V* (R*) M     (Z )(R )(H*)
K       (T )    P      (P )   (K )     T     (M )      P     K      (T*)    M      (M*)   (K )     T*    (M )      M*
(H*)(V ) V      T*  Z   V*    (H )(V*) D      T    Z*  C*    (D*)(K ) Z     K*  T   Z*    (D )(K*) H      K   T*   H*

C01C
   CCC           CC1             01C              C11            1CC          101           110           111
R*  K*  V      (P ) K  (K*)    R* (V*) Z     (M )(V )(T*)    V*  D*  V     (Z ) D  (K*)   V* (R*) Z     (Z )(R )(T*)
K       (T )    P      (P )   (K )     T     (M )      P     K      (T*)    P      (M*)   (K )     T*    (M )      M*
(H*)(R ) V      T*  P   V*    (H )(R*) D      T    M*  C*    (C*)(C ) Z     K*  H   Z*    (D )(D*) H      K   H*   H*

C011
   CCC           CC1             C1C              C11            1CC          101           110           111
R*  K*  V      (P ) K  (K*)    R* (V*) Z     (M )(V )(T*)    V*  C*  V     (Z ) D  (K*)   V* (R*) Z     (Z )(R )(T*)
K       (T )    P      (P )   (K )     T     (M )      P     K      (T*)    M      (M*)   (K )     T*    (M )      M*
(H*)(V ) V      T*  Z   V*    (H )(V*) D      T    Z*  C*    (D*)(K ) Z     K*  T   Z*    (D )(K*) H      K   T*   H*
```

STIMULUS PATTERNS FOR   EXPERIMENT 3   6, 9, 17, 4, 2, 1, 1

```
0100
   000          001           010           011           100           101           110           111
R*  T*  R    (M ) T  (C*)   R* (Z*) M    (M )(Z )(H*)   V*  H*  R    (Z ) H  (D*)   V* (M*) M    (Z )(M )(H*)
K    (T )    M       (M )   (K )    T    (M )     M     K     (T*)   M      (M*)   (K )    T*   (M )    M*
(M*)(R ) V   Z*  M   V*     (M )(R*) D    Z   M*  C*   (R*)(D ) Z    V*  H   Z*    (R )(D*) H    V   H*  H*

0101
   000          001           010           011           100           101           110           111
R*  T*  R    (M ) T  (C*)   R* (Z*) M    (M )(Z )(H*)   V*  H*  R    (Z ) H  (D*)   V* (M*) M    (Z )(M )(H*)
K    (T )    M       (M )   (K )    T    (M )     M     K     (T*)   M      (M*)   (K )    T*   (M )    M*
(M*)(V ) V   Z*  Z   V*     (M )(V*) D    Z   Z*  C*   (R*)(K ) Z    V*  T   Z*    (R )(K*) H    V   T*  H*

0110
   000          001           010           011           100           101           110           111
R*  T*  V    (M ) T  (K*)   R* (Z*) Z    (M )(Z )(T*)   V*  H*  V    (Z ) H  (K*)   V* (M*) Z    (Z )(M )(T*)
K    (T )    M       (M )   (K )    T    (M )     M     K     (T*)   M      (M*)   (K )    T*   (M )    M*
(M*)(R ) V   Z*  M   V*     (M )(R*) D    Z   M*  C*   (R*)(D ) Z    V*  H   Z*    (R )(D*) H    V   H*  H*

0111
   000          001           010           011           100           101           110           111
R*  T*  V    (M ) T  (K*)   R* (Z*) Z    (M )(Z )(T*)   V*  H*  V    (Z ) H  (K*)   V* (M*) Z    (Z )(M )(T*)
K    (T )    M       (M )   (K )    T    (M )     M     K     (T*)   M      (M*)   (K )    T*   (M )    M*
(M*)(V ) V   Z*  Z   V*     (M )(V*) D    Z   Z*  D*   (R*)(K ) Z    V*  T   Z*    (R )(K*) H    V   T*  H*
```

STIMULUS PATTERNS FOR EXPERIMENT 3 6, 9, 17, 4, 2, 1, 1

```
1000
  000         001          010           011           100          101          110           111
D*  K*  R    (H ) K  (D*)  D* (V*) M    (H )(V )(H*)  K*  D*  R    (T ) D  (D*)  K* (R*) M     (T )(R )(H*)
K*     (K )   M*     (R )  (K*)     K   (M*)     R    K*     (K*)  M*      (R*)  (K*)     K*   (M*)      R*
(H*)(R )(V )  T*  M  (V*)  (H )(R*)(D )  T   M* (D*)  (D*)(D )(Z )  K*  H  (Z*)  (D )(D*)(H )   K    H* (H*)

1001
  000         001          010           011           100          101          110           111
D*  K*  R    (H ) K  (D*)  D* (V*) M    (H )(V )(H*)  K*  D*  R    (T ) D  (D*)  K* (R*) M     (T )(R )(H*)
K*     (K )   M*     (R )  (K*)     K   (M*)     R    K*     (K*)  M*      (R*)  (K*)     K*   (M*)      R*
(H*)(V )(V )  T*  Z  (V*)  (H )(V*)(D )  T   Z* (D*)  (D*)(K )(Z )  K*  T  (Z*)  (D )(K*)(H )   K    T* (H*)

1010
  000         001          010           011           100          101          110           111
D*  K*  V    (H ) K  (K*)  D* (V*) Z    (H )(V )(T*)  K*  D*  V    (T ) D  (K*)  K* (R*) Z     (T )(R )(T*)
K*     (K )   M*     (R )  (K*)     K   (M*)     R    K*     (K*)  M*      (R*)  (K*)     K*   (M*)      R*
(H*)(R )(V )  T*  M  (V*)  (H )(R*)(D )  T   M* (D*)  (D*)(D )(Z )  K*  H  (Z*)  (D )(D*)(H )   K    H* (H*)

1011
  000         001          010           011           100          101          110           111
D*  K*  V    (H ) K  (K*)  D* (V*) Z    (H )(V )(T*)  K*  D*  V    (T ) D  (K*)  K* (R*) Z     (T )(R )(T*)
K*     (K )   M*     (R )  (K*)     K   (M*)     R    K*     (K*)  M*      (R*)  (K*)     K*   (M*)      R*
(H*)(V )(V )  T*  Z  (V*)  (H )(V*)(D )  T   Z* (D*)  (D*)(K )(Z )  K*  T  (Z*)  (D )(K*)(H )   K    T* (H*)
```

STIMULUS PATTERNS FOR EXPERIMENT 3 6, 9, 17, 4, 2, 1, 1

```
1100
   CCO          CO1           010           C11           1CO           101           110           111
D* T*  R     (H ) T  (C*)   D* (Z*) M    (H )(Z )(H*)   K*  H*  R     (T ) H  (D*)   K* (M*) M     (T )(M )(H*)
K*    (K )   M*      (R )   (K*)    K    (M*)     R     K*     (K*)   M*      (R*)   (K*)    K*    (M*)     R*
(M*)(R )(V ) Z*  M   (V*)   (M )(R*)(D ) Z     M* (C*)  (R*)(D )(Z )  V*  H   (Z*)   (R )(D*)(H )  V    H* (H*)

1101
   CCC          CC1           C1C           C11           1CC           101           110           111
D* T*  R     (H ) T  (C*)   D* (Z*) M    (H )(Z )(H*)   K*  H*  R     (T ) H  (D*)   K* (M*) M     (T )(M )(H*)
K*    (K )   M*      (R )   (K*)    K    (M*)     R     K*     (K*)   M*      (R*)   (K*)    K*    (M*)     R*
(M*)(V )(V ) Z*  Z   (V*)   (M )(V*)(D ) Z     Z* (C*)  (R*)(K )(Z )  V*  T   (Z*)   (R )(K*)(H )  V    T* (H*)

1110
   CCC          CO1           010           C11           1CC           101           110           111
D* T*  V     (H ) T  (K*)   C* (Z*) Z    (H )(Z )(T*)   K*  H*  V     (T ) H  (K*)   K* (M*) Z     (T )(M )(T*)
K*    (K )   M*      (R )   (K*)    K    (M*)     R     K*     (K*)   M*      (R*)   (K*)    K*    (M*)     R*
(M*)(R )(V ) Z*  M   (V*)   (M )(R*)(D ) Z     M* (D*)  (R*)(D )(Z )  V*  H   (Z*)   (R )(D*)(H )  V    H* (H*)

1111
   CCC          CO1           C1C           011           1CC           101           110           111
D* T*  V     (H ) T  (K*)   D* (Z*) Z    (H )(Z )(T*)   K*  H*  V     (T ) H  (K*)   K* (M*) Z     (T )(M )(T*)
K*    (K )   M*      (R )   (K*)    K    (M*)     R     K*     (K*)   M*      (R*)   (K*)    K*    (M*)     R*
(M*)(V )(V ) Z*  Z   (V*)   (M )(V*)(D ) Z     Z* (D*)  (R*)(K )(Z )  V*  T   (Z*)   (R )(K*)(H )  V    T* (H*)
```

```
STIPLLLS PATTERNS FCR        EXPERIPENT 4     11, 1C, 11, 4, 2, 1, 1

CCCC
   CCC             CC1            01C            011            1CC             101           110           111
D* (C ) T*    (C*)(R ) Z*     T*  D*  K*    (T*) R*  V*     R* (H )(T )    (R*)(M )(Z )   Z*  H* (K )   (Z*) M* (V )
V*      H*     Z        T    (V*)     M*    (Z )      Z     R*     (H*)    M        (T )  (R*)     (M*)  (M )    (Z )
T   Z*  V      M  (V*) V      T   M   K      M  (R ) K     (T*) Z*  M*    (M*)(V*) M*    (T*) M    H*   (M*)(R ) H*

CC01
   CCC             CC1            C1C            C11            1CC             101           110           111
D* (C ) T*    (C*)(R ) Z*     T*  D*  K*    (T*) R*  V*     R* (H )(T )    (R*)(M )(Z )   Z*  H* (K )   (Z*) M* (V )
V*      D*     Z        K    (V*)     R*    (Z )      V     R*     (D*)    M        (K )  (R*)     (R*)  (M )    (V )
T   Z*  V      M  (V*) V      T   M   K      M  (R ) K     (T*) Z*  M*    (M*)(V*) M*    (T*) M    H*   (M*)(R ) H*

C01C
   CCC             CC1            01C            011            1CC             101           110           111
D* (C ) T*    (C*)(R ) Z*     T*  D*  K*    (T*) R*  V*     R* (H )(T )    (R*)(M )(Z )   Z*  H* (K )   (Z*) M* (V )
K*      H*     T        T    (K*)     M*    (T )      Z     D*     (H*)    H        (T )  (D*)     (M*)  (H )    (Z )
T   Z*  V      M  (V*) V      T   M   K      M  (R ) K     (T*) Z*  M*    (M*)(V*) M*    (T*) M    H*   (M*)(R ) H*

C011
   CCC             CC1            C1C            C11            1CC             101           110           111
D* (C ) T*    (C*)(R ) Z*     T*  D*  K*    (T*) R*  V*     R* (H )(T )    (R*)(M )(Z )   Z*  H* (K )   (Z*) M* (V )
K*      C*     T        K    (K*)     R*    (T )      V     D*     (D*)    H        (K )  (D*)     (R*)  (H )    (V )
T   Z*  V      M  (V*) V      T   M   K      M  (R ) K     (T*) Z*  M*    (M*)(V*) M*    (T*) M    H*   (M*)(R ) H*
```

STIMULUS PATTERNS FOR   EXPERIMENT 4   11, 10, 11, 4, 2, 1, 1

```
 010C
    CCC             CC1              C1C              C11              1C0              101              110              111
D* (C ) H*     (C*)(R ) M*      T*  D*  D*      (T*) R*  R*      R* (H )(H )      (R*)(M )(M )      Z*  H* (D )      (Z*) M* (R )
V*       H*     Z        T      (V*)     M*      (Z )     Z      R*      (H*)      M       (T )     (R*)    (M*)      (M )     (Z )
T   Z* (V )     M   (V*)(V )     T   M  (K )      M  (R )(K )    (T*) Z* (M*)      (M*)(V*)(M*)      (T*) M  (H*)      (M*)(R )(H*)

 0101
    CCC             CC1              C1C              C11              1C0              101              110              111
D* (C ) H*     (C*)(R ) M*      T*  D*  D*      (T*) R*  R*      R* (H )(H )      (R*)(M )(M )      Z*  H* (D )      (Z*) M* (R )
V*       C*     Z        K      (V*)     R*      (Z )     V      R*      (D*)      M       (K )     (R*)    (R*)      (M )     (V )
T   Z* (V )     M   (V*)(V )     T   M  (K )      M  (R )(K )    (T*) Z* (M*)      (M*)(V*)(M*)      (T*) M  (H*)      (M*)(R )(H*)

 0110
    CCC             CC1              010              C11              1C0              101              110              111
D* (C ) H*     (C*)(R ) M*      T*  D*  D*      (T*) R*  R*      R* (H )(H )      (R*)(M )(M )      Z*  H* (D )      (Z*) M* (R )
K*       H*     T        T      (K*)     M*      (T )     Z      D*      (H*)      H       (T )     (D*)    (M*)      (H )     (Z )
T   Z* (V )     M   (V*)(V )     T   M  (K )      M  (R )(K )    (T*) Z* (M*)      (M*)(V*)(M*)      (T*) M  (H*)      (M*)(R )(H*)

 0111
    0C0             CC1              C1C              C11              1C0              101              110              111
D* (C ) H*     (C*)(R ) M*      T*  D*  D*      (T*) R*  R*      R* (H )(H )      (R*)(M )(M )      Z*  H* (D )      (Z*) M* (R )
K*       D*     T        K      (K*)     R*      (T )     V      D*      (D*)      H       (K )     (D*)    (R*)      (H )     (V )
T   Z* (V )     M   (V*)(V )     T   M  (K )      M  (R )(K )    (T*) Z* (M*)      (M*)(V*)(M*)      (T*) M  (H*)      (M*)(R )(H*)
```

```
STIMULUS PATTERNS FOR         EXPERIMENT 4     11, 1C, 11, 4, 2, 1, 1

 1C00
    CCC              CC1              01C              C11              1CC              101              110              111
 D  (K ) T*     (C )(V ) Z*      T   K*   K*      (T ) V*   V*      R  (T )(T )    (R )(Z )(Z )     Z   T* (K )     (Z ) Z* (V )
 V*      H*      Z        T     (V*)      H*      (Z )      Z       R*       (H*)   H        (T )   (R*)     (H*)    (H )     (Z )
 K   T*  V       R  (K*) V       K   H    K        R  (D ) K       (K*) T*  H*     (R*)(K*) H*     (K*) H    H*     (R*)(D ) H*

 1C01
    CCC              C01              01C              C11              1C0              101              110              111
 D  (K ) T*     (C )(V ) Z*      T   K*   K*      (T ) V*   V*      R  (T )(T )    (R )(Z )(Z )     Z   T* (K )     (Z ) Z* (V )
 V*      C*      Z        K     (V*)      R*      (Z )      V       R*       (C*)   H        (K )   (R*)     (R*)    (H )     (V )
 K   T*  V       R  (K*) V       K   H    K        R  (D ) K       (K*) T*  H*     (R*)(K*) H*     (K*) H    H*     (R*)(D ) H*

 101C
    CCC              CC1              01C              C11              1CC              101              110              111
 D  (K ) T*     (C )(V ) Z*      T   K*   K*      (T ) V*   V*      R  (T )(T )    (R )(Z )(Z )     Z   T* (K )     (Z ) Z* (V )
 K*      H*      T        T     (K*)      H*      (T )      Z       D*       (H*)   H        (T )   (D*)     (H*)    (H )     (Z )
 K   T*  V       R  (K*) V       K   H    K        R  (D ) K       (K*) T*  H*     (R*)(K*) H*     (K*) H    H*     (R*)(D ) H*

 1011
    CCC              CC1              01C              C11              1CC              101              110              111
 D  (K ) T*     (C )(V ) Z*      T   K*   K*      (T ) V*   V*      R  (T )(T )    (R )(Z )(Z )     Z   T* (K )     (Z ) Z* (V )
 K*      C*      T        K     (K*)      R*      (T )      V       D*       (D*)   H        (K )   (D*)     (R*)    (H )     (V )
 K   T*  V       R  (K*) V       K   H    K        R  (D ) K       (K*) T*  H*     (R*)(K*) H*     (K*) H    H*     (R*)(D ) H*
```

STIMULUS PATTERNS FOR EXPERIMENT 4 11, 10, 11, 4, 2, 1, 1

```
1100
   000            001              010             011              100               101             110             111
D  (K ) H*    (D )(V ) M*     T    K*   D*    (T ) V*   R*    R  (T )(H )     (R )(Z )(M )    Z    T* (D )    (Z ) Z* (R )
V*      H*    Z         T     (V*)      M*    (Z )      Z     R*      (H*)    M       (T )    (R*)    (M*)    (M )    (Z )
K   T* (V )   R   (K*)(V )    K    H   (K )   R   (D )(K )    (K*) T* (M*)    (R*)(K*)(M*)    (K*) H  (H*)    (R*)(D )(H*)

1101
   000            001              010             011              100               101             110             111
D  (K ) H*    (D )(V ) M*     T    K*   D*    (T ) V*   R*    R  (T )(H )     (R )(Z )(M )    Z    T* (D )    (Z ) Z* (R )
V*      D*    Z         K     (V*)      R*    (Z )      V     R*      (D*)    H       (K )    (R*)    (R*)    (M )    (V )
K   T* (V )   R   (K*)(V )    K    H   (K )   R   (D )(K )    (K*) T* (M*)    (R*)(K*)(M*)    (K*) H  (H*)    (R*)(D )(H*)

1110
   000            001              010             011              100               101             110             111
D  (K ) H*    (D )(V ) M*     T    K*   D*    (T ) V*   R*    R  (T )(H )     (R )(Z )(M )    Z    T* (D )    (Z ) Z* (R )
K*      H*    T         T     (K*)      M*    (T )      Z     D*      (H*)    H       (T )    (D*)    (M*)    (H )    (Z )
K   T* (V )   R   (K*)(V )    K    H   (K )   R   (D )(K )    (K*) T* (M*)    (R*)(K*)(M*)    (K*) H  (H*)    (R*)(D )(H*)

1111
   000            001              010             011              100               101             110             111
D  (K ) H*    (D )(V ) M*     T    K*   D*    (T ) V*   R*    R  (T )(H )     (R )(Z )(M )    Z    T* (D )    (Z ) Z* (R )
K*      D*    T         K     (K*)      R*    (T )      V     D*      (D*)    H       (K )    (D*)    (R*)    (H )    (V )
K   T* (V )   R   (K*)(V )    K    H   (K )   R   (D )(K )    (K*) T* (M*)    (R*)(K*)(M*)    (K*) H  (H*)    (R*)(D )(H*)
```

STIMULUS PATTERNS FOR   EXPERIMENT 5.   2, 1C, 2C, 4, 2, 1, 1

```
CC00
    CCC               CC1               01C              C11              1CC              101              110              111
Z*  M* (C )     (V ) H*   V*      T*  V* (D )    (K ) K*   V*      Z*  M   (D )    (V ) H    V*     T*  V  (D )     (K ) K    V*
H*       T*     H*      (C*)      (Z*)     T     (Z*)     (D )     D*        T*    C*      (D*)     (V*)     T      (V*)     (D )
R* (K )(R )     Z  (Z ) K*        (R*) K* (R )   (Z ) Z*   K*      R* (K )(R )     Z  (Z ) K*     (R*) K* (R )     (Z ) Z*   K*

CC01
    CCC               CC1               01C              C11              1CC              101              110              111
Z*  M* (H )     (V ) H*   Z*      T*  V* (H )    (K ) K*   Z*      Z*  M   (H )    (V ) H    Z*     T*  V  (H )     (K ) K    Z*
H*       T*     H*      (C*)      (Z*)     T     (Z*)     (C )     D*        T*    D*      (D*)     (V*)     T      (V*)     (D )
R* (K )(R )     Z  (Z ) K*        (R*) K* (R )   (Z ) Z*   K*      R* (K )(R )     Z  (Z ) K*     (R*) K* (R )     (Z ) Z*   K*

C010
    CCC               CC1               C1C              C11              1CC              101              110              111
Z*  M* (C )     (V ) H*   V*      T*  V* (D )    (K ) K*   V*      Z*  M   (D )    (V ) H    V*     T*  V  (D )     (K ) K    V*
H*       T*     H*      (C*)      (Z*)     T     (Z*)     (D )     D*        T*    D*      (D*)     (V*)     T      (V*)     (D )
D* (K )(R )     T  (Z ) K*        (C*) K* (R )   (T ) Z*   K*      D* (K )(R )     T  (Z ) K*     (D*) K* (R )     (T ) Z*   K*

C011
    COC               C01               01C              C11              1CC              101              110              111
Z*  M* (H )     (V ) H*   Z*      T*  V* (H )    (K ) K*   Z*      Z*  M   (H )    (V ) H    Z*     T*  V  (H )     (K ) K    Z*
H*       T*     H*      (C*)      (Z*)     T     (Z*)     (C )     D*        T*    C*      (C*)     (V*)     T      (V*)     (D )
D* (K )(R )     T  (Z ) K*        (C*) K* (R )   (T ) Z*   K*      D* (K )(R )     T  (Z ) K*     (D*) K* (R )     (T ) Z*   K*
```

STIMULUS PATTERNS FOR EXPERIMENT 5 2, 10, 20, 4, 2, 1, 1

```
0100
   CCO            C01          C1C          C11           1CO          101          110          111
Z•  M• (C )   (V ) H•  V•    T•  V• (D )  (K ) K•  V•   Z•  M  (D )  (V ) H   V•   T•  V  (D )  (K ) K   V•
H•      T•     H•     (C•)  (Z•)     T    (Z•)    (C )  D•      T•   D•     (C•)  (V•)     T    (V•)    (D )
R• (C )(M )   Z  (M ) T•    (R•) D• (M )  (Z ) M•  T•   R• (D )(M )  Z  (M ) T•   (R•) D• (M )  (Z ) M•  T•

C101
   CCC            CC1          C1C          C11           1CC          101          110          111
Z•  M• (H )   (V ) H•  Z•    T•  V• (H )  (K ) K•  Z•   Z•  M  (H )  (V ) H   Z•   T•  V  (H )  (K ) K   Z•
H•      T•     H•     (C•)  (Z•)     T    (Z•)    (C )  D•      T•   D•     (C•)  (V•)     T    (V•)    (D )
R• (C )(M )   Z  (M ) T•    (R•) D• (M )  (Z ) M•  T•   R• (D )(M )  Z  (M ) T•   (R•) D• (M )  (Z ) M•  T•

C11C
   CCC            C01          01C          011           1CC          101          110          111
Z•  M• (C )   (V ) H•  V•    T•  V• (D )  (K ) K•  V•   Z•  M  (D )  (V ) H   V•   T•  V  (D )  (K ) K   V•
H•      T•     H•     (C•)  (Z•)     T    (Z•)    (C )  D•      T•   C•     (D•)  (V•)     T    (V•)    (D )
D• (C )(M )   T  (M ) T•    (C•) D• (M )  (T ) M•  T•   D• (D )(M )  T  (M ) T•   (C•) D• (M )  (T ) M•  T•

C111
   CCC            CC1          C1C          C11           1CC          101          110          111
Z•  M• (H )   (V ) H•  Z•    T•  V• (H )  (K ) K•  Z•   Z•  M  (H )  (V ) H   Z•   T•  V  (H )  (K ) K   Z•
H•      T•     H•     (C•)  (Z•)     T    (Z•)    (C )  C•      T•   D•     (C•)  (V•)     T    (V•)    (D )
D• (C )(M )   T  (M ) T•    (C•) D• (M )  (T ) M•  T•   D• (D )(M )  T  (M ) T•   (D•) D• (M )  (T ) M•  T•
```

STIMULUS PATTERNS FOR EXPERIMENT 5 2, 10, 20, 4, 2, 1, 1

```
1000
 000           001              010              011              100             101              110              111
M* (M*)(D )    (R )(H*) V*      H* (V*)(D )      (D )(K*) V*      M* (M )(D )     (R )(H ) V*      H* (V )(D )      (D )(K ) V*
H        Z*    H        (R*)    (Z )       Z     (Z )     (R )    D         Z*    D        (R*)    (V )       Z     (V )     (R )
R* (K )(R )    Z  (Z ) K*       (R*) K* (R )     (Z ) Z*  K*      R* (K )(R )     Z  (Z ) K*       (R*) K* (R )     (Z ) Z*  K*

1001
 000           001              010              011              100             101              110              111
M* (M*)(H )    (R )(H*) Z*      H* (V*)(H )      (D )(K*) Z*      M* (M )(H )     (R )(H ) Z*      H* (V )(H )      (D )(K ) Z*
H        Z*    H        (R*)    (Z )       Z     (Z )     (R )    D         Z*    D        (R*)    (V )       Z     (V )     (R )
R* (K )(R )    Z  (Z ) K*       (R*) K* (R )     (Z ) Z*  K*      R* (K )(R )     Z  (Z ) K*       (R*) K* (R )     (Z ) Z*  K*

1010
 000           001              010              011              100             101              110              111
M* (M*)(D )    (R )(H*) V*      H* (V*)(D )      (D )(K*) V*      M* (M )(D )     (R )(H ) V*      H* (V )(D )      (D )(K ) V*
H        Z*    H        (R*)    (Z )       Z     (Z )     (R )    D         Z*    D        (R*)    (V )       Z     (V )     (R )
D* (K )(R )    T  (Z ) K*       (D*) K* (R )     (T ) Z*  K*      D* (K )(R )     T  (Z ) K*       (D*) K* (R )     (T ) Z*  K*

1011
 000           001              010              011              100             101              110              111
M* (M*)(H )    (R )(H*) Z*      H* (V*)(H )      (D )(K*) Z*      M* (M )(H )     (R )(H ) Z*      H* (V )(H )      (D )(K ) Z*
H        Z*    H        (R*)    (Z )       Z     (Z )     (R )    D         Z*    D        (R*)    (V )       Z     (V )     (R )
D* (K )(R )    T  (Z ) K*       (D*) K* (R )     (T ) Z*  K*      D* (K )(R )     T  (Z ) K*       (D*) K* (R )     (T ) Z*  K*
```

```
STIMULUS PATTERNS FOR        EXPERIMENT 5      2, 1C, 2C, 4, 2, 1, 1

 1100
    CCC              CC1              01C              C11              1CC              101              110              111
M* (M*)(D )    (R )(H*) V*      H* (V*)(D )    (D )(K*) V*      M* (M )(D )    (R )(H ) V*      H* (V )(D )    (D )(K ) V*
H        Z*    H        (R*)    (Z )      Z    (Z )     (R )    D         Z*   D        (R*)    (V )      Z    (V )     (R )
R* (D )(M )    Z  (M ) T*       (R*) D* (M )   (Z ) M*   T*     R* (D )(M )    Z  (M ) T*       (R*) D* (M )   (Z ) M*   T*

 1101
    CCC              CC1              01C              C11              1CC              101              110              111
M* (M*)(H )    (R )(H*) Z*      H* (V*)(H )    (D )(K*) Z*      M* (M )(H )    (R )(H ) Z*      H* (V )(H )    (D )(K ) Z*
H        Z*    H        (R*)    (Z )      Z    (Z )     (R )    D         Z*   D        (R*)    (V )      Z    (V )     (R )
R* (D )(M )    Z  (M ) T*       (R*) D* (M )   (Z ) M*   T*     R* (D )(M )    Z  (M ) T*       (R*) D* (M )   (Z ) M*   T*

 1110
    CCC              CC1              01C              C11              1CC              101              110              111
M* (M*)(D )    (R )(H*) V*      H* (V*)(D )    (D )(K*) V*      M* (M )(D )    (R )(H ) V*      H* (V )(D )    (D )(K ) V*
H        Z*    H        (R*)    (Z )      Z    (Z )     (R )    D         Z*   D        (R*)    (V )      Z    (V )     (R )
D* (D )(M )    T  (M ) T*       (D*) D* (M )   (T ) M*   T*     D* (D )(M )    T  (M ) T*       (D*) D* (M )   (T ) M*   T*

 1111
    CCC              CC1              01C              C11              1CC              101              110              111
M* (M*)(H )    (R )(H*) Z*      H* (V*)(H )    (D )(K*) Z*      M* (M )(H )    (R )(H ) Z*      H* (V )(H )    (D )(K ) Z*
H        Z*    H        (R*)    (Z )      Z    (Z )     (R )    D         Z*   D        (R*)    (V )      Z    (V )     (R )
D* (D )(M )    T  (M ) T*       (D*) D* (M )   (T ) M*   T*     D* (D )(M )    T  (M ) T*       (D*) D* (M )   (T ) M*   T*
```

STIMULUS PATTERNS FOR   EXPERIMENT 6   11, 10, 11, 4, 2, 1, 1

```
CCOO
    CCC          CC1          C1C          C11          1CC          101          110          111
Z   C*  K     Z   Z*  C    V  (C*) V*   V  (Z*) R*   (M*) D* (K )  (M*) Z* (D )  (R*)(D*)(V*)  (R*)(Z*)(R*)
T*     Z*     Z*     (V )  H*      Z*   M*     (V )  (T )     M*   (Z )    (R )  (H )     M*   (M )    (R )
R*  C  (K*)   R  (C )(V*)  K*  H   K    K  (H ) V      M*  R* (T*)  M  (R*)(Z*)   T*  M*  T     T  (M*) Z

CCC1
    CCC          CC1          01C          C11          1C0          101          110          111
Z   C*  T     Z   Z*  H    V  (D*) Z*   V  (Z*) M*   (M*) D* (T )  (M*) Z* (H )  (R*)(D*)(Z*)  (R*)(Z*)(M*)
T*     Z*     Z*     (V )  H*      Z*   M*     (V )  (T )     M*   (Z )    (R )  (H )     M*   (M )    (R )
R*  C  (K*)   R  (C )(V*)  K*  H   K    K  (H ) V      M*  R* (T*)  M  (R*)(Z*)   T*  M*  T     T  (M*) Z

C01C
    CCC          CC1          01C          C11          1CC          101          110          111
T   C*  K     T   Z*  C    K  (D*) V*   K  (Z*) R*   (H*) D* (K )  (H*) Z* (D )  (D*)(D*)(V*)  (D*)(Z*)(R*)
T*     Z*     Z*     (V )  H*      Z*   M*     (V )  (T )     M*   (Z )    (R )  (H )     M*   (M )    (R )
R*  C  (K*)   R  (C )(V*)  K*  H   K    K  (H ) V      M*  R* (T*)  M  (R*)(Z*)   T*  M*  T     T  (M*) Z

C011
    CCC          C01          01C          C11          1CC          101          110          111
T   C*  T     T   Z*  H    K  (D*) Z*   K  (Z*) M*   (H*) D* (T )  (H*) Z* (H )  (D*)(D*)(Z*)  (D*)(Z*)(M*)
T*     Z*     Z*     (V )  H*      Z*   M*     (V )  (T )     M*   (Z )    (R )  (H )     M*   (M )    (R )
R*  C  (K*)   K  (C )(V*)  K*  H   K    K  (H ) V      M*  R* (T*)  M  (R*)(Z*)   T*  M*  T     T  (M*) Z
```

```
STIMULCS PATTERNS FCR        EXPERIMENT 6       11, 1C, 11, 4, 2, 1, 1

 0100
     C00          001           010           C11           1CC           101           110           111
 Z   C*  K      Z  Z*  C      V  (D*) V*    V  (Z*) R*    (M*) D* (K )  (M*) Z* (D )  (R*)(D*)(V*)  (R*)(Z*)(R*)
 T*      Z*     Z*     (V )   M*      Z*    M*     (V )   (T )     M*   (Z )    (R )  (H )      M*  (M )    (R )
(R*) C  (D*)   (R )(C )(R*)  (K*) H   D    (K )(H ) R    (M*) R* (H*)  (M )(R*)(M*)  (T*) M*   H   (T )(M*) M

 0101
     CCC          C01           01C           C11           1CC           101           110           111
 Z   C*  T      Z  Z*  H      V  (D*) Z*    V  (Z*) M*    (M*) D* (T )  (M*) Z* (H )  (R*)(D*)(Z*)  (R*)(Z*)(M*)
 T*      Z*     Z*     (V )   H*      Z*    M*     (V )   (T )     M*   (Z )    (R )  (H )      M*  (M )    (R )
(R*) C  (C*)   (R )(C )(R*)  (K*) H   D    (K )(H ) R    (M*) R* (H*)  (M )(R*)(M*)  (T*) M*   H   (T )(M*) M

 0110
     CCC          CC1           01C           C11           1CC           101           110           111
 T   C*  K      T  Z*  C      K  (C*) V*    K  (Z*) R*    (H*) D* (K )  (H*) Z* (D )  (D*)(D*)(V*)  (D*)(Z*)(R*)
 T*      Z*     Z*     (V )   H*      Z*    M*     (V )   (T )     M*   (Z )    (R )  (H )      M*  (M )    (R )
(R*) C  (C*)   (R )(C )(R*)  (K*) H   D    (K )(H ) R    (M*) R* (H*)  (M )(R*)(M*)  (T*) M*   H   (T )(M*) M

 0111
     CC0          CC1           010           011           1C0           101           110           111
 T   C*  T      T  Z*  H      K  (D*) Z*    K  (Z*) M*    (H*) D* (T )  (H*) Z* (H )  (D*)(D*)(Z*)  (D*)(Z*)(M*)
 T*      Z*     Z*     (V )   H*      Z*    M*     (V )   (T )     M*   (Z )    (R )  (H )      M*  (M )    (R )
(R*) C  (C*)   (R )(C )(R*)  (K*) H   D    (K )(H ) R    (M*) R* (H*)  (M )(R*)(M*)  (T*) M*   H   (T )(M*) M
```

```
STIMULUS PATTERNS FOR        EXPERIMENT 6       11, 1C, 11, 4, 2, 1, 1

1CCC
   CCO          CC1          O1O           C11           1CC            101             110             111
Z  C  K      Z  Z  C      V  (D ) V*    V  (Z ) R*    (M*) D  (K )   (M*) Z  (C )    (R*)(D )(V*)    (R*)(Z )(R*)
K*    T*     V*   (K )    C*      T*    R*    (K )    (K )    H*     (V )    (D )    (D )     H*     (R )    (D )
R*  K (K*)   R  (K )(V*)  K*  T   K     K  (T ) V      M*  V* (T*)    M  (V*)(Z*)    T*  Z*  T       T  (Z*) Z

1001
   CCC          CC1          C1C           C11           1CC            101             110             111
Z  C  T      Z  Z  H      V  (D ) Z*    V  (Z ) M*    (M*) D  (T )   (M*) Z  (H )    (R*)(D )(Z*)    (R*)(Z )(M*)
K*    T*     V*   (K )    C*      T*    R*    (K )    (K )    H*     (V )    (C )    (D )     H*     (R )    (D )
R*  K (K*)   R  (K )(V*)  K*  T   K     K  (T ) V      M*  V* (T*)    M  (V*)(Z*)    T*  Z*  T       T  (Z*) Z

1C1C
   CCO          CO1          O1C           O11           1CO            101             110             111
T  C  K      T  Z  C      K  (D ) V*    K  (Z ) R*    (H*) D  (K )   (H*) Z  (D )    (D*)(D )(V*)    (D*)(Z )(R*)
K*    T*     V*   (K )    C*      T*    R*    (K )    (K )    H*     (V )    (D )    (D )     H*     (R )    (D )
R*  K (K*)   R  (K )(V*)  K*  T   K     K  (T ) V      M*  V* (T*)    M  (V*)(Z*)    T*  Z*  T       T  (Z*) Z

1011
   CCC          CC1          C1C           C11           1CC            101             110             111
T  C  T      T  Z  H      K  (D ) Z*    K  (Z ) M*    (H*) D  (T )   (H*) Z  (H )    (D*)(D )(Z*)    (D*)(Z )(M*)
K*    T*     V*   (K )    C*      T*    R*    (K )    (K )    H*     (V )    (D )    (D )     H*     (R )    (D )
R*  K (K*)   R  (K )(V*)  K*  T   K     K  (T ) V      M*  V* (T*)    M  (V*)(Z*)    T*  Z*  T       T  (Z*) Z
```

STIMULUS PATTERNS FOR EXPERIMENT 6 11, 10, 11, 4, 2, 1, 1

```
1100
   000           001          010           011           100           101           110           111
Z  0  K      Z  Z   D     V  (D ) V*    V  (Z ) R*    (M*) D  (K )  (M*) Z  (D )  (R*)(D )(V*)  (R*)(Z )(R*)
K*    T*     V*    (K )   D*      T*    R*     (K )   (K )    H*    (V )    (D )  (D )     H*   (R )    (D )
(R*) K (D*)  (P )(K )(R*) (K*) T   D    (K )(T ) R    (M*) V* (H*)  (M )(V*)(M*)  (T*) Z*  H    (T )(Z*) M

1101
   000           001          010           011           100           101           110           111
Z  0  T      Z  Z   H     V  (D ) Z*    V  (Z ) M*    (M*) D  (T )  (M*) Z  (H )  (R*)(D )(Z*)  (R*)(Z )(M*)
K*    T*     V*    (K )   D*      T*    R*     (K )   (K )    H*    (V )    (D )  (D )     H*   (R )    (D )
(R*) K (D*)  (R )(K )(R*) (K*) T   D    (K )(T ) R    (M*) V* (H*)  (M )(V*)(M*)  (T*) Z*  H    (T )(Z*) M

1110
   000           001          010           011           100           101           110           111
T  0  K      T  Z   D     K  (D ) V*    K  (Z ) R*    (H*) D  (K )  (H*) Z  (D )  (D*)(D )(V*)  (D*)(Z )(R*)
K*    T*     V*    (K )   D*      T*    R*     (K )   (K )    H*    (V )    (D )  (D )     H*   (R )    (D )
(R*) K (D*)  (R )(K )(R*) (K*) T   D    (K )(T ) R    (M*) V* (H*)  (M )(V*)(M*)  (T*) Z*  H    (T )(Z*) M

1111
   000           001          010           011           100           101           110           111
T  0  T      T  Z   H     K  (D ) Z*    K  (Z ) M*    (H*) D  (T )  (H*) Z  (H )  (D*)(D )(Z*)  (D*)(Z )(M*)
K*    T*     V*    (K )   D*      T*    R*     (K )   (K )    H*    (V )    (D )  (D )     H*   (R )    (D )
(R*) K (D*)  (R )(K )(R*) (K*) T   D    (K )(T ) R    (M*) V* (H*)  (M )(V*)(M*)  (T*) Z*  H    (T )(Z*) M
```

STIMULUS PATTERNS FOR EXPERIMENT 7 20, 1C, 2, 4, 2, 1, 1

```
CCOC
   CCC          CC1          C1C          C11          1CC          101          110          111
V*  K* (T )   V*  V* (T )   V   H* (T*)   V   M* (T*)  (M*) K   V   (M*) V   V   (M ) H   V*  (M ) M   V*
R       Z*    R        Z*   (R )     Z*   (R )     Z*   H*     (H )  H*     (H )  (H*)    (H )  (H*)    (H )
R* (F )(D*)   R  (F )(D*)   (R*) T (M*)   (R ) T (M*)   Z* (M*) K*   Z  (M*) K*   (Z*) Z*  Z*   (Z ) Z*  Z*

CC01
   CCC          CC1          01C          C11          1CC          101          110          111
K*  K* (T )   K*  V* (T )   K   H* (T*)   K   M* (T*)  (H*) K   V   (H*) V   V   (H ) H   V*  (H ) M   V*
R       Z*    R        Z*   (R )     Z*   (R )     Z*   H*     (H )  H*     (H )  (H*)    (H )  (H*)    (H )
R* (F )(D*)   R  (F )(D*)   (R*) T (M*)   (R ) T (M*)   Z* (M*) K*   Z  (M*) K*   (Z*) Z*  Z*   (Z ) Z*  Z*

C01C
   CCC          CC1          01C          C11          1C0          101          110          111
V*  K* (T )   V*  V* (T )   V   H* (T*)   V   M* (T*)  (M*) K   V   (M*) V   V   (M ) H   V*  (M ) M   V*
V       Z*    V        Z*   (V )     Z*   (V )     Z*   T*     (H )  T*     (H )  (T*)    (H )  (T*)    (H )
R* (H )(D*)   R  (F )(D*)   (R*) T (M*)   (R ) T (M*)   Z* (M*) K*   Z  (M*) K*   (Z*) Z*  Z*   (Z ) Z*  Z*

C011
   CCC          CC1          01C          C11          1CC          101          110          111
K*  K* (T )   K*  V* (T )   K   H* (T*)   K   M* (T*)  (H*) K   V   (H*) V   V   (H ) H   V*  (H ) M   V*
V       Z*    V        Z*   (V )     Z*   (V )     Z*   T*     (H )  T*     (H )  (T*)    (H )  (T*)    (H )
R* (F )(D*)   R  (F )(D*)   (R*) T (M*)   (R ) T (M*)   Z* (M*) K*   Z  (M*) K*   (Z*) Z*  Z*   (Z ) Z*  Z*
```

STIMULUS PATTERNS FOR EXPERIMENT 7 20, 10, 2, 4, 2, 1, 1.

```
0100
    000            001           010           011          100          101          110          111
V*  K* (T )    V*  V* (T )    V    H* (T*)   V    M* (T*)   (M*) K   V   (M*) V   V   (M ) H  V*   (M ) M   V*
R       V*     R       V*     (R )     V*    (R )     V*    H*     (D )  H*     (D )  (H*)   (D )   (H*)    (D )
D*  (H )(C*)   C   (H )(C*)   (C*) T  (M*)   (D ) T  (M*)   T* (M*) K*   T  (M*) K*   (T*) Z* Z*   (T ) Z*  Z*

0101
    000            001           010           011          100          101          110          111
K*  K* (T )    K*  V* (T )    K    H* (T*)   K    M* (T*)   (H*) K   V   (H*) V   V   (H ) H  V*   (H ) M   V*
R       V*     R       V*     (R )     V*    (R )     V*    H*     (D )  H*     (D )  (H*)   (D )   (H*)    (D )
D*  (H )(C*)   C   (H )(C*)   (C*) T  (M*)   (D ) T  (M*)   T* (M*) K*   T  (M*) K*   (T*) Z* Z*   (T ) Z*  Z*

0110
    000            001           010           011          100          101          110          111
V*  K* (T )    V*  V* (T )    V    H* (T*)   V    M* (T*)   (M*) K   V   (M*) V   V   (M ) H  V*   (M ) M   V*
V       V*     V       V*     (V )     V*    (V )     V*    T*     (D )  T*     (D )  (T*)   (D )   (T*)    (D )
D*  (H )(C*)   C   (H )(C*)   (C*) T  (M*)   (D ) T  (M*)   T* (M*) K*   T  (M*) K*   (T*) Z* Z*   (T ) Z*  Z*

0111
    000            001           010           011          100          101          110          111
K*  K* (T )    K*  V* (T )    K    H* (T*)   K    M* (T*)   (H*) K   V   (H*) V   V   (H ) H  V*   (H ) M   V*
V       V*     V       V*     (V )     V*    (V )     V*    T*     (D )  T*     (D )  (T*)   (D )   (T*)    (D )
D*  (H )(C*)   C   (H )(C*)   (C*) T  (M*)   (D ) T  (M*)   T* (M*) K*   T  (M*) K*   (T*) Z* Z*   (T ) Z*  Z*
```

STIMULUS PATTERNS FOR EXPERIMENT 7 20, 10, 2, 4, 2, 1, 1

```
 1000
    000          001          010          011          100          101          110          111
V* (K*)(H )   V* (V*)(H )   V  (H*)(H*)   V  (M*)(H*)   (M*)(K ) R   (M*)(V ) R   (M )(H ) R*   (M )(M ) R*
R        Z*   R        Z*   (R )     Z*   (R )     Z*   H*     (H )   H*     (H )   (H*)   (H )   (H*)   (H )
R* (D )(D )   R  (D )(D )   (R*) K  (M )  (R ) K  (M )  Z* (R*) K    Z  (R*) K    (Z*) V*  Z    (Z ) V*  Z

 1001
    000          001          010          011          100          101          110          111
K* (K*)(H )   K* (V*)(H )   K  (H*)(H*)   K  (M*)(H*)   (H*)(K ) R   (H*)(V ) R   (H )(H ) R*   (H )(M ) R*
R        Z*   R        Z*   (R )     Z*   (R )     Z*   H*     (H )   H*     (H )   (H*)   (H )   (H*)   (H )
R* (D )(D )   R  (D )(D )   (R*) K  (M )  (R ) K  (M )  Z* (R*) K    Z  (R*) K    (Z*) V*  Z    (Z ) V*  Z

 1010
    000          001          010          011          100          101          110          111
V* (K*)(H )   V* (V*)(H )   V  (H*)(H*)   V  (M*)(H*)   (M*)(K ) R   (M*)(V ) R   (M )(H ) R*   (M )(M ) R*
V        Z*   V        Z*   (V )     Z*   (V )     Z*   T*     (H )   T*     (H )   (T*)   (H )   (T*)   (H )
R* (D )(D )   R  (D )(D )   (R*) K  (M )  (R ) K  (M )  Z* (R*) K    Z  (R*) K    (Z*) V*  Z    (Z ) V*  Z

 1011
    000          001          010          011          100          101          110          111
K* (K*)(H )   K* (V*)(H )   K  (H*)(H*)   K  (M*)(H*)   (H*)(K ) R   (H*)(V ) R   (H )(H ) R*   (H )(M ) R*
V        Z*   V        Z*   (V )     Z*   (V )     Z*   T*     (H )   T*     (H )   (T*)   (H )   (T*)   (H )
R* (D )(D )   R  (D )(D )   (R*) K  (M )  (R ) K  (M )  Z* (R*) K    Z  (R*) K    (Z*) V*  Z    (Z ) V*  Z
```

```
STIMULUS PATTERNS FOR              EXPERIMENT 7     2C, 1C, 2, 4, 2, 1, 1


 110C
     CCC              CC1             01C              C11              1CC             101              110              111
V* (K*)(H )      V* (V*)(H )     V  (H*)(H*)      V  (M*)(H*)     (M*)(K ) R      (M*)(V ) R      (M )(H ) R*     (M )(M ) R*
R         V*     R         V*    (R )     V*      (R )     V*      H*       (D )    H*       (D )   (H*)      (D )   (H*)      (D )
D* (D )(D )      D  (D )(D )     (D*) K  (M )     (D ) K  (M )     T* (R*) K       T  (R*) K      (T*) V*  Z      (T ) V*  Z


 1101
     CCC              CC1             01C              C11              1CC             101              110              111
K* (K*)(H )      K* (V*)(H )     K  (H*)(H*)      K  (M*)(H*)     (H*)(K ) R      (H*)(V ) R      (H )(H ) R*     (H )(M ) R*
R         V*     R         V*    (R )     V*      (R )     V*      H*       (D )    H*       (D )   (H*)      (D )   (H*)      (D )
D* (D )(D )      D  (D )(D )     (D*) K  (M )     (D ) K  (M )     T* (R*) K       T  (R*) K      (T*) V*  Z      (T ) V*  Z


 111C
     CCC              C01             01C              C11              1CC             101              110              111
V* (K*)(H )      V* (V*)(H )     V  (H*)(H*)      V  (M*)(H*)     (M*)(K ) R      (M*)(V ) R      (M )(H ) R*     (M )(M ) R*
V         V*     V         V*    (V )     V*      (V )     V*      T*       (D )    T*       (D )   (T*)      (D )   (T*)      (D )
D* (D )(D )      D  (D )(D )     (D*) K  (M )     (D ) K  (M )     T* (R*) K       T  (R*) K      (T*) V*  Z      (T ) V*  Z


 1111
     CCC              CC1             C1C              C11              1CC             101              110              111
K* (K*)(H )      K* (V*)(H )     K  (H*)(H*)      K  (M*)(H*)     (H*)(K ) R      (H*)(V ) R      (H )(H ) R*     (H )(M ) R*
V         V*     V         V*    (V )     V*      (V )     V*      T*       (D )    T*       (D )   (T*)      (D )   (T*)      (D )
D* (D )(D )      D  (D )(D )     (D*) K  (M )     (D ) K  (M )     T* (R*) K       T  (R*) K      (T*) V*  Z      (T ) V*  Z


**** ALL INPUT DATA HAVE BEEN PROCESSED.
     AT LOCATION 10572
```

BIBLIOGRAPHY

A compendium of titles either cited or of significant influence on this study.

Berry, P.C.. An exploration of the interrelations among some non-intellectual predictors of achievement in problem solving. Technical Report No. 4, Department of Industrial Administration and of Psychology. Yale University, 1958.

Bouling, Kenneth E. The Image: Knowledge in Life and Society. Ann Arbor, University of Michigan Press, 1956.

Bourne, L.E. and R.C. Haygood. The role of stimulus redundancy in concept identification. Journal of Experimental Psychology, 58, 1959, 232-238.

Bourne, L.E. and F. Restle. Mathematical theory of concept identification. Psychological Review, 66, 1959, 278-296.

Bower, G. An association model for response and training variables in paired-associate learning. Psychological Review, 69, 1962, 34-53.

Bruner, J.S., J.J. Goodnow, and G.A. Austin. A Study of Thinking. New York, Wiley, 1956.

Bruner, J.S. and L. Postman. On the perception of incongruity: a paradigm. Journal of Personality, 18, 1949, 206-223.

Bush, R. and F. Mosteller. Stochastic Models for Learning. New York, Wiley, 1955.

Cherry, Colin. On Human Communication. New York, Science Editions, 1961.

Chomsky, N. Syntactic Structures. M. Cravenhenge, Amsterdam, 1953.

Estes, W.K. Toward a statistical theory of learning. Psychological Review, 57, 1950, 94-107.

Feigenbaum, E.A. The simulation of verbal learning behavior. Proceedings, Western Joint Computer Conference, 1961, 121-132.

Goss, A. Verbal mediating responses and concept formation. Psychological Review, 68, 1961, 248-274.

Hebb, D.O. The Organization of Behavior. New York, Wiley, 1949.

Hilgard, E.R. Theories of Learning. New York, Appleton, 1956.

Hovland, C.I. A "communication analysis" of concept learning. Psychological Review, 59, 1952, 461-472.

Hovland, C.I. and W. Weiss. Transmission of information concerning concepts through positive and negative instances. Journal of Experimental Psychology, 45, 1953, 175-182.

Hull, C.L. Quantitative aspects of the evolution of concepts. Psychological Monographs, 28, No. 1, 1920.

__________. Principles of Behavior. New York, Appleton, 1943.

Hunt, Earl B. Concept Learning: An Information Processing Problem. New York, Wiley, 1962.

Hunt, Earl B. and C.I. Hovland. Programming a model of human concept formulation. Proceedings, Western Joint Computer Conference, 1960, 145-155.

__________. Order of consideration of different types of concepts. Journal of Experimental Psychology, 59, 1960, 220-225.

Kendler, H.H. and T.S. Kendler. Vertical and horizontal processes in problem solving. Psychological Review, 69, 1962, 1-16.

Kochen, M. An experimental program for the selection of "disjunctive hypotheses." Proceedings, Western Joint Computer Conference, 1961, 571-578.

Lee, Dorothy. Lineal and Nonlineal Codifications of Reality. Psychosomatic Medicine, Vol. XII, March-April, 1950.

Logan, F.A. Incentive. New Haven, Yale University Press, 1960.

Luce, R.D. Individual Choice Behavior. New York, Wiley, 1959.

Luce, R. Duncan and Raiffa, Howard. Games and Decisions. New York, Wiley, 1957.

Miller, George A. Language and Communication. New York, McGraw-Hill, 1951.

__________. The magical number seven, plus or minus two: some limits on our capacity to process information. Psychological Review, 63, 1956, 81-97.

Miller, G.A., E. Galanter, and K. Pribram. Plans and the Structure of Behavior. New York, Holt, 1960.

Newell, A., J.C. Shaw, and H.A. Simon. Report on a general problem solving program. Proceedings, 1st International Conference on Information Processing, Paris, UNESCO House, 1959.

Oseas, L. and B.J. Underwood. Studies of distributed practice vs. learning and retention of concepts. Journal of Experimental Psychology, 43, 1952, 143-148.

Piaget, J. Logic and Psychology. New York, Basic Books, 1957.

Polya, G. Vol. 1: Induction and Analogy in Mathematics.
Vol. 2: Patterns of Plausible Inference.
Princeton: Princeton University Press, 1954.

Rapoport, A. Strategy and Conscience. New York, Harper and Row, 1964.

Riordan, J. An Introduction to Combinatorial Analysis. New York, Wiley, 1958.

Samuel, A. Some Studies in Machine Learning, Using the Game of Checkers. IBM Journal of Research and Development, Vol. 3, No. 3, July, 1959.

Selfridge, O. Pandemonium, a paradigm for learning. Proceedings, Symposium on the Mechanization of Thought Processes. London, H.M. Stationery Office, 1959.

Spence, K. Behavior Theory and Conditioning. New Haven, Yale University Press, 1957.

Taylor, D.W. and O. McNemar. Problem solving and thinking. Annual Review of Psychology, 6, 1955, 455-482.

Thompson, D'Arcy. On Growth and Form. Bonner, John Taylor, ed., Cambridge University Press, 1961.

Uhr, L. and C. Vossler. A pattern recognition program that generates, evaluates, and adjusts its own operators. Proceedings, Western Joint Computer Conference, 1961, 555-570.

Watanabe, S. Information theoretic aspects of inductive and deductive inference. IBM Journal of Research and Development, 4, 1960, 208-231.

Wells, H.H. Transfer and stimulus effects in disjunctive concept learning. Unpublished doctoral dissertation, Yale University. Cited in: Hunt, Earl B. Concept Learning: An Information Processing Problem. New York, Wiley, 1962.

Wertheimer, M. Productive Thinking. 2nd Edition. New York, Harper, 1959.

Whorf, Benjamin. Language, Thought, and Reality. Cambridge, Massachusetts, Technology Press, 1956.

RULES COVERING USE OF MANUSCRIPT THESES
IN THE UNIVERSITY OF MICHIGAN LIBRARY
AND THE GRADUATE SCHOOL OFFICE

Manuscript copies of theses submitted for the doctor's degree and deposited in the University of Michigan Library and in the Office of the Graduate School are open for inspection, but are to be used only with due regard to the rights of the authors. For this reason it is necessary to require that a manuscript thesis be read within the Library or the Office of the Graduate School. If the thesis is borrowed by another library, the same rules should be observed by it. Bibliographical references may be noted, but passages may be copied only with the permission of the authors, and proper credit must be given in subsequent written or published work. Extensive copying or publication of the thesis in whole or in part must have the consent of the author as well as of the Dean of the Graduate School.

This thesis by ....
has been used by the following persons, whose signatures attest their acceptance of the above restrictions.

A Library which borrows this thesis for use by its readers is expected to secure the signature of each user.

NAME AND ADDRESS	DATE