

THE EFFECT OF STIMULUS EMPHASIS ON
STRATEGY SELECTION IN THE ACQUISITION
AND TRANSFER OF CONCEPTS

Thesis for the Degree of Ph. D.
MICHIGAN STATE UNIVERSITY
Thomas Robert Trabasso
1961

This is to certify that the
thesis entitled
THE EFFECT OF STIMULUS EMPHASIS ON STRATEGY
SELECTION IN THE ACQUISITION AND
TRANSFER OF CONCEPTS

presented by

Thomas Robert Trabasso

has been accepted towards fulfillment
of the requirements for

Ph. D. degree in Psychology


Major professor

Date November 11, 1961

O-169



ABSTRACT

THE EFFECT OF STIMULUS EMPHASIS ON STRATEGY SELECTION IN THE ACQUISITION AND TRANSFER OF CONCEPTS

by Thomas Robert Trabasso

This dissertation reports the results of a theoretical and experimental analysis of attention in a simple concept formation task.

An analysis of the literature suggested that attention in learning can be studied objectively. Aspects of the stimulus situation can be arranged so that they control or affect the locus of attention. The speed of learning is an indicator of the probability that a relevant dimension is attended to and serves as an index of the effect of an attempt to direct S's attention. When a problem involves more than one relevant cue, or more than one irrelevant cue, or a relevant cue which can be diminished or enlarged, measurements of transfer-of-training can be used to gain more detailed information regarding the distribution of attention over the parts of the stimulus pattern.

Salient stimuli, such as (1) color or (2) a large difference between discriminanda, which affect the direction of the S's attention to cues, were varied in their contextual rôle in order to study their effect on learning and transfer of concepts. A stimulus which increases the probability of attending to a relevant stimulus is called a stimulus "emphasizer" while one which directs attention away from a relevant stimulus is called a "counter-emphasizer."

A mathematical model of discrimination learning (Restle, 1961a; 1961b) is used in the formulation of the experimental problems. The model explicitly treats the mechanisms by which S selects strategies in cue-learning and defines the probability of solution as equal to the proportion

of correct strategies in the problem. The goodness-of-fit and predictive value of the theory was tested against the data. The attention-value of a stimulus aspect was evaluated by using the model to estimate the measure of a set of strategies based on that aspect.

A transfer-of-training design known as "easy-to-hard" transfer was used to study (1) the rôle of attention and (2) efficiency in concept formation. The degree of efficiency was hypothesized to depend upon stimulus emphasis and relationships of the original learning (easy) and transfer (hard) problems.

The stimuli were complex flower patterns and the correct responses (two-choice) depended upon one or two aspects of the pattern. The relevant dimension of the hard problem was the angle of the leaves to the stem of the flower.

Nine groups, of 20 Ss each, worked on different original learning problems and were all transferred to the same hard problem after criterion in original learning. A tenth group had the hard problem as its original learning problem and served as a control. All comparisons of acquisition and transfer were relative to this control group.

In two problems, emphasis of the relevant angle dimension was achieved by either (1) doubling the difference between discriminanda or (2) removing irrelevant cues during original learning. Both groups were highly efficient: they learned very rapidly and showed nearly perfect transfer.

Color on the angle of the leaves to the stem constituted an "emphasizer." When a constant color was used on all trials, the effect was not strong; red had a detectable effect, but green did not facilitate learning at all. When color varied from trial to trial in a third problem, and was an irrelevant dimension, the net effect was slight facilitation of learning. In these three groups, color could not serve as the basis for a correct strategy and transfer to the hard problem was perfect.

Two problems had color added as a redundant and relevant dimension during original learning. Both problems were learned faster than problems with only one dimension relevant, an example of "additivity of strategies." In one problem, the color was also an emphaser and transfer to the hard problem was somewhat positive. In a second problem, color was a counter-emphasizer, appearing over the flowers during original learning, and transfer to the hard problem was slightly negative.

Two control problems had color relevant and the angle dimension fixed. Color was found to be more salient as a cue than the angle. There was no evidence for transfer of an "observing response" to the angle in these groups.

The stochastic properties of the data were consistent with the expectations of the Strategy Selection Theory. Analyses of Ss' performances before criterion indicated that errors occurred at random with probability near one-half, constant and independent of how long S was in the pre-solution phase. Fitted theoretical error distributions yielded good approximations in eleven of twelve cases. There was some evidence that Ss use "wrong" as well as irrelevant strategies in the pre-solution phase. Since wrong strategies depend upon the same cues as correct strategies, it was predicted that estimates of the measure of wrong strategies would be about the same as estimates of correct strategies. This quantitative prediction was verified. Wrong strategies were detected and their measure correlated with the measure of correct strategies. Inter-correlations between practice, original learning and transfer problems indicated no stable individual differences.

By taking account of stimulus emphasis, and using the Strategy Selection Theory, the additivity of relevant strategies and additivity of irrelevant strategies were accurately predicted. The degree of transfer was predicted in three ways: (1) number of Ss showing perfect transfer,

(2) mean errors in transfer and (3) cumulative distributions of error scores in transfer. All predictions were based on parameters which had been estimated from original learning data and independent groups. Seven of eight predictions on transfer were accurate.

Efficiency in concept learning was discussed in relation to the present and other findings. The question of the precise rôle of a stimulus emphasize was examined and further investigations on emphasizees suggested.

References:

- Restle, F. The selection of strategies in cue learning. Psychol. Rev., 1961a (in press).
- Restle, F. Statistical methods for a theory of cue learning. Psychometrika, 1961b, 26, 291-306.

Approved: Frank Restle
Frank Restle, Major Professor

Date: 11-11-61

THE EFFECT OF STIMULUS EMPHASIS ON STRATEGY
SELECTION IN THE ACQUISITION AND
TRANSFER OF CONCEPTS

By

Thomas Robert Trabasso

A THESIS

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

DOCTOR OF PHILOSOPHY

Department of Psychology

1961

5 20920
5/24/92

DEDICATION

To Sue

ACKNOWLEDGMENT

The author wishes to express his sincere gratitude to Dr. Frank Restle, chairman of his committee, who willingly gave time and energy to the planning, execution and development of this manuscript.

In addition, he wishes to thank Drs. T. M. Allen, A. B. Barch, D. M. Johnson, and M. R. Denny, members of his committee, who have lent their criticism and advice during the preparation of this thesis.

Finally, he thanks John N. Schneider, who helped collect some of the data in the experiment.

TABLE OF CONTENTS

CHAPTER	Page
I. INTRODUCTION	1
II. THE STRATEGY SELECTION THEORY.	24
III. METHOD	38
IV. STOCHASTIC PROPERTIES OF THE DATA AND TEST OF THE STRATEGY SELECTION THEORY	47
V. EXPERIMENTAL RESULTS.	65
VI. DETAILED PREDICTIONS.	76
VII. DISCUSSION	91
VIII. SUMMARY	96
REFERENCES	99

LIST OF TABLES

TABLE	Page
3.1 Experimental Groups and Problems	45
4.1 Test of Stationarity: 1st versus 2nd Half Errors in the Pre-solution Phase	49
4.2 Maximum-likelihood Estimates of c	53
4.3 Maximum Discrepancies Between Theoretical and Observed Error Distributions	59
4.4 Test of Consecutive Errors: Proportion of \bar{S} s Making More, As Many, or Fewer Errors Following Errors than Correct Responses Following Errors.	61
4.5 Estimates of w , the Proportion of Wrong Strategies, Compared with Estimates of c	62
4.6 Inter-correlations Between Practice (P), Original Learning (OL), and Transfer (T) Problems.	64
5.1 Comparison of Large Angle and Angle Only with Angle on Original Learning	65
5.2 Transfer to the Angle Problem by Large Angle and Angle Only Compared with Original Learning of Angle	66
5.3 Comparison of Red Angle and Green Angle with Angle on Original Learning	67
5.4 Transfer to the Angle Problem by Red Angle and Green Angle Compared with Original Learning of Angle	67
5.5 Comparison of Angle + Angle Color Irrelevant with Angle on Original Learning and Transfer	68

LIST OF TABLES - Continued

TABLE	Page
5.6 Comparison of Angle + Angle Color and Angle + Flower Color with Angle on Original Learning. . . .	69
5.7 Transfer to the Angle Problem by Angle + Angle Color and Angle + Flower Color Compared with Original Learning of Angle	70
5.8 Comparison of Angle Color Control and Flower Color Control with Angle on Original Learning	71
5.9 Transfer to the Angle Problem by Angle Color Control and Flower Color Control Compared with Original Learning of Angle	72
6.1 Formulas of c for Experimental Groups Used in Predictions	78
6.2 Estimates of Sets of Strategies Used in Predictions .	78

LIST OF FIGURES

FIGURE	Page
2.1 Stochastic Structure of Strategy Selection	27
2.2 A Geometric Distribution showing the probability of obtaining k tails before the first head in tossing a fair coin six times	29
2.3 Hypothetical geometric and normal distributions of error scores in a learning experiment	29
3.1 Examples of flower patterns used in experimental problems	42
4.1 The probability of an error, conditional on <u>S</u> being in the pre-solution phase	51
4.2 Observed and theoretical cumulative distributions of	
4.3 the proportion of <u>Ss</u> making n or fewer errors . . .	55-58
4.4	
6.1 Predicted and observed cumulative distributions of the proportion of <u>Ss</u> making n or fewer errors . . .	86

CHAPTER I

INTRODUCTION

The present study addresses itself to three central problems in current learning theory:

1. What makes a problem easy or difficult to learn?
2. Under what conditions will learning transfer to new problems?
3. What roles does attention play in learning and transfer?

Answers to these questions have a practical value (e.g. for use in educational devices such as teaching machines) and are relevant to basic theoretical issues (e.g. the question of continuity versus non-continuity of learning).

Attention, as will be shown, has been neglected in learning theory. The term carries several meanings so that its usage is frequently not precise. Attention, as used in the present study, is defined as:

the active selection of, and emphasis on, one component of a complex experience and the narrowing of the range of objects to which the organism is responding; the maintenance of a perceptual set for one object and disregard for other. (English and English, 1958, p. 49)

An illustration of the role of S's attention to cues in a learning situation is given by the following example:

An instructor in a biology laboratory wishes to train a student to discriminate a micro-organism from the background debris on a slide. To facilitate the student's identification of the micro-organism the instructor might (a) color the micro-organism or (b) point to it with a black line. The instructor's intervention into the stimulus situation via the color or line elicits an attending response of the student to the micro-organism. The additional stimulus in the situation has the effect

of "pushing" another stimulus, the micro-organism, to the fore.

To facilitate learning, the biology instructor's method of pointing out the micro-organism must make it easy to identify. Later, for the training to be judged effective, the student must be able to identify the micro-organism without the added stimulus. An effective teaching method should facilitate learning and lead to good transfer of training.

The learning problem can be complicated by requiring the student to learn to differentiate the micro-organism not only from the background but also from other micro-organisms. This requires further training, which might consist of the student looking at a set of slides, in a series, with each slide containing a different micro-organism. The student's job would now be to (1) discriminate each micro-organism from the background and (2) discriminate each micro-organism from the others in the collection and correctly label them. The task has now become one involving conceptualization of the stimuli as well as discrimination from background cues. The training procedure in this example is analogous to one commonly used in concept formation.

In this thesis, aspects of the stimulus situation which can be used to direct the S's attention are studied. An aspect, such as the color or black line in the example, which increases the probability of attending to a relevant stimulus is called a stimulus "emphasizer." The emphasizee itself may or may not be a cue to solution and thus, its own rôle can vary. The effect of an emphasizee is to facilitate learning by increasing the likelihood that S uses the relevant stimulus.

A second rôle of a salient stimulus such as color may be to distract S away from the relevant cue. Suppose the instructor inadvertently colored a smudge on the slide. The student's attention would be directed by the color away from the micro-organism. Identification would then be retarded.

An aspect of the stimulus situation which decreases the probability of the S attending to the relevant cue is called a "counter-emphasizer."

In the example, if the colored smudge served as a counter-emphasizer and the student responded to it as if it were the micro-organism, it is likely that no transfer of learning would occur on a test situation which contained the micro-organism and no smudge.

This example, with salient stimuli as emphasizees and counter-emphasizers, suggests a way in which attention and its effect on learning can be investigated in a set of concept formation problems.

Concept formation has become a focal point of experimental effort during the last decade (see Kendler, 1960, for a good review). Methods of investigating concept formation are formally like those used in the study of discrimination learning. In general, S is required to learn the same response to objects of the same class, but a different response or no response at all to objects belonging to other classes.

Hull (1920) used what may be called a "modified memory" method in an early study of efficiency of concept learning. Nonsense names were assigned to Chinese characters and S was required to learn to name each character. The relevant aspect of the character was called a "radical" and the radicals were embedded in many compound characters. Each character contained one of twelve radicals and all characters with the same radical were assigned the same name. By use of a memory drum, packs of 12 radicals were presented serially and after a practice trial, learning was by the anticipation method.

A second approach is that of Heidebreder (1946) who had Ss sort a pack of 144 cards into nine piles. Each pack contained three kinds of objects, three kinds of forms and three kinds of number groups. Objects were easiest and number the hardest to learn.

In both the Hull and Heidebreder approaches, the S had to search for a basis of classification within each stimulus pattern. In order to perform the task accurately, S must resort to some form of "conceptual behavior."

Concept formation in the Hull approach resembles learning of paired associates. One of the most successful approaches toward understanding of the learning of paired associates is the idea that errors arise through confusion between stimuli. Similar stimuli within a list are difficult to learn (provided different responses are required) and errors confusing two similar stimuli are more frequent than errors confusing two dissimilar stimuli. Factors in the difficulty of learning paired associates can be interpreted as depending upon a tendency to make the same response to two stimuli which are similar, i. e., to "generalization." The main statement of this approach is by Gibson (1940).

The similarity of concept formation to discrimination learning allows the application of theory and advanced experimental techniques of discrimination learning to the study of concept formation. At the same time, the similarity revives theoretical issues which have been raised in discrimination learning; namely, continuity of learning, the nature of selectivity and, as mentioned above, the rôle of attention.

Since the position has been taken that attention is an important part of the learning process and that attention has been neglected as a factor in learning, a review of some of the thinking on this problem by learning theorists is made first (question 3 above). Then, stimuli which influence the ease or difficulty of a problem are considered (question 1). Finally, the discussion is centered on those studies which have dealt with efficiency (question 2) via the "easy-to-hard" transfer paradigm. This review will attempt to show that divergent results on transfer can be obtained from seemingly similar procedures. These topics, (1) attention in learning theory, (2) stimulus sources of difficulty, and (3) experiments on easy-to-hard transfer, together provide the setting for the present experimental problem. Chapter II contains a detailed treatment of the experimental problem formulated in terms of a theory of cue learning called the Strategy Selection Theory (Restle, 1961a; 1961b).

Attention and Learning Theory

The complaint that attention has been neglected is not new nor infrequent. William James (1890), in an introduction to his chapter on attention, had the following to say:

Strange to say, so patent a fact as the perpetual presence of selective attention has received hardly any notice from psychologists of the English empiricist school. The Germans have explicitly treated of it, either as a faculty or as a resultant, but in the pages of such writers as Locke, Hume, Hartley, and the Mill's and Spencer, the word hardly occurs, or if it does, it is parenthetically and as if by inadvertance. (p. 402)

James defined attention as a process with a locus in the "mind," where one out of several possible objects was selected and made clear or vivid. Attention was not an isolate, for:

The immediate effects of attention are to make us: (a) perceive-- (b) conceive-- (c) distinguish-- (d) remember-- better than we otherwise could-- both more successive things and each thing more clearly. It also (e) shortens 'reaction time'. (p. 425)

The set of problems which have arisen out of the study of attention, before and since James, has been summarized by Woodworth and Schlosberg (1954). In general, the experiments deal with the problem of S making some selective response such as looking at one of several simultaneously presented objects. The response depends on physically defined stimuli and organismic variables such as past experience or set. Several problems are distinguished; namely, (1) the stimulus determinants of attention, (2) shifting and fluctuation of attention, (3) distraction, (4) divided attention--doing two things at once, and (5) span of attention. In the recent Zeitgeist on attention, experimental effort has been directed toward effects of simultaneous presentation of two stimuli and the question of which is attended to first (Broadbent, 1958) and on internal states of arousal (Berlyne, 1960).

In the concept formation task (e.g. Hull's) a number of stimuli are competing for the S's attention at the same time. Some are relevant and can be used as a basis for solution while others are irrelevant and do not lead to a correct strategy. Stimuli which are salient would be the ones most likely to be attended to and be used as the basis for a strategy. Thus, a salient stimulus which was relevant would be more likely to facilitate learning than one which was not salient but relevant.

Hull, in his 1920 monograph on concept information, did not neglect the rôle of attention. He considered the use of a salient stimulus in posing the question:

What is the relative efficiency of evolving functional concepts from concrete cases in which the attention of the subject is continuously attracted to the significant common element, as compared with the ordinary simple-to-complex method? (p. 51).

Hull assumed that a saturated red on the common element (radical) would attract the S's attention to the relevant aspect and lead to faster learning. For each S, each list of twelve Chinese characters had the same six radicals colored red, while the remaining six were left black. Colored and black radicals were counter-balanced over Ss. The red symbols were learned faster and showed more transfer on a test series with no red than those symbols which were black in the training series. Hull concluded that there was a distinct advantage where the attention of S is attracted to the common element in situ.

Despite Hull's early demonstration of the rôle of attention in a learning task, S-R theorists have given attention no detailed treatment. Attention was excluded from the study of learning largely because it had "mentalistic" import. Instead, attentional theories (e.g. Lashley, 1929) became a source of criticism of behaviorism. S-R theorists continued to concentrate on observable responses and their reinforcement, precluding stimulus factors which direct the S's attention.

The classic S-R approach of Hull and Spence regards learning mainly as the manifestation of approach responses to reinforced stimuli and avoidance responses to non-reinforced or punishing stimuli. Spence (1937; 1940) recognized that during discrimination training, S learns responses which increase the likelihood of his being stimulated by the relevant stimuli. Spence called these responses "receptor-orienting" but did not attempt to account for their development. Berlyne (1951) made a plea for the inclusion of perception and, particularly, attention, into S-R theory. Berlyne suggested that the perceptual process be treated as an intervening variable in Hull's system. The selective nature of learning was stressed since it reveals the direction of attention. Stimulus variables, such as (1) intensity, (2) change, (3) postural adjustments and preparatory sets, and (4) organization of the perceptual field, were mentioned as some determinants of attention.

A stimulus determinant of attention which has been singled out for investigation in the present study is a "salient" stimulus, used as an emphasisizer of another stimulus. The question, What is salience? may now be asked. Salience may be regarded as a symptom of attention as well as an influence on its direction. William Stern defined salience phenomenologically as the degree to which an experience stands out sharply and is relatively disconnected from the rest of the experience. An antonym of salience is "embeddedness." A salient stimulus, then, is one which is relatively prominent in the psychological field in relation to other stimuli. Salience is not intensity of the stimulus (though they may be related) but it is a "distinctiveness" and a sort of immediately perceived importance (English and English, 1958, p. 471). Color, then, as used by Hull (1920) seems to fit this definition of salience.

Wyckoff (1952), following Spence's receptor-orienting notion, published a probability model calling the learned response an "observing response." The observing response, R_0 , is learned by the principle of

secondary reinforcement. The probability, p_0 , of the observing response bears a circular relation to the speed of learning. The observing response is learned to the extent that its occurrence increases the probability of reinforcement, but the observing response can increase the probability of reinforcement only if S is responding above chance, i. e. learning. Wyckoff tested his theory on pigeons (reported in Prokasy, 1956) where S was required to learn to step on a pedal (observing response) to the flashing of a white light in order to be presented with two colors to be discriminated and receive reinforcement. The pedal press response was learned without direct reinforcement. Prokasy (1956) showed that rats would develop right or left turning habits (observing responses) to the side where consistent reinforcement of a black-white discrimination could be obtained. The observing response notion of Wyckoff offers some possibility of a rapprochement between S-R and attentional theorists. However, Spence and Wyckoff seem more concerned with the form of the observing response and the reinforcement of it, than with either the stimulus conditions (cues) to observing or the detailed changes in the stimulus input which may be brought about by observing.

Lawrence (1949; 1950) offered a mediating process called "acquired distinctiveness of cues" which takes into account the previous experience of S in determining the selection of cues and rate of learning in new situations. Lawrence trained rats on simultaneous and then on successive discriminations with the same stimulus dimensions. New instrumental responses were learned faster when the cues were familiar (had been previously reinforced). Selectivity of responding was demonstrated with respect to the relevant stimuli which would become associated in the new learning situation. Lawrence stated that discrimination learning consisted of two processes: (1) a change in the perceptual character of the stimuli which was brought about by prior learning, and (2) then the association of stimuli with instrumental responses. With human Ss, Kurtz (1955)

showed positive transfer of a response where the training stimuli were not identical to the test stimuli but where they were distinguished by the same property as the latter; and negative transfer where the training stimuli were not identical to the test stimuli and were distinguished by a different property. This study supported Lawrence's notion of acquired distinctiveness, but the cues were relations rather than specific stimulus dimensions. Lawrence's concept of a mediating process which causes "acquired distinctiveness of cues" brings out changes in the effectiveness of stimuli due to training but says little about the cues which give rise to the mediating process.

A third body of experimental literature has dealt with learning to "ignore" stimuli. Hammer (1955) tried to determine whether human Ss learn to not attend to irrelevant stimuli. In one group, letters which were irrelevant in a training problem were made relevant in the transfer problem. In a second group, irrelevant letters of the training problem were retained as irrelevant stimuli in the transfer problem. If Ss learned to not attend to irrelevant stimuli, the first group should show negative and the second, positive transfer. No differences in transfer performance were found and no evidence for Ss learning to ignore irrelevant stimuli during training.

LaBerge and Smith (1957) derived a hypothesis from stimulus sampling theory (see Estes, 1959), and found that Ss who respond asymptotically ignore "background common elements" which were associated with partial reinforcement. Blank trials were inserted in the training series as a test and when the partial reinforcement schedule associated with the background stimuli changed, Ss at asymptote did not change their responses but those who were not at asymptote did change on the blank trials. Hughes and North (1955), on rats, found that Ss attend to partially correlated cues after learning had taken place on other cues, a result in opposition to the LaBerge and Smith finding. Hughes and North first

trained the rats to discriminate form. After criterion was reached, during a series of overlearning trials, color (black-white) was partially reinforced (75% versus 25%). Training was then given on a black-white discrimination problem and transfer to this problem was found to be positive.

Restle (1955), in a mathematical theory of discrimination learning, explicitly describes a dual process by which S learns to make correct responses. Cues which are relevant become conditioned to the response while those which are irrelevant become "adapted." Restle (1959) tested the effect of making one of two relevant cues irrelevant (by "scrambling") after conditioning and adaptation had taken place. He found that about 50% of the previously adapted cues became unadapted and interfered on transfer. Restle claimed that "background cues" which are irrelevant throughout the experiment are neutralized or adapted during original learning. Neutralization is not a form of "learning to ignore" cues, for a cue will remain neutralized only if the relevant cues remain present and relevant. Thus an irrelevant cue is neutralized with respect to some relevant cue. Restle (1958) has applied the same reasoning to learning set data and found some good approximations of his model to the data. The results were used to explain Hammer's failure to find a transfer of neutralization since relevant cues were not carried from one problem to the next. Thus, Restle's notion of adaptation of irrelevant cues gives a quantitative account of changes in the stimuli due to a sort of observing process. His hypothesis that the rate of adapting depends upon the proportion of relevant cues relates these changes in perception to characteristics of the stimulus presented. The connection is, however, an ad hoc assumption and makes the S somewhat prescient.

In a different approach to selectivity, Harlow (1959) operationally defines "hypotheses" that monkeys use in forming concepts and discriminating objects. Harlow calls this analysis "error factor theory."

For example, a "stimulus perseveration error factor" is said to occur when S makes repeated choices of the incorrect stimulus object. Moon and Harlow (1955) studied a number of error factors and found that these responses extinguish progressively throughout the course of learning set formation. Harlow and Hicks (1957), in a "uniprocess learning theory," describe discrimination learning as a process of eliminating error factors. By "uniprocess" these writers mean that S is not trying to learn correct responses but is learning to not make incorrect responses. Although Harlow's error factor theory describes the gradual removal of erroneous stimulus-response connections, it does not give either the cue to observing or the changes in stimuli which might form a basis for the removal of error factors.

Restle (1961a; 1961b) has recently published a new discrimination learning theory (the Strategy Selection Theory, see Chapter II) which represents a complete revision of his position in the 1955 model. In the new theory, explicit assumptions regarding the mechanisms by which S solves cue problems are made. The S is assumed to select "hypotheses" or "strategies" at random from a set of strategies (which are determined by the stimulus situation) and to respond accordingly. If the response is correct, S continues to use the same strategy; if the response is in error, S resamples a new strategy (with replacement) and continues testing. Sampling of strategies may be (a) one-at-a-time, (b) all-at-once, or (c) a random sample of strategies. The result is the same set of probabilities of success or failure. Learning is described as discontinuous and the model is similar to the earlier attentional theories of Lashley (1929; 1938) and Krechevsky (1932). The theory handles most of the quantitative results as accurately as the 1955 model and its applications (Restle, 1957; 1958; 1959).

Lashley (1929) emphasized the perceptual and selective nature of discrimination learning. He criticized S-R theory by stating:

The description of discrimination as a mere combination of a positive and a negative reaction misses the essential features of the process, which are isolation of figure, the discovery of differences and the generalizing characteristic of the responses. These are prior to and not a result of training. (p. 184)

Krechevsky (1932; 1938) in his emphasis on the use of "hypotheses" by rats was influenced by the thinking of Lashley with respect to the discontinuity and the selective nature of learning. Krechevsky described the discrimination process as one where the S selects out of the situation certain stimuli to which he attends and continues to do so until he learns that they are not correct. The S then gives up responding to these stimuli and proceeds to select another set of stimuli to respond to. Krechevsky invoked certain Gestalt principles to account for the formation of hypotheses, such as

the stimulus configuration forcing a specific response. . . . A hypothesis is the individual's interpretation of the data; it is not a phenomenon deriving from the presented data alone. (p. 532, 1932).

Lashley and Wade (1946), again opposing S-R theories, credited the stimulus situation with an important role in the determination of selective responding. The concept of "perceptual dominance" was used to describe the fact that one stimulus dimension (e.g. color) may predominate over another (e.g. size) in the field even though both are relevant.

If a monkey is trained to choose a large red circle and avoid a small green one, he will usually choose any red object and avoid any green but will make chance scores when like colored large and small circles are presented. (p. 82)

Warren (1954) tested this statement by training monkeys on problems where two or more stimuli were relevant and redundant (either one or both could be used as the basis for a correct strategy) and then transferred them to problems where one of the relevant dimensions was removed. Although color controlled responses more than other dimensions, the monkeys showed positive transfer to other dimensions, indicating that

they learned something about the second redundant dimension during training. Although Lashley and Wade concentrate on the distribution of attention over stimuli, they do not take much account of how modifiable this distribution is or of the details of the stimuli involved.

In summary, S-R theories, most of which assume learning to be continuous, have neglected attention in learning and offer no detailed explanation for it as a part of the discrimination process. On the other hand, attentional theories, most of which assume learning to be discontinuous, have been used to criticize S-R theories, but are vague and not precise as to mechanisms and conditions for selectivity of responding.

Stimulus Sources of Difficulty

Consider a two-choice learning problem. The S must attempt to perform two tasks simultaneously. Using Woodworth and Schlosberg's (1954) familiar notation, the double task may be represented by the formula:

$$R_1 R_2 = f(O_1 O_2, S_1 S_2).$$

The question is whether R_1 can be connected with S_1 and R_2 with S_2 . If other stimuli, S_1' and S_2' are present and they call for R_1 and R_2 respectively, then similarity between S_1 and S_1' facilitates learning but similarity between S_2 and S_1 retards learning.

The degree of difference between two discriminanda, S_1 and S_2 , is a well-established variable which can influence the ease or difficulty of learning. A reduction in the difference (by making S_1 and S_2 more similar on the same continuum) results in slower learning. This relation holds for identifiable physical continua such as brightness, color, size, shape, pitch, number, etc. Murdock (1960) has recently defined "distinctiveness" on the basis of difference between discriminanda. In the easy-to-hard

training procedure, one of the main methods for making a hard discrimination easier is to manipulate the magnitude of the difference between discriminanda (with color, Lawrence, 1952; with pitch, Baker and Osgood, 1954; and with size, Restle, 1955).

Early experimenters on concept identification (Hull, 1920) and discrimination learning (Lashley, 1929; 1938) used "complexity" of the stimulus as a source of difficulty. Hull, in his monograph, made the radicals more difficult to identify by embedding the radical in a context with several irrelevant stimuli. Lashley trained rats on embedded figures and tested them on simple ones in an effort to determine to which aspects the rats were responding.

The modern view of complexity of the stimulus is analytic, and systematic research has dealt with the major variables of relevant and irrelevant dimensions. Archer, Bourne and Brown (1955) and Bourne and Restle (1959) report cases where difficulty increases as a function of the number of irrelevant dimensions in the problem. Archer, Bourne and Brown used information theory to explain the result, contending that each added irrelevant dimension increases the alternatives and slows learning. Bourne and Restle assumed that learning rate depends upon the proportion of relevant cues and this proportion decreases as irrelevant cues are added.

Relevant dimensions, when added and made redundant, lead to faster learning (Eninger, 1952; Warren, 1953; 1954; Restle, 1959; Bourne and Restle, 1959; Trabasso, 1960). This effect of faster learning with added redundant relevant cues has been described as "additivity of cues" (Restle, 1955). Additivity of relevant and irrelevant cues can be handled quantitatively, even though their effects are opposite (Bourne and Restle, 1959).

The nature of the concept to be learned is related to difficulty. Heidbreder (1946), over a long series of studies, has consistently

demonstrated that object, form and number concepts are ordered in difficulty with object concepts the easiest. In more recent studies, where the stimuli are binary dimensions of color, form, size, etc., the rate of learning on each of the several dimensions appears to be about equal, except that color is somewhat faster (Bourne and Restle, 1959). Warren (1953) has also confirmed that color is a more salient cue than form or size for monkeys, although form and size are about equal. However, Hara and Warren (1961) have indicated that inequalities of the strength of dimensions occur because the stimuli have not been psychophysically scaled. Scaling the stimuli on the basis of discriminability with cats, Hara and Warren were able to show that, for form, size and brightness, discriminations combining equally detectable differences in different sensory continua produced faster learning in cats (additivity of cues). Ss were also trained on problems where two cues were correct (small black figures versus large white ones). Critical tests of equivalence were then made by posing small white figures (+ form and - brightness) against large black ones (- form and + brightness). No preferences were shown when the cues were equated in terms of discriminability for an S and then opposed in these critical tests and averaged over Ss. Hiedbreder concluded that the function of the concept, or the S's familiarity with it, controlled the difficulty of learning. This later work suggests that the difference in difficulty in her concepts were the result of unequal discriminability.

Separation of the relevant aspect from the background in the stimulus situation leads to faster learning. Object (3-dimensional) discriminations are easier to learn than pattern (2-dimensional) discriminations, even where the relevant cues are identical (Harlow, 1945). A similar result has been shown with mentally retarded children (House and Zeaman, 1960). In these studies, by making the patterns three dimensional, the investigators were adding stimuli to the situation which lead to a

clearer separation of the relevant cue (figure) from the irrelevant cues (ground). Hull achieved a similar effect by coloring the radicals of the Chinese characters red. Blazek and Harlow (1955) made discrimination problems easier by increasing the color area on a two dimensional surface. North (1959) achieved the same result with rats by filling in forms such as triangles and by making bars over which rats had to crawl, thicker. Warren (1953) found that larger forms were easier to learn even though the forms to be discriminated were of the same size. Restle (1958, p. 88) described this result as an effect of an increase in the proportion of "valid" cues, and derived a theoretical function to fit Warren's data. The interpretation here is that these results depend on the use of a stimulus emphasize. In each case, some stimulus is added which makes the relevant aspect more salient and increases the learning rate.

Easy-to-Hard Transfer

The review of the work on sources of difficulty in learning suggests that any discrimination or concept formation problem can be made easier by changing the stimulus situation. For example, instead of a small difference between discriminanda, one inserts a larger difference and the problem becomes easier. But, what is the effect on subsequent transfer to the hard problem which involves the same relevant dimension and a reduced discriminative difference?

Earlier theories such as those of Thorndike (1914) and Guthrie (1935) would expect that the transfer be positive since there are "identical elements" or a number of common "conditioners" in the two tasks. The degree of transfer could not be specified in detail from either of these theoretical positions. Precisely what is meant by "identity" is not clear in the Thorndikian viewpoint but an interpretation is that it means

any clearly discriminable aspect which is the same in the two tasks (McGeoch and Irion, 1952, p. 343).

Estes' (1959) stimulus sampling theory, which is formally similar to the model of Bush and Mosteller (1951), represents a modern version of Guthrian theory. Estes assumes that the probability of a response is equal to the proportion of stimulus elements in the trial sample which are connected to the response. In the transfer problem, the stimulus sample would presumably contain elements which became connected to the correct response in the easy problem. Transfer would then be positive. The transfer situation in the hard problem would resemble what Estes has termed "stimulus compounding." In a study by Schoeffler (1954), human Ss were trained to first discriminate two "disjoint" (non-overlapping) sets of signal lights. After Ss reached 100% discrimination in responding, they were tested on new combinations of the lights. Schoeffler made very accurate predictions on the proportion of responses to sets of stimuli which represented various combinations of the previously discriminated sets.

S-R theorists, notably Gibson (1940) and Spence (1938), rely on stimulus generalization as a basic working assumption. For Gibson, if the stimuli in the two problems are similar and call for the same response then transfer would be positive. Spence and Hull (1950) base their expectations on the concept of generalization gradients. Spence assumes that each response which is reinforced to the positive stimulus results in some increment in habit strength and non-rewarded responses to the negative stimulus add to the habit of not responding to that stimulus. These tendencies of positive and negative responses generalize to similar stimuli along the continuum.

Both the "identical elements" and gradient theories agree that transfer between two tasks is greater the more similar the tasks. It seems to follow that the most efficient way to teach any task is to give training on

that task itself. If training is given on any other task, there will be a loss in transfer. The theoretical result is that any program of training on a task A and transfer on task B must take longer, or produce more errors, than a program in which all training is on task B.

Lawrence (1952) trained rats on a black-white discrimination problem and found that it is more efficient to train the S first on an easier discrimination along a dimension and shift to a harder discrimination on the dimension, than to give all the training on the hard problem. This easy-to-hard transfer result is in conflict with the expectation that transfer is most positive when the training and test discriminations are identical. Lawrence (1955) has shown that to account for his results by generalization gradients, one must postulate very specific gradients and make unlikely assumptions about how habit strengths add.

In a review of Lawrence's work, Estes (1956) commented:

These findings seem intuitively reasonable, but I do not see (and neither, evidently does Lawrence (1955)) that they can be handled in detail by any available theory. (p. 23)

Estes failed to note that Restle's (1955) theory, reviewed in another context, did handle Lawrence's transfer data.

In the Lawrence 1952 study, the easy problem was a black versus white discrimination and the hard problem pitted dark-gray against light-gray so that the easy-to-hard transfer constituted what Lawrence called "transfer along a stimulus continuum." Transfer from one easy problem (at the end of 30 fixed trials) to the hard was found to be positive but a marked disruption in performance occurred at the point of transfer. A second group, which had a problem of intermediate difficulty, was transferred at the end of 50 trials and showed positive transfer with no disruption. A third group, which worked through a series of gradually more difficult discriminations showed the best positive transfer and very little disruption.

Baker and Osgood (1954) trained groups of human Ss on discrimination problems involving pairs of tones which differed only in frequency. The design was similar to Lawrence's. A fixed number of trials was used for training and the dependent measure was a difference in errors between a pre- and post-test. Positive transfer was obtained where the test was approached through a series of problems which became more difficult gradually. Ss who were trained on a very easy pitch discrimination and then shifted to the test showed a deterioration (not significant) in performance on the test problem. This result suggests that practical application of such a training design to human learning may be limited by the possibility that fine and difficult discriminations are often required in real life and that transfer from an artificially easy problem to a hard one may be poor.

Restle (1955), in the same study in which he accurately predicted Lawrence's transfer data, trained human Ss to criterion on an easy discrimination (large vs. small black squares) and transferred them to a hard one where the difference in size was reduced. Transfer was positive with some disruption at the point of transfer. Again, the model correctly predicted the amount of transfer.

North (1959) performed two easy-to-hard transfer experiments on rats. Ss were trained first on very "distinctive" forms (solid black triangles or thick bars) and then transferred to less distinctive forms (striated triangles or thin bars). The relevant relation was horizontal vs. vertical position of the forms. In the triangle experiment, Ss learned the solid triangle problem faster than controls with the striated triangles. Transfer, after 40 interpolated overlearning trials, was almost perfect. In the bar experiment, two groups were run, one with a gradual transition (big-medium-small thickness) and one with an abrupt transition (medium-small). No overlearning trials were interpolated. The gradual transition group from big to medium showed nearly perfect transfer but performance

on the small bars was disrupted somewhat. The abrupt transition group made nearly perfect transfer from the medium to the small bars. In both experiments, North assumed that the filled triangles and thicker bars furnished "richer cues for discrimination" and were "more structured" than the striated triangles or thinner bars. The results were interpreted as a demonstration of Lawrence's "acquired distinctiveness of form stimuli."

House and Zeaman (1960) trained mentally retarded children on a hard pattern problem using an easy-to-hard design. Ss in two groups were trained to criterion on objects (3-dimensional) and then transferred to patterns (2-dimensional). In one group, the pattern was the same in both the object and pattern problems so that relevant cues were identical in the training and transfer problems. In a second group, the set of forms and colors which appeared in the object problem were different from those in the transfer problem. A control group had the hard pattern problem without the benefit of prior training. The two experimental groups were about equal performance on the object problems. The group which had the same cues on transfer showed nearly perfect transfer. The group which had different cues in the object and pattern problems made more errors on transfer. The control group performed worst of all. An application of Restle's 1955 model to the transfer data of the identical relevant cue object-to-pattern group failed to predict the amount of transfer. House and Zeaman interpreted their results as supporting both Lawrence and Wyckoff. The interpretation was that an observing response was transferred. The observing response was defined as consisting of "looking at the color and form cues."

In two further experiments on easy-to-hard transfer, Ss were trained first on an easy problem which contained two relevant and redundant cues. Transfer was to problems where the number of relevant cues was reduced. Warren (1954) trained monkeys on six types of problems, using combinations of color, form and size as stimuli. Restle (1959) trained

human Ss on problems involving consonant letter patterns. In both experiments, Ss made a large number of errors on transfer to reduced cue problems. Overall transfer tended to be slightly positive. The easy-to-hard program as a whole took more trials and errors than direct training on the hard, single cue relevant, problem.

These easy-to-hard transfer experiments can be classified into three procedural types with differing results:

1. Both abrupt and gradual movement in a physical stimulus dimension produces faster learning and sizeable positive transfer. Performance is somewhat disrupted at the point of transfer when the transition is abrupt. The easy-to-hard program produces less total errors than the hard problem given along. (Lawrence, 1952; Baker and Osgood, 1954; Restle, 1955).

2. Separation of the relevant cue from irrelevant cues (background) by increasing the salience of the relevant cue during training leads to faster learning. Transfer to a problem involving the same relevant cue without the salient aspect is nearly perfect. This easy-to-hard program is much shorter than the hard problem given alone. (North, 1959; House and Zeaman, 1960).

3. Addition of relevant and redundant cues makes the training problem easier to learn but the removal of one of the redundant cues on transfer produces disruption at the point of transfer. Overall transfer is only slightly positive. This easy-to-hard program produces as many or more errors than the hard problem given alone. (Warren, 1954; Restle, 1959).

Applying the notion of a stimulus emphasisizer to these data, the supposition is that Lawrence's method of working in on a dimension has a dual effect. The increased difference between the discriminanda makes the relevant dimension more salient and draws the attention of the S to the relevant cue, thereby increasing the probability of the S using that

dimension as the basis for a correct strategy. In this case, the stimulus emphasize is a part of the relevant dimension. The black-white discrimination draws the rat's attention to the brightness dimension by making it "stand out" at the expense of other, irrelevant stimuli. At the same time, the quality of the dimension has been changed somewhat. A kind of observing response oriented to these grosser aspects of the differential cue is thereby produced. On transfer, some rats overlook the now subtle difference between the cues and errors occur. An identical supposition is made with respect to the findings of Baker and Osgood and Restle (1955).

In the North and House and Zeaman studies, stimulus emphasis is achieved in a different way. In their investigations, the emphasize was not an integral part of the relevant dimension (i. e. it was not relevant itself) but a stimulus aspect (3-dimensionality) was added which made the relevant cues salient. The problem was made easier because the emphasize drew the attention of S directly to the relevant cues or relevant relations which appear in the hard problem. On transfer, the emphasize was removed but had no disruptive effect. This nearly perfect transfer indicates that the emphasize itself was not a basis for correct strategies in the easy problem.

With regard to the Warren and Restle (1959) results, an alternative way of solving the problem was inserted by the addition of a relevant and redundant cue, making the training problem easier. If the S attends to one relevant dimension and solves the problem with strategies based on that dimension, he may not learn that the other dimension is also relevant. If he solves on the cue which is removed on transfer, the effect is to have distracted the S from the cue which is required for solution in the hard problem. If this is so, no transfer for this S would occur. The extra relevant cue during original learning would thus serve as a counter-emphasizer with respect to the relevant cue of the hard problem.

If the foregoing analysis is correct, optimal easy-to-hard transfer would be obtained by using what may be termed a "pure" emphasizez. A pure emphasizez is one which increases the probability of S attending to the relevant cue, but at the same time, is not a basis for a correct strategy. The emphasizez when removed on transfer would not interfere with transfer since it is not a basis for a correct strategy.

A stimulus emphasizez might, in some cases, be a relevant and redundant cue. If so, the rôles might compete; the S would attend to the emphasizez as a cue rather than to the relevant dimension being emphasized.

The analysis of the literature suggests that attention in learning can be studied objectively. Aspects of the stimulus situation can be arranged so that they control or at least affect the locus of attention. This possibility affords a powerful and reasonably precise independent variable for investigation. The speed of learning is an indicator of the probability that a relevant dimension is attended to, and serves as an index of an attempt to direct the S's attention. When a problem involves more than one relevant cue, or more than one irrelevant cue, or a relevant cue which can be diminished or enlarged, measurements of transfer-of-training can be used to gain more detailed information regarding the distribution of attention over the parts of the stimulus situation. With objective and practical independent and dependent variables, the main questions are open to experimental investigation. This dissertation reports the results of such a theoretical and experimental analysis of attention in a simple concept formation task.

CHAPTER II

THE STRATEGY SELECTION THEORY

In the present study, a number of flower designs are presented to the S in a series. On each trial, the S is required to make one of two responses and then is told the correct choice. Each pattern is complex and the correctness of the S's response depends upon one or two aspects of the pattern. The S must, in each instance, try to determine which aspect or aspects of the patterns are relevant.

To resolve the discrimination, the S might use "strategies" or hypotheses which are based upon some stimulus aspect. A strategy is any consistent pattern of behavior to the cues in the situation. The stimulus patterns give rise to a number of strategies, some of which conflict. The S has difficulty in solving whenever he uses strategies which conflict with the strategy intended by the experimenter.

If all patterns were presented at once, the S could make systematic tests of strategies in order to discover the relevant aspect or aspects.

Since in a serial presentation only one pattern is present at the time of reinforcement, the S must remember what has been presented and what was reinforced.

In many tasks, the capacity of the S to remember aspects is exceeded. One approach is to assume that S can remember only the strategy or strategies that he is using. If the strategy is wrong, he then shifts to a new strategy. The shift from one strategy to another is assumed to be random, permitting the S flexible rather than stereotyped responding.

The concept that S uses strategies is similar to the attentional views of Lashley (1929) and Krechevsky (1932) on the learning process. The S selects a strategy and makes the indicated response. If the response is rewarded, then the same strategy is used again on the next trial. If the strategy is in error, then the S chooses at random from the set of strategies available to him in the problem.

Restle (1961a; 1961b) has worked out a number of consequences of these ideas where sampling of strategies is with replacement. Restle's theory is referred to as the Strategy Selection Theory. This model, as explicitly stated by Restle, determines the detailed stochastic properties of the data and gives a rational theory of "learning rate." We shall proceed with the case where S is assumed to use only one strategy per trial.¹

When applied to the two-choice concept formation task, the theory postulates a total set of strategies, H , of which a subset C always lead to a correct response, a subset W always lead to a wrong response and the remainder I lead to correct and wrong responses half the time at random. Let c , w and i represent the proportion of each type and $c + w + i = 1$. The speed of learning depends upon the probability of selecting a correct strategy, c .

Stochastic Properties of the Data

The S begins the task in a "starting state" where he chooses a strategy with probabilities:

c , it is a correct strategy

¹Restle (1961a) has shown that S may (a) use only one strategy per trial, (b) use all strategies at once and attempt to narrow down to a correct one, or (c) use a random sample of all strategies and attempt to narrow down his sample, and each approach leads to the same set of probabilities in acquisition.

w, it is a wrong strategy and

i, it is an irrelevant strategy.

If S samples a correct strategy, he is in the "solution phase" and makes no more errors. If he samples either a wrong or an irrelevant strategy, he is in the "pre-solution phase" and will make at least one more error.

A wrong strategy always leads to errors whereas an irrelevant strategy leads to an error half the time. Each time he makes an error, he is returned to the starting state, where he resamples (with replacement) a strategy with probabilities c , w , and i that it is correct, wrong or irrelevant. As long as S is in the pre-solution phase (using wrong or irrelevant strategies) he shows no net gain or improvement since he will always make at least one more error and return to the starting state. The sequence of correct and wrong responses is "stationary" in the sense that the probability of an error does not change during the pre-solution phase.

This analysis of S's behavior is illustrated by a "tree" diagram in Fig. 2.1.

A second stochastic property of the model is that the transition probability, c , from the pre-solution to the solution phase, is independent of the number of trials S has been in the pre-solution phase.

Let T_a be the total errors made by subject a . Since the probability of no more errors is c , the probability of exactly k errors is $\Pr(T_a = k)$ which is $\Pr(\text{at least one error}) \cdot \Pr(\text{at least one more error}) \dots (k \text{ times}) \dots \Pr(\text{no more errors}) =$

$$(1-c)^k (c), \text{ a geometric distribution.} \quad (2.1)$$

The mean of a geometric distribution of error scores is

$$E(k) = \bar{T} = (1-c) / (c) \quad (2.2)$$

and the variance is

$$\text{var}(k) = (1-c) / c^2. \quad (2.3)$$

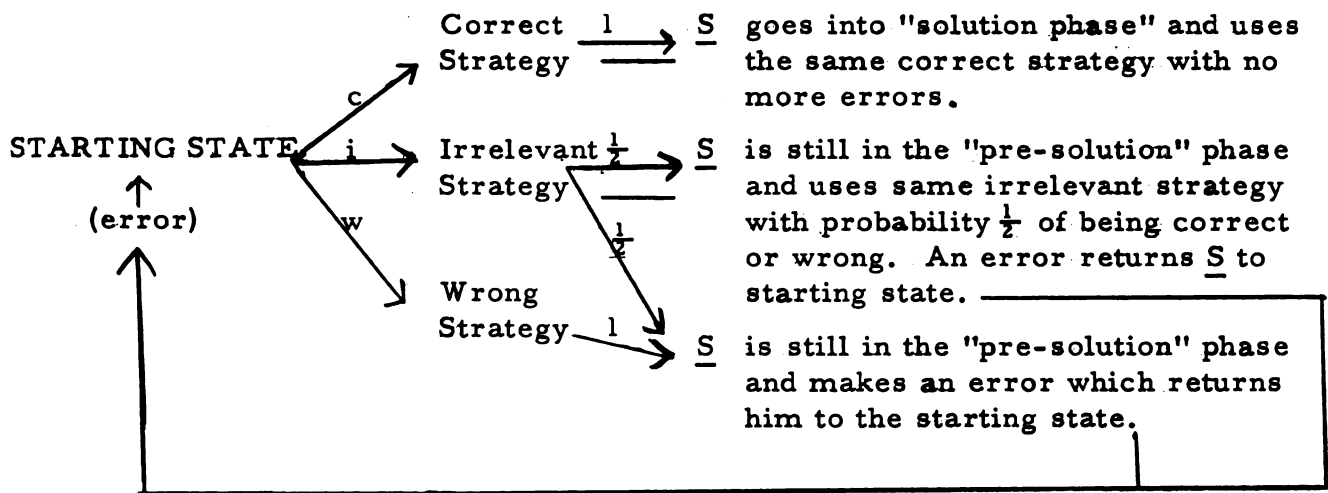


Figure 2.1. Stochastic Structure of Strategy Selection.

Compared with the normal distribution of error scores, a geometric distribution is positively skewed and its mean is positively correlated with its variance. To illustrate the shape and characteristics of the geometric distribution, consider a problem of obtaining the first head in tossing a coin. If the coin is fair, the probability of a head is $\frac{1}{2}$ on each trial. On trial 1, the probability is $\frac{1}{2}$ of obtaining a head and $(1-\frac{1}{2})$ of a tail. The conditional probability of the last tail on trial 1 is $(1-\frac{1}{2})(\frac{1}{2})$. The conditional probability of the last tail (error) on trial 2 is $(1-\frac{1}{2})(1-\frac{1}{2})(\frac{1}{2})$, etc. The probability of obtaining k tails is, in general, $(1-\frac{1}{2})^k(\frac{1}{2})$, a geometric distribution. The expected number of tails before the first head (Eq. 2.2) is $(1-\frac{1}{2})/(\frac{1}{2}) = 1$ and the variance (Eq. 2.3) is $(1-\frac{1}{2})/(\frac{1}{2})^2 = 2$.

A geometric distribution of the number of tails until the first head occurs is shown in Fig. 2.2 for six tosses.

In the present experiment, learning scores are expected to be geometrically distributed. Fig. 2.3 compares a hypothetical geometric distribution of error scores with one that is normal.

Sets of Strategies and Parameters of Learning

According to the theory, the probability of solving the problem on any trial is c , the proportion of correct strategies. Correct and wrong strategies depend upon stimulus aspects which vary and are relevant during training. Irrelevant strategies depend upon those aspects which are fixed or vary and are uncorrelated with reward, and on other factors such as outcomes of previous trials, boredom, etc.

The relevant and most of the irrelevant strategies in the present experiment arise from flower patterns. Elementary notions of set theory can be applied to the stimulus situation to permit

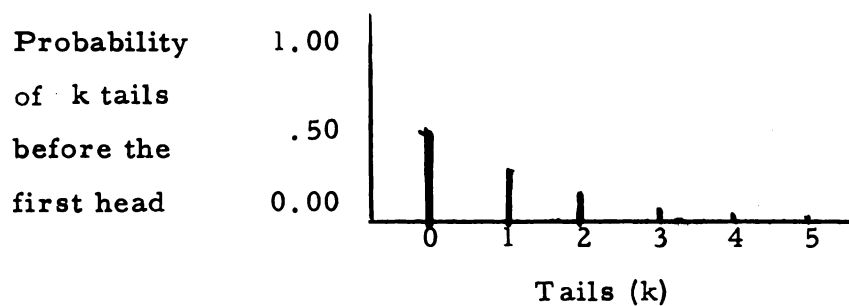


Fig. 2.2. A geometric distribution showing the probability of obtaining k tails before the first head in tossing a fair coin six times.

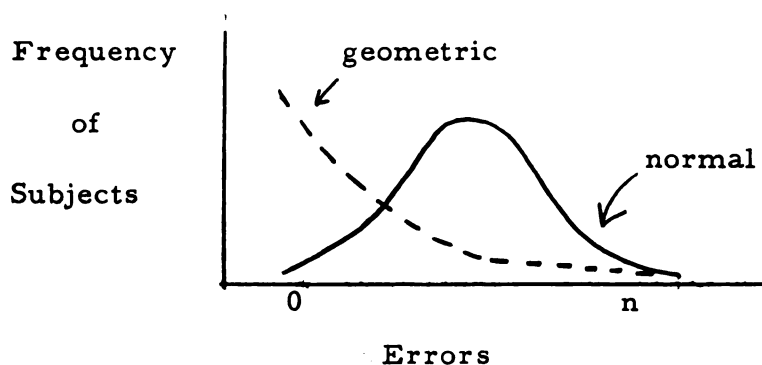


Fig. 2.3. Hypothetical geometric and normal distributions of error scores in a learning experiment.

1. a logical analysis of the problem in terms of the strategies present and

2. estimation of the measure of each set which S_s use in solving.

The measure of a set reflects the set's "weight" in influencing the direction of the S 's attention and his speed of learning.

Suppose that the strategies which arise from the angle of the leaves to the stem of the flower are correct. Such strategies from the angle dimension are called A . The number of strategies in the set A is written $m(A)$ and is called its "measure."²

If S samples from a set of H total strategies, with measure, $m(H)$, then the probability that the strategy chosen is in the correct set A is

$$P(A) = m(A)/m(H). \quad (2.4)$$

If A is the only set of strategies which is correct, the rate of learning parameter, c is the probability of choosing a strategy in A . Therefore,

$$c = P(A) = m(A)/m(H). \quad (2.5)$$

The total strategy set, H , may be subdivided into subsets which arise from dimensions of the stimulus pattern. Suppose that in the pattern there are subsets A (correct angle strategies), A^* (wrong angle strategies), L (leaf strategies) and I (all other irrelevant strategies).

The set H may be written as the union (\cup) of the subsets,

$H = (A \cup A^* \cup L \cup I)$ and $m(H) = m(A \cup A^* \cup L \cup I)$, which is read "the measure of the set of H strategies which contains all the strategies in the sets A , A^* , L , and I and in all combinations of the subsets."

If the subsets are assumed to be disjoint(have no common strategies), then

²Elementary set theory, its rules and relationship to probability theory may be found in Feller (1950), Kemeny, Snell and Thompson (1957) or Restle (1961c).

$$m(H) = [m(A) + m(A^*) + m(L) + m(I)].$$

The probability of selecting correct strategies in A is now

$$c = P(A) = m(A)/[m(A) + m(A^*) + m(L) + m(I)]. \quad (2.6)$$

Eq. 2.6 is a basic equation in the analysis of acquisition and it is used as the cornerstone for the following theoretical development.

Acquisition

Analysis of the research on what makes a problem either easy or difficult (Chapter I) indicated that the probability c of \underline{S} hitting upon a correct strategy can be influenced by

1. A removal of irrelevant dimensions.
2. An increase in the number of relevant and redundant dimensions.
3. An increase in the salience of a relevant dimension by either
 - a. emphasizing it in some way or
 - b. making it a larger cue.

1. A removal of irrelevant dimensions. Suppose that a problem is constructed with the set of strategies in A relevant and all others irrelevant. The probability of selecting strategies in the set A is given by Eq. 2.6, above.

If the irrelevant dimensions of the leaves are removed by holding the leaves constant from trial to trial, all strategies based on aspects from the leaves are removed. Then only the sets A and A^* and I are present, so that the new value of c is

$$c = m(A)/[m(A) + m(A^*) + m(I)]. \quad (2.7)$$

Learning of this problem should be faster since the denominator of Eq. 2.7 has been reduced by the removal of the irrelevant set L and the numerator is unchanged.

Suppose one wished to estimate $m(A)$ in this experiment. Let the estimate of c (written \hat{c}) be .05. Since the sets A and A^* both depend on the same relevant dimension, let $m(A) = m(A^*)$. From Eq. 2.7,

$$\hat{c} = .05 = m(A)/[2m(A) + m(I)], \text{ or rearranging terms,}$$

$$\begin{aligned} m(A) &= .05 [2m(A) + m(I)] \\ &= .10 m(A) + .05 m(I). \end{aligned}$$

Solving for $m(A)$,

$$m(A) = \frac{.05}{.90} m(I) = .056 m(I).$$

The weight of the set, $m(A)$, is estimated relative to $m(I)$, the number of irrelevant strategies in the problem.³

2. An increase in the number of relevant dimensions. Now consider a problem where both A and L are sets of correct strategies, with A , A^* , L , L^* , and I present. The \underline{S} may solve on strategies in A or in L and the sets are assumed disjoint. The probability of hitting upon a correct strategy in A or in L is

$$c = P(A \cup L) = [m(A) + m(L)]/[m(A) + m(A^*) + m(L) + m(L^*) + m(I)]. \quad (2.8)$$

Comparing Eqs. 2.6 and 2.8, c in 2.8 is larger, provided c is less than $\frac{1}{2}$, since $m(L)$ is added and the numerator has been increased relatively more than the denominator. Learning is faster when two cues are relevant and redundant according to the theory. This represents a case of "additivity of cues" (Restle, 1955; Trabasso, 1960).

3. An increase in the salience of a relevant dimension by

³The measures of sets of strategies can be determined only up to an arbitrary unit of measurement. The data only give estimates of probabilities (ratios of measures). Multiplying all measures by a constant, K , would have no effect on the theoretical predictions, since when ratios are formed, K would cancel out. Hence $m(A)$ can be specified only relative to the measure of some standard set. It is convenient to use $m(I)$ as a standard since the experiment is designed so that the set of irrelevant strategies, I , is common to all problems.

a. emphasizing it in some way. The measure of a set, $m(A)$, is the saliency measure or attentional value of the set A . The attention value is represented by the measure of the set instead of the probability, since the probability of attending to a particular aspect depends both on its attention value (measure) and on the number and attention values of other aspects in the situation. If the set is emphasized in a problem, its measure, $m(A)$, is multiplied. Multiplication increases the measure of a set without changing its quality.

One way of making a dimension salient is to use another stimulus to emphasize it. Suppose that a problem is constructed with the set A relevant, but the dimensions in A are colored red so that they "stand out" on the pattern. If the color emphasizes A (multiplier $r > 1$), the probability of selecting strategies in A is now

$$c = P(A) = r.m(A)/[r.m(A) + r.m(A^*) + m(L) + m(I)]. \quad (2.9)$$

As in Eq. 2.8, the rate of learning, c , is increased over the rate of Eq. 2.6 where A is not emphasized, provided $c < \frac{1}{2}$ and $r > 1$.

3. An increase in the salience of a relevant dimension by making it

b. a larger cue. If the difference between discriminanda is increased, the effect is two-fold: (1) new strategies, B , are added by the change of the stimulus qualities, and (2) the strategies which already existed are emphasized. Suppose that the angle difference in the problem is doubled. Then the new measure of angle strategies is written $d.m(A) + m(B)$, where $d > 1$. The probability of selecting a correct strategy in A or B is now

$$c = P(A \cup B) = \frac{d.m(A) + m(B)}{[d.m(A) + d.m(A^*) + m(B) + m(B^*) + m(L) + m(I)]} \quad (2.10)$$

The rate of learning in this problem would be faster than the one in Eq. 2.6, provided $c < \frac{1}{2}$, since the numerator is increased relatively more than the denominator.

Transfer

A general hypothesis of this thesis is that the degree of transfer depends upon the stimulus relationship between the training (problem 1) and transfer (problem 2) situations. The hypothesis may be written as a conditional probability of solution on problem 2 after solution on problem 1. The proportion of Ss who transfer perfectly after mastery of problem 1, with the set of correct strategies C_1 , to problem 2, with the set of correct strategies C_2 is

$$P(C_2/C_1) = m(C_1 \cap C_2)/m(C_1). \quad (2.11)$$

The set $C_1 \cap C_2$ is the "intersection" (\cap) of the sets, C_1 and C_2 , and denotes the common strategies.

If the S is using one strategy at a time, solution of problem 1 would ensure that the strategy being used is in C_1 . If $C_1 = C_2$ (the problems have the same relevant dimensions), then Eq. 2.11 equals 1.00. Transfer in this case is perfect for all Ss.

Suppose that problem 1, the training problem, has two redundant sets of correct strategies, A and L , but on transfer L is removed. After problem 1 is learned, it is known that S is using a strategy in the set $C_1 = (A \cup L)$. The conditional probability that the strategy is in the set A , where $C_2 = A$, is

$$P(A/A \cup L) = m(A)/[m(A) + m(L)]. \quad (2.12)$$

Transfer from problem 1, where two redundant sets of correct strategies are present, might vary according to the emphasize rôle.

If A is emphasized during training on problem 1, Eq. 2.12 becomes

$$P(A/A \cup L) = r.m(A)/[r.m(A) + m(L)], \quad (2.13)$$

and more Ss are expected to transfer.

If A is counter-emphasized by coloring another set L , Eq. 2.12 now becomes

$$P(A/A \cup L) = m(A)/[m(A) + r.m(L)], \quad (2.14)$$

and fewer Ss show perfect transfer.

In summary, Eqs. 2.6 to 2.14 formulate, in mathematical terms, the hypotheses regarding stimulus emphasis and the relationships of original learning (easy) to transfer (hard) problems. Eq. 2.6 is the basic formula for c , the probability of solving the original learning problem, and is developed in Eqs. 2.7 to 2.10 to apply to the various acquisition problems. Eq. 2.11 is the basic formula for transfer of training, and is developed in Eqs. 2.12 to 2.14 to apply to experimental variations on transfer. These formulas summarize the hypotheses to be tested by experiment.

Estimation of the Parameter c ⁴

When learning is complete (all Ss solve), the maximum-likelihood estimator of c is

$$\hat{c} = 1/[\bar{T} + 1] \quad (2.15)$$

and its variance

$$\text{var}(\hat{c}) = \hat{c}^2(1-\hat{c})/N \text{ (Restle, 1961a)} \quad (2.16)$$

where \bar{T} is the mean total error score and N is the number of Ss.

In the present experiment, a learning criterion of 10 successive correct responses and a fixed number of trials are used. Some Ss might fail to solve within the allotted time. These non-solvers produce what is known as a "censored" distribution of error scores. The estimate of c must be modified in order to take into account the presence of non-solvers in a group.

Consider a group of N Ss who make a total of T errors. Of these, N_s are observed to reach learning criterion within the fixed number of

⁴This section is technical and the general reader may skip to the next chapter (Method) if he so wishes without any loss of information.

trials, making a total of X_1 errors, and the other $N - N_s$ \underline{S} s make a total of X_2 errors. The likelihood of such an outcome as a function of the possible values of the parameter c is

$$L(\text{set of data with } X_1 \text{ errors, } N_s \text{ solvers; } c) = (1-c)^{X_1} c^{N_s}, \text{ and} \\ (\text{set of data with } X_2 \text{ errors, } N - N_s \text{ non-solvers; } c) = (1-c)^{X_2}.$$

The joint likelihood of solvers and non-solvers is

$$L(c) = [(1-c)^{X_1} c^{N_s}] = [(1-c)^T c^{N_s}]. \quad (2.17)$$

The value of L is to be maximized with respect to c . We shall maximize $\log(L)$.

$$\log(L) = [T \cdot \log(1-\hat{c}) + N_s \cdot \log(\hat{c})].$$

Taking the derivative of $\log(L)$ with respect to c , and setting it equal to 0,

$$\frac{d \log(L)}{d(c)} = 0 = \left[\frac{-T}{(1-\hat{c})} + \frac{N_s}{\hat{c}} \right].$$

Solving for \hat{c} ,

$$\hat{c} = N_s / [T + N_s]. \quad (2.18)$$

If all \underline{S} s solve, then $\hat{c} = 1/\bar{T} + 1$, which is the maximum-likelihood estimator of c obtained by Restle (1961a) and given in Eq. 2.15.

It can be shown that the derived maximum-likelihood estimator of c in Eq. 2.18 is biased and an unbiased estimate of c is

$$\hat{c} = (N_s - 1) / [\bar{T} + N_s - 1] \quad (2.19)$$

However, in the present study, Eq. 2.18, the maximum-likelihood estimator, is used to permit statistical tests of hypotheses such as likelihood ratio tests.

Variance of c

With large samples of \underline{S} s, the sampling variance of a maximum-likelihood estimate (provided the estimates are normally distributed) can be calculated by the formula,

$$\text{var}(c) = \frac{-1}{E \left[\frac{d^2 \log(L)}{d(c)^2} \right]}, \text{ which in the present case is}$$

$$\text{var}(c) = \frac{-1}{E \left[\frac{-T}{(1-c)^2} + \frac{N_s}{c^2} \right]}.$$

In a censored distribution with number of trials = 64,

$$\text{var}(c) = [c^2(1-c)] / N [1 - (1-c)^{32} - 33c(1-c)^{32} + 32c(1-c)^{33} + 32c(1-c)^{31}].$$

This is very close to

$\text{var}(c) = [c^2(1-c)] / N_s$ and is approximated by substituting \hat{c} for c so that our estimate of the variance of c is

$$\text{var}(\hat{c}) = [\hat{c}^2(1-\hat{c})] / N_s. \quad (2.20)$$

Estimation of w , the Proportion of Wrong Strategies

A high frequency of consecutive errors in the pre-solution phase is an indication of a relatively large proportion of wrong strategies, whereas a small frequency of consecutive errors is an indication that there are relatively few wrong strategies. To estimate w , the proportion of wrong strategies, one counts the "trial zero" (initial starting state) as an error. Then for each \underline{S} , the number of errors which are followed by correct responses M_0 , and the number of errors which are followed by errors, M_1 , are obtained. Then averaging $\sum M_0$ and $\sum M_1$ for the group, all \underline{S} s all \underline{S} s

$$\hat{w} = (\bar{M}_1 - \bar{M}_0 + 1) / (\bar{M}_0 + \bar{M}_1) \quad (2.21)$$

which is a corrected maximum likelihood estimate of w (Restle, 1961b, p. 299).

CHAPTER III

METHOD

Subjects

The Ss were 215 students in elementary psychology courses at Michigan State University who received credit for participation in experiments. Ten groups, each composed of 11 or 12 men and 9 or 8 women, were formed by assigning Ss in a haphazard, pre-arranged order. Fifteen Ss failed to solve their original learning problems within the allowed 64 trials. Each group with non-solvers was supplemented by more Ss until 20 solvers were obtained, so that 20 Ss, all of whom had solved the original learning problem, were available for testing on the transfer problem.

Apparatus

From S's point of view, the apparatus consisted of a 2 x 2 ft. black screen with a centered 4 x 5 in. window. Stimulus patterns were presented in the window on a hinged card holder. S classified the patterns by saying aloud "A" or "B" while E recorded each pattern and response. S self-paced his responses. Correct pattern classifications were signaled by a lighted letter, A or B, located 2 in. below and to the left and right of the window, respectively. Duration of reinforcement was about 2 sec. and patterns were removed from view approximately 3 sec. after the light was turned off. S sat facing the screen with his head approximately 20 in. from the window. Except during the reading of instructions, E was shielded from S's view throughout the course of the experiment.

Procedure

The same instructions were read to all groups. Each S was told that the study intended to find out how college students form concepts. S's job was to divide correctly a set of cards into two categories, A and B. The cards could be classified on the basis of a simple principle. Each card would contain a different figure but all the cards within a category have something in common. S could take as much time as he wished to say aloud his classification, either A or B. After he responded, one of two letters would light, telling him the correct class of the card, A or B. Each time, he was to remember the basis upon which he classified the card and whether or not he was correct. Guessing of the response sequence was discouraged by informing S that the cards were shuffled.

Ss were told that two problems would be given. The first, a practice one, would be unrelated to the second (original learning-transfer). On the practice problem, each card would contain a single geometric figure. After S solved the practice problem, he was told that all cards with triangles were A's and those with circles were B's. After the correct solution for the practice problem was stated, the second problem was introduced. In the second problem, the patterns would be floral designs, each card bearing a flower, stem and leaves. Instructions regarding classification and use of a simple principle were repeated.

The practice problem consisted of a sequence of geometric forms and was given to ensure that the procedure was understood and to orient the S to the task. In the problem, form (circle vs. triangle) was relevant and color (black vs. white) and size (large vs. small) were irrelevant. Single stimulus figures were drawn in India ink on white 3 x 5 in. file cards. Large and small figures could be inscribed within 2 and 1 in. squares, respectively. The stimulus deck contained eight different cards, representing all combinations of the three dimensions. Cards were

shuffled before and during learning at the end of 8-trial blocks until S reached a criterion of 8 successive correct responses.

After practice, each S received an original learning problem for 64 trials or until 10 successive correct responses were given, whichever occurred first. Those Ss who solved the problem were continued without interruption to the transfer problem which was the same for all groups. One group learned the "transfer" problem as its original learning problem and served as a control; all other groups had different original learning problems. Comparison of each group with the control constituted an example of the transfer-of-training paradigm, $X \longrightarrow Y$ versus Y .

Selection of the Learning Problem

A two-choice concept formation task was designed to meet several considerations:

1. Two distinct response choices in order to prevent response generalization.
2. A completely specified set of stimuli with some degree of complexity and spatial separation of parts within each pattern.
3. Continuous variation of a relevant dimension above psycho physical threshold.
4. Variation in level of difficulty by addition of relevant or irrelevant dimensions.
5. A distinctive dimension (color) as a stimulus emphasize.
6. An intrinsically interesting and plausible problem.

To meet the above requirements, flower patterns were abstracted from those published by Hovland (1953). These proved to be satisfactory in a pilot study.

Stimuli

The stimuli in all experimental problems were flower patterns. Figure 3.1 shows four examples used in original learning and transfer problems.

To ensure uniformity of the drawings, separate templates of the main dimensions were used. Patterns were drawn in India ink on white 3 x 5 in. file cards. Each figure could be inscribed within a 1 x 3 in. rectangle. The base was 1 in. long, stems were 2 or $2\frac{1}{2}$ in. high and leaf stems were $\frac{3}{4}$ in. long. The vertical angle of each of the three leaves to the stem was the same on any card.

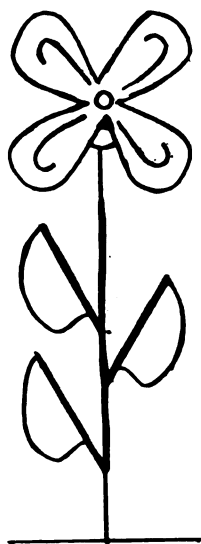
The five dimensions of the patterns were:

1. The angles of the leaves to the stem was relevant and had two values, either 30° versus 60° or 15° versus 75° . When the angle dimension was constant, all angles equalled 45° .
2. The flower dimension had four values: tulip, daisy, pansy, and fleur-de-lis.
3. The leaf shape dimension had two values: smooth or scalloped.
4. The leaf position dimension had two values: 2 left and 1 right or 1 left and 2 right.

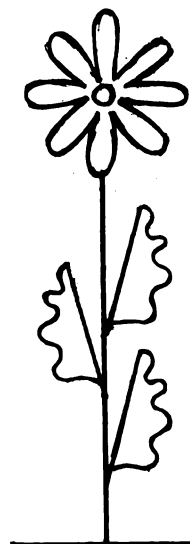
In nine problems, the flower, leaf shape and leaf position dimensions were varied independently of one another and were irrelevant. On a tenth problem, they were fixed and only the angle dimension, which was relevant was varied (30 vs. 60 degrees).

5. Color was used as an "emphasizer" or "counter-emphasizer" with respect to the angles. When color varied, it had two values: red versus green.

A different order of presentation was given to each S by shuffling the deck before each session. Cards were reshuffled at the end of 32 trials of each problem, if S had not yet reached criterion.

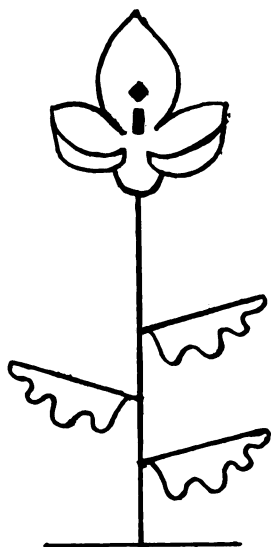


A: 30°



B: 15°

C: 75°



D: 60°



Fig. 3.1 Examples of flower patterns used in experimental problems (actual size). The degrees indicate the size of the angle of the leaves to the stem.

Experimental Groups

Group Angle, which served as a control for acquisition and transfer, had a problem with angle relevant and flower, leaf shape and leaf position irrelevant. No color appeared in this problem. Patterns with 30° angles were A's and those with 60° angles were B's. This same Angle problem was the transfer problem for all groups.

Groups Red Angle and Green Angle had the Angle problem but the angles and stem included were always colored: red in the Red Angle problem and green in the Green Angle problem. Color did not vary from trial to trial and could not be used to solve the problem directly. The constant color thus served as a "pure" emphasize. Since color was spatially contiguous with the relevant angles, it was intended to make the angles "stand out" and emphasize that relevant dimension.

Group Angle + Angle Color had the Angle problem but either red or green appeared on the angles. The A patterns had 30° red angles and the B patterns had 60° green angles, making angle and color relevant and redundant. S could learn that A was either a 30° angle or a red angle, and that B was either a 60° angle or a green angle. Color was presumed to be an emphasize and a redundant relevant dimension.

Group Angle Color Control had a problem with angles all at 45° during original learning. Patterns with red angles were A's and green angles were B's, making color relevant. Correct responses depended upon the color of the angle.

Group Angle + Flower Color had the Angle problem but either red or green appeared on the flowers. The A patterns had 30° angles and red flowers and the B patterns had 60° angles and green flowers, making angle and color relevant and redundant. S could learn either that A was a 30° angle or red flower and that B was either a 60° angle or green flower. Color was presumed to be a counter-emphasizer with respect to the angle dimension, and also a redundant relevant dimension.

Group Flower Color Control had a problem with all angles fixed at 45° during original learning. Patterns with red flowers were A's and patterns with green flowers were B's, making color relevant. Correct responses depended upon the color of the flower.

Group Angle + Angle Color Irrelevant had the Angle problem but either red or green appeared on both angles. Ss could not use the irrelevant color to solve the problem, but color was spatially contiguous with the angles and was intended to serve as an emphasize.

Group Large Angle had the Angle problem but the angle difference was increased from 30 to 60 degrees. Patterns with 15° angles were A's and those with 75° angles were B's. The larger angle difference was intended to make the relevant angle dimension "stand out" and serve as another kind of emphasize.

Group Angle Only had a problem where all the irrelevant dimensions of the Angle problem were fixed. The 30° angles card was an A and the 60° angles card was a B. The removal of all the irrelevant dimensions by fixing them was intended to draw the attention of the S to the relevant angle and make the problem easier to learn.

Table 3.1 summarizes these ten experimental groups, and describes the problems in detail.

Data

The data used in the individual comparisons (Chapter V) are total errors made by individual Ss in reaching learning criterion or within 64 trials, whichever occurred first. For comparisons of acquisition during original learning, the data are from the first 20 Ss in each group. For comparisons on transfer, the data are from the 20 Ss who solved their original learning problems and were transferred to the Angle problem.

Table 3.1. Experimental Groups and Problems

Group	Theoretical Significance	Relevant Dimensions	Irrelevant Dimensions	Number of Patterns
Angle	Control	Angles (30-60)	Flower, leaf shape and position	32
Red Angle	Constant Emphasizer	Angles (30-60)	Flower, leaf shape and position	32
Green Angle	Constant Emphasizer	Angles (30-60)	Flower, leaf shape and position	32
Angle + Angle Color	Emphasizer Relevant	Angles(30-60) Color(red-green)	Flower, leaf shape and position	32
Angle Color Control	Emphasizer Relevant Control	Color(red-green)	Flower, leaf shape and position	32
Angle + Flower Color	Counter-Emphasizer Relevant	Angles(30-60) Color(red-green)	Flower, leaf shape and position	32
Flower Color Control	Counter-Emphasizer Relevant Control	Color(red-green)	Flower, leaf shape and position	32
Angle + Angle Color Irrelevant	Emphasizer Irrelevant	Angles (30-60)	Flower, leaf shape and position	64
Large Angle	Lawrence Easy-Hard	Angles (15-75)	Flower, leaf shape and position	32
Angle Only	Irrelevant Cues Constant	Angles (30-60)	none	2

Statistical Tests

In the experiment, mean errors are asymptotically normal with large numbers of \underline{S} s so that individual comparison \underline{t} -tests are used.

The program described in Chapter II is feasible only if reliable estimates of the probability of sampling given sets of strategies can be obtained. Predictions on the rate of learning and transfer are to be made (Chapter VI), so it is desirable to make them as accurately as possible and statistically test any discrepancies between the predicted and obtained results.

To test the hypotheses:

1. $c = c_0$ or
2. $c_1 = c_2$

the theory offers likelihood ratio tests which use the maximum-likelihood estimates of c . In a likelihood ratio test one maximizes the likelihood over a restricted parameter subspace (the null hypothesis) and also over the entire space of logical possibilities. The ratio of these two likelihoods is called λ . With large samples, the value of $-2 \ln(\lambda)$ is distributed approximately as χ^2 , provided the null hypothesis is true. The degrees of freedom are equal to the number of free parameters (see Restle, 1961b, pp. 301-304).

If \hat{c} is assumed normally distributed with $\text{var}(\hat{c}) = \hat{c}^2(1-\hat{c})/N_s$, where N_s is the number of \underline{S} s who solve, then the difference between c_1 and c_2 , if independent, is normally distributed with variance equal to $\text{var}(c_1) + \text{var}(c_2)$. Normal distribution \underline{z} -tests of differences between two parameters are made in Chapter V.

CHAPTER IV

STOCHASTIC PROPERTIES OF THE DATA AND TEST OF THE STRATEGY SELECTION THEORY

According to the Strategy Selection Theory, the responses of an S are composed of a sequence of correct and wrong responses in an irregular order (the "pre-solution phase"), followed by an infinite sequence of correct responses (the "solution phase"). In the experiment, 10 correct responses in a row represented the learning criterion and constituted the "solution phase." Some Ss failed to solve and consequently had no solution phase.

The model may be said to describe the data well if

1. the responses before the last error do constitute a statistically stationary sequence of correct and wrong responses and
2. transition from the pre-solution to the solution phase occurs with a probability which is constant and independent of how long S was in the pre-solution phase. These two questions about the stochastic structure of the data can be answered by considering, first, behavior during the pre-solution phase, and second, the distribution of errors made in the pre-solution phase.

Analysis of the Pre-solution Data

The pre-solution phase contains all the trials before the last error observed. This last error may be followed by a criterion run of 10 consecutive correct responses or, in the case of non-solvers, may be the error at which training was terminated by E. Let the trial of last error for subject a be m_a . Let $X_{a,n}$ be a random variable which takes on

the value 1 if S makes an error on trial n and 0 otherwise. Then the proportion of errors during the pre-solution phase is

$$p_a = \frac{\sum_{i=1}^{m_a-1} X_{a,i}}{(m_a-1)} . \quad (4.1)$$

If w , the proportion of wrong strategies, is small, then the proportion of errors during pre-solution trials (p_a) should be near $\frac{1}{2}$ for all Ss, and in fact should have a binomial distribution,

$$\Pr \left\{ p_a = k / (m_a - 1) \right\} = \binom{m_a - 1}{k} \left(\frac{1}{2} \right)^{m_a - 1} . \quad (4.2)$$

To obtain a statistical test, approximate the binomial distribution in Eq. 4.2 by a normal distribution with mean = $\frac{1}{2}$ and variance

$$pq/N = \left(\frac{1}{2} \right)^2 / (m_a - 1) . \quad \text{With this approximation, the statistic}$$

$$D^2(a) = 4(m_a - 1)(p_a - \frac{1}{2})^2 \quad (4.3)$$

has a χ^2 distribution with 1 degree of freedom. $D^2(a)$ was computed for each S with 2 or more errors and the values summed for each group and then over all groups. These pooled values should have an approximately χ^2 distribution with df corresponding to the number of Ss involved.

In original learning, 163 Ss made 2 or more errors and the pooled $D^2 = 147.32$. Since the value is less than the degrees of freedom, it is not significant. In transfer, 76 Ss made at least 2 errors and the pooled $D^2 = 63.94$, again less than the df.

No statistically significant deviation of individual error proportions from $\frac{1}{2}$ during the pre-solution phase was observed, a result in accord with the strategy selection theory. Furthermore, pooled values of D^2 were not significant for any one of the 10 original learning groups and the 4 transfer groups, which contained Ss with 2 or more errors.

If the data showed gradual improvement in the probability of a correct response during the pre-solution phase, then the proportion of

errors should be less than .50. In fact, the observed mean proportion of errors was .55 for original learning and .51 for transfer pre-solution trials. Since these values were numerically larger than .50, Ss may have used wrong as well as irrelevant strategies (see below). In any case, there is no indication, either numerical or statistical for a gradual elimination of errors during the pre-solution trials.

Although the hypothesis that $p_a = \frac{1}{2}$ could not be rejected, Ss could have made disproportionately more errors early or late in the pre-solution phase. A test of stationarity, whether or not the probability of an error went up or down during the pre-solution phase, was made. If the probability of an error was constant, then error frequencies during halves of the pre-solution trials should be equal for each S. To make a statistical test, each S's pre-solution trials, $m_a - 1$, was divided in half, discarding the middle trial where $m_a - 1$ was odd, and errors per half were counted. For Ss who had at least 2 errors remaining, errors in the first half were compared with those in the second half by a sign test. Table 4.1 shows proportions of Ss making more, as many or fewer errors during the first half pre-solution trials compared with the second half.

Table 4.1. Test of Stationarity: 1st versus 2nd Half Errors in the Pre-solution Phase.

Condition	Proportion of Ss Making More, As Many or Fewer 1st Half Errors			Number of Ss
	More	As Many	Fewer	
Original learning	.27	.36	.37	158
Transfer	.38	.21	.41	72

First half errors did not differ significantly from second half errors, either during original learning ($z = -1.59$) or transfer ($z = -0.13$) at the .05 level. These results indicate that the probability of an error during the pre-solution trials did not change, and support the theory.

Whether learning is continuous or discontinuous during the pre-solution phase can be shown by construction of a "backwards learning curve" for each group. If the individual learning curve is discontinuous, the backward group learning curve shows performance at a chance level and a sharp jump to better than chance performance (Hayes, 1953).

Since criterion was not reached by all Ss, such curves for each experimental group could not be constructed. However, if the probability of an error is constant and independent of how long the S is in the pre-solution phase, the proportion of errors made by Ss who are still in the pre-solution phase should be constant and near .50 over trials. The proportions were obtained by forming, for each trial k , a ratio of total errors on trial k to the number of Ss who were still in the pre-solution phase on trial k . The probability of an error, conditional on S being in the pre-solution phase, is given for original learning (A) and transfer (B) in Fig. 4.1.

Both curves in Fig. 4.1 appear to be reasonably flat over the 70 trials where errors were made, even where the S frequencies grew small. Variation of the proportion was nearly all contained within the .40 to .60 range and the fluctuation centered around .50. If gradual learning occurred during the pre-solution phase, these curves would diverge from .50 toward zero over trials.

Errors Before Solution

The second aspect of the Strategy Selection Theory is the assumption that the probability of moving from the pre-solution to the solution phase is a constant, c , on any trial the S responds erroneously.

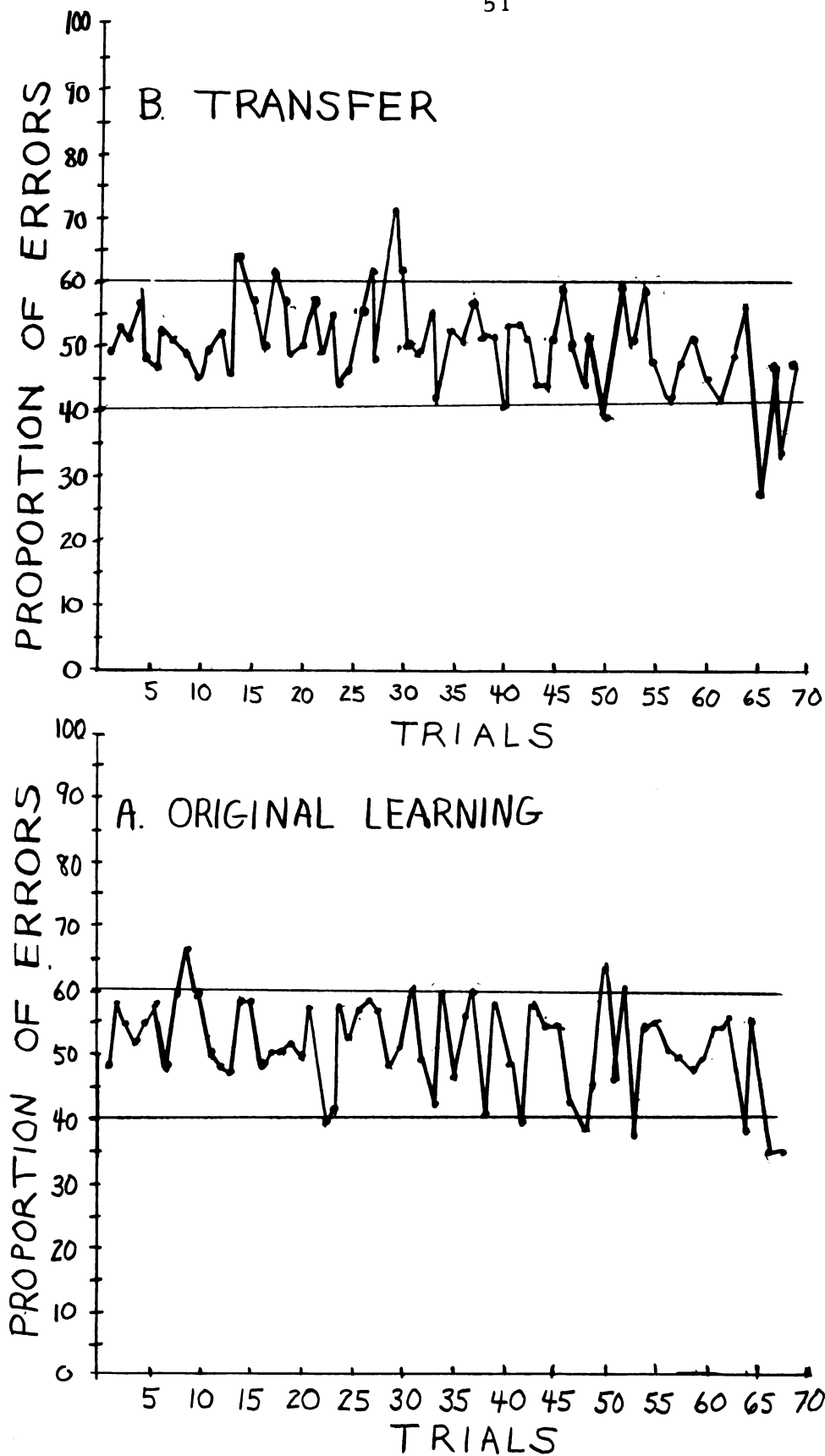


Fig. 4.1. The probability of an error, conditional on \underline{S} being in the pre-solution phase, for original learning (A) and transfer (B).

The probability of solving with zero errors is c , with one error is $c(1-c)$, with 2 errors is $c(1-c)^2$, and with n errors is $c(1-c)^n$, a geometric distribution. The cumulative geometric gives the cumulative proportion of Ss who solve making n or less errors, that is

$$p(n) = \sum_{i=1}^n c(1-c)^i = 1 - (1-c)^{n+1} \quad (4.4)$$

for a given value of c .

The second part of the test of the stochastic properties of the data is to compare obtained distributions with theoretical distributions of error scores using Eq. 4.4. This, in turn, requires an estimate of the parameter c for each group.

Estimation of c , the Probability of Selecting a Correct Strategy

A maximum-likelihood estimator of c , by Eq. 2.18, which takes into account the presence of non-solvers in a group, is

$\hat{c} = N_s / [T + N_s]$, where N_s is the number of solvers and T is the total errors made by the group.

Using Eq. 2.18, estimates of c were calculated for all groups in original learning. Six groups made zero or near-zero errors on transfer, leaving estimates for four transfer groups. Two of the transfer groups (Angle + Angle Color and Angle + Flower Color) should show transfer effects and their distributions would not be geometric. Thus, a total of ten estimates on original learning and two on transfer data were calculated. These values, their standard deviations, corresponding mean errors, and number of solvers in each group are given in Table 4.2.

Table 4.2. Maximum-likelihood Estimates of c

Group	Original Learning			
	Mean Errors	Number of Solvers	\bar{c}	Standard Deviation of \bar{c}^{**}
Angle	19.50	14	.035	.0092
Red Angle	12.45	18	.067	.0152
Green Angle	19.40	14	.035	.0092
Angle + Angle Color	3.45	20	.225	.0443
Angle Color Control	4.05	19	.190	.0392
Angle + Flower Color	2.40	20	.294	.0552
Flower Color Control	3.40	19	.218	.0442
Angle + Angle Color Irrelevant	14.65	17	.055	.0130
Large Angle	8.35	18	.097	.0217
Angle Only	2.60	20	.278	.0528
Transfer				
Angle Color Control	20.10	14	.034	.0089
Flower Color Control	19.75	14	.034	.0089

^{**} The standard deviation of \bar{c} was obtained by taking the square root of the value obtained by substituting \bar{c} for each group into Eq. 2.20 which is $\text{var}(\bar{c}) = \bar{c}^2(1-\bar{c})/N_s$, the variance of \bar{c} .

Goodness-of-fit Distribution Tests

Using the obtained estimates of c , given in Table 4.2, the theoretical cumulative proportion of \underline{Ss} making n or fewer errors were computed by Eq. 4.4 for each group. Empirical cumulative proportions were also obtained. Non-solvers in some groups produced a piling up of error scores near 32. Their error scores were assumed to be binomially distributed with mean, $64(\frac{1}{2}) = 32$, and variance, $64(\frac{1}{2})(\frac{1}{2}) = 16$. Thus, the theoretical distribution of \underline{Ss} making n or fewer errors is a summed geometric distribution combined with this binomial distribution of non-solvers. Figures 4.2, 4.3 and 4.4 present the observed and theoretical distributions for ten original learning and two transfer groups.

The fit of theoretical to obtained distributions is very close in each of the twelve graphs. The degree of approximation was tested by the Kolmogorov-Smirnov one-sample test since expected values of low error scores were too small to justify the χ^2 test (Walker and Lev, 1953, p. 443). Maximum discrepancies between theoretical and obtained distributions are given in Table 4.3.

Maximal discrepancies tended to be small, averaging .169 for the twelve comparisons. In ten cases, the p value of the observed maximum discrepancy was larger than .20 and in one case, larger than .05. Only in Angle Color Control was the theoretical distribution rejected at the .05 but not at the .01 level. This group learned rapidly and had more \underline{Ss} making exactly 2 errors than was expected. The failure to reject the approximations in eleven of twelve cases agrees with the assumption that c , the probability of moving from the pre-solution to the solution phase, is a constant.

Figs. 4.2, 4.3 and 4.4. Observed and theoretical cumulative distributions of the proportion of Ss making n or fewer errors. Theoretical distributions are computed from equations and parameter estimates given in the text.

Fig. 4.2

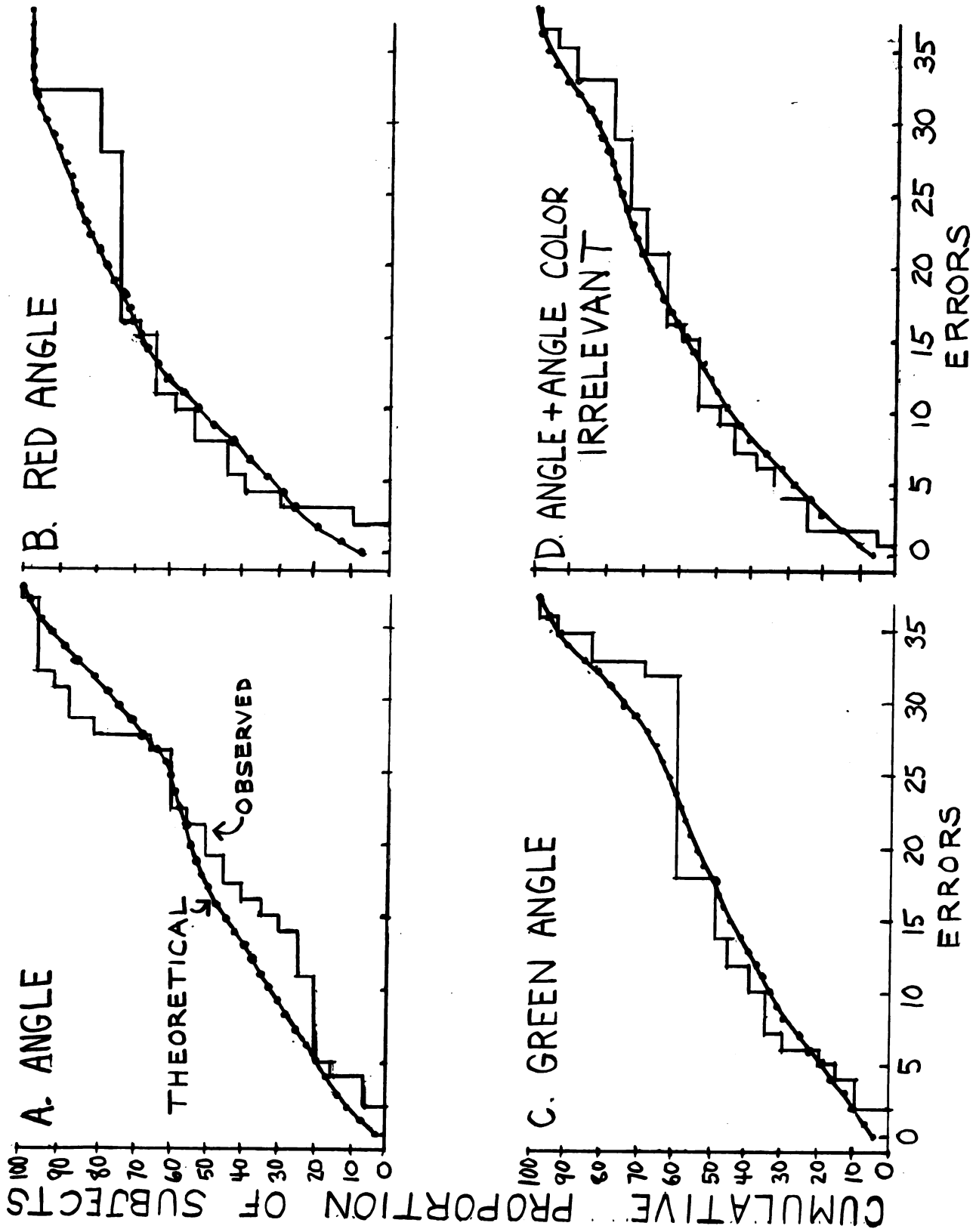


Fig. 4.3

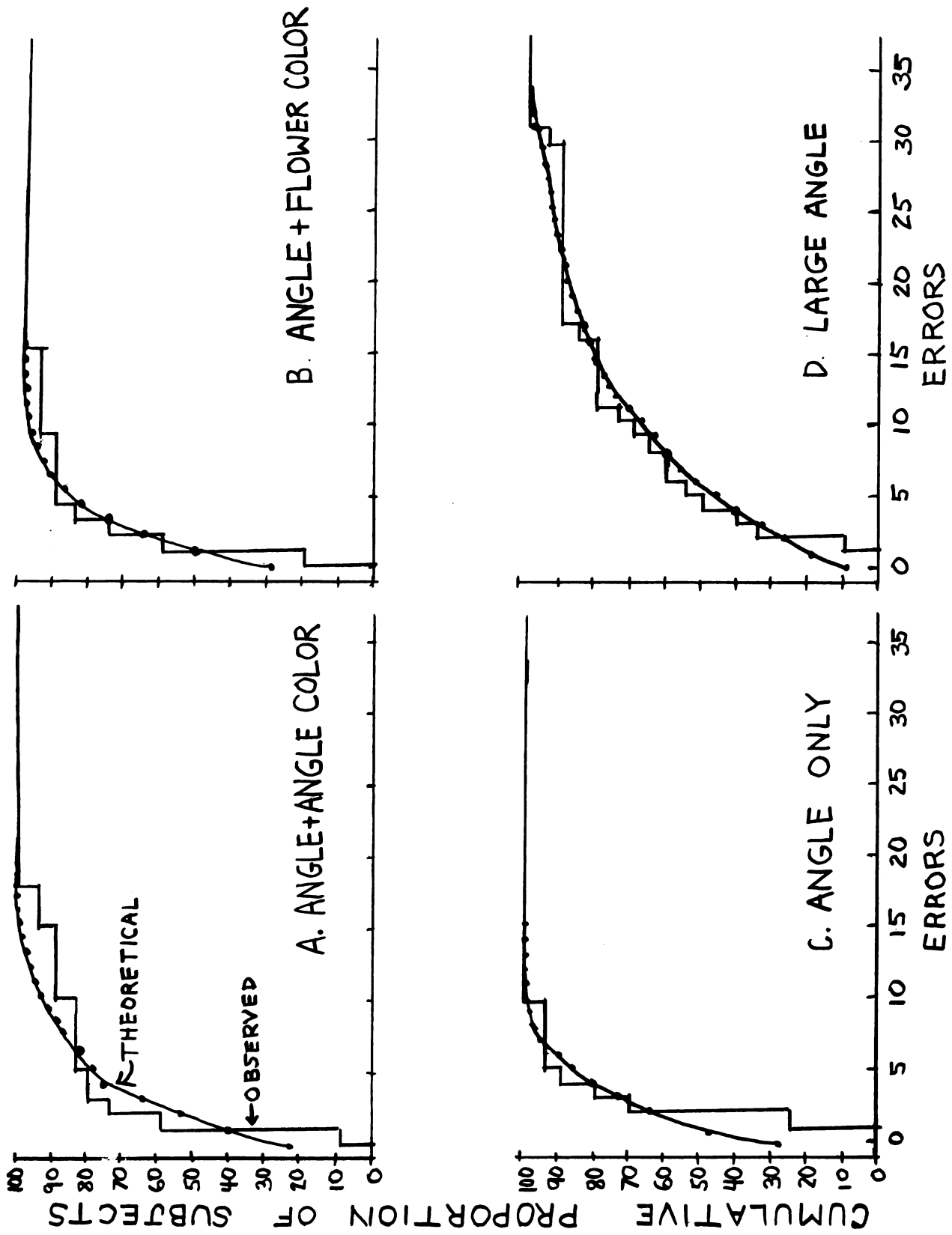


Fig. 4.4

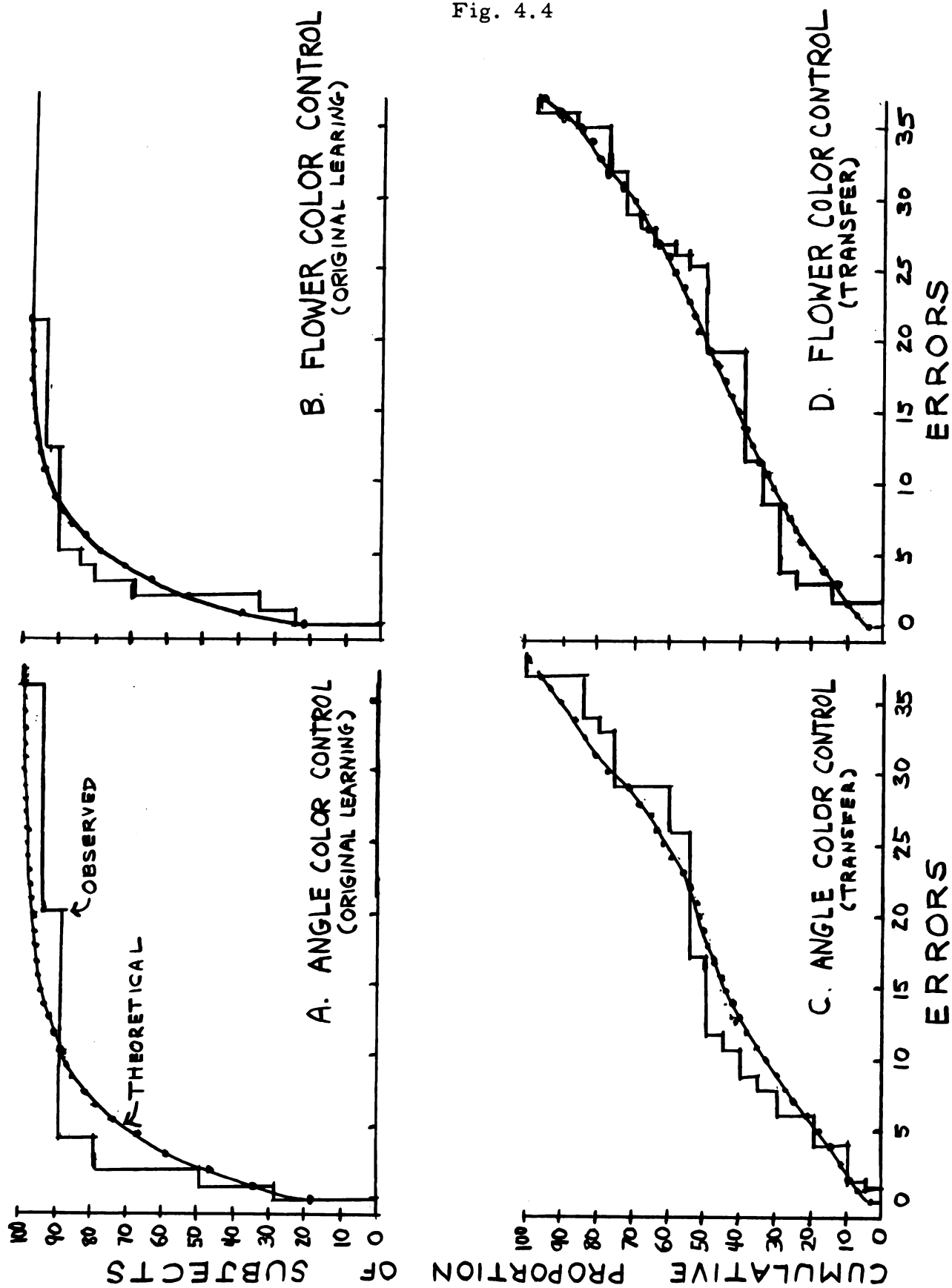


Table 4.3. Maximum Discrepancies Between Theoretical and Observed Error Distributions

Group	Original Learning Maximum Discrepancy
Angle	.114
Red Angle	.125
Green Angle	.216
Angle + Angle Color Irrelevant	.105
Large Angle	.100
Angle Only	.278*
Angle + Angle Color	.215
Angle Color Control	.332**
Angle + Flower Color	.101
Flower Color Control	.178
Transfer	
Angle Color Control	.114
Flower Color Control	.150

* P: .10 > P > .05

** P: .05 > P > .01

Estimation of w , the Proportion of Wrong Strategies

The mean proportion of errors in the pre-solution phase, \bar{p}_a , was reported above to be .55 for original learning and .51 for transfer data. The fact that these values are larger than .50 suggests that w is larger than zero, since a high frequency of consecutive errors in the pre-solution phase indicates a large proportion of wrong strategies and chance frequency of consecutive errors is an indication of relatively few wrong strategies.

If w is small or zero, then errors should be as frequent after errors as after correct responses. To test this, frequencies of errors followed by errors and correct responses followed by errors were tabulated for each S during the pre-solution phase. Chi² tests were performed on the resulting 1 x 2 table for each S where the expected frequency of each cell was at least 5. These values were summed over S s within and then, over groups. During original learning, 69 S s showed a pooled chi² of 59.85 which is not significant. During transfer, the pooled chi² was 60.72 with 55 df, again not significant at the .05 level ($z = 0.58$). These tests indicate that the S 's probability of making an error is independent of whether or not he made an error on the previous trial, and that w is probably small.

Comparison of individual S 's proportion of errors following errors and correct responses following errors would also show any tendency to use wrong strategies. This comparison was made and tested by a sign test on the pre-solution trials of all S s who made 2 or more errors. The proportions of S s making more, as many or fewer errors following errors than correct responses following errors are given in Table 4.4.

Table 4.4. Test of Consecutive Errors: Proportions of Ss Making More, as Many, or Fewer Errors Following Errors than Correct Responses Following Errors

Condition	Proportion of <u>Ss</u>			Number of <u>Ss</u>
	More	As Many	Fewer	
Original Learning	.50	.19	.31	163
Transfer	.43	.08	.49	76

Ss made significantly more errors following errors than correct responses following errors during original learning ($z = 2.79$, $P < .01$). The difference was not significant on transfer ($z = -0.36$, $P > .05$).

Estimates of w were obtained by Eq. 2.21.

$w = [\bar{M}_1 - \bar{M}_0 + 1] / [\bar{M}_1 + \bar{M}_0] = [\bar{M}_1 + \bar{M}_0 + 1] / \bar{T}$, a correction of Restle's Eq. 38 (1961b, p. 299), where \bar{M}_0 is the mean number of errors followed by correct responses and \bar{M}_1 is the mean number of errors followed by errors. \bar{T} , the mean total error score of a group, is equal to $\bar{M}_0 + \bar{M}_1$. In the computation of \bar{M}_0 and \bar{M}_1 , the last error in each S's sequence is taken into account. The obtained estimates of w are reported in Table 4.5 and are compared with estimates of c .

The estimates of w were larger than the estimates of c in all twelve cases. A positive relationship between w and c is expected from the theory since both correct and wrong hypotheses depend upon the relevant dimension. Inspection of the pairs of values in Table 4.5 indicates a positive relationship. The rank order correlation between the two sets of estimates was $+ .79$, ($P < .01$), indicating that w is significantly related to c .

Since all estimates of w were larger than zero, there is an indication that Ss use both wrong and irrelevant strategies in solving the problems and do not merely guess.

Table 4.5. Estimates of w , the Proportion of Wrong Strategies, Compared With Estimates of c . (Groups are ranked according to \hat{c}).

Group	Original Learning	
	\hat{w}	\hat{c}
Angle	.092	.035
Green Angle	.129	.035
Angle + Angle Color Irrelevant	.188	.055
Red Angle	.100	.067
Large Angle	.138	.097
Angle Color Control	.383	.190
Flower Color Control	.353	.218
Angle + Angle Color	.275	.225
Angle Only	.462	.278
Angle + Flower Color	.542	.294
Transfer		
Angle Color Control	.149	.034
Flower Color Control	.104	.034

Individual Differences

The assumption that the probability of selecting a correct strategy is a constant and is the same for all Ss implies that individual differences are random and not stable. To test this, inter-correlations were computed on error scores between the practice problem (P), original learning (OL) and transfer (T) for all groups where possible. The data for the correlations of P-OL was from the first 20 Ss and for correlations between P-T and OL-T, from the 20 solvers in each group. The correlations are presented in Table 4.6.

Of the 18 correlations shown, only one was significant. This correlation, for Angle + Angle Color Irrelevant between P and OL, dropped from 0.52 to 0.34 (not significant) when three additional solvers of the group were added. The set of non-significant intercorrelations obtained shows the absence of any stable individual differences on problems, a result in accord with the Strategy Selection Theory.

The evidence in this chapter is consistent with the Strategy Selection Theory. An analysis of the pre-solution performance indicated that Ss distribute their errors randomly with probability of an error near .50. The probability was found to be stationary in the pre-solution phase by (a) evidence that frequencies of errors during the first and second halves of the pre-solution phase were equal and (b) the proportion of errors, conditional on S being in the pre-solution phase, was constant and near .50 over trials where errors were made. Fitted error distributions from the theory yielded good approximations in eleven of twelve cases. Evidence for Ss using wrong strategies was given, although the proportion of wrong strategies, w , was indicated by independence tests to be small. A high positive correlation between estimates of w and c , the proportion of correct strategies was found, indicating a relationship expected from the theory. Inter-correlations between practice, original learning and transfer problems indicated no stable individual differences, as expected from the theory.

Table 4.6. Inter-correlations Between Practice (P), Original Learning (OL), and Transfer (T) Problems. (N = 20 Ss).

Group	P-OL	P-T	OL-T
Angle	-0.03	--	--
Red Angle	-0.24	--	--
Green Angle	+0.06	--	--
Angle + Angle Color	+0.04	-0.07	-0.32
Angle Color Control	-0.34	-0.15	+0.24
Angle + Flower Color	-0.13	-0.35	+0.03
Flower Color Control	-0.31	-0.12	-0.22
Angle + Angle Color Irrelevant	+0.52**	--	--
Large Angle	+0.12	--	--
Angle Only	+0.23	--	--

** .05 > P > .01

CHAPTER V

EXPERIMENTAL RESULTS

To evaluate the effect of the training conditions on acquisition, the original learning performance of each experimental group is compared with that of group Angle. The effect of training on transfer is studied by comparing each experimental group's performance on the Angle problem with the original learning of group Angle. Together, these comparisons constitute examples of the transfer-of-training paradigm, $X \rightarrow Y$ versus Y .

Discriminanda Difference Increase and Removal of Irrelevant Dimensions

Large Angle and Angle Only both learned significantly faster than group Angle. Doubling the angle size difference in Large Angle decreased mean errors by slightly more than one-half. Removal of the irrelevant dimensions by holding them constant made acquisition of the angle concept considerably easier to learn in Angle Only. See Table 5.1.

Table 5.1. Comparison of Large Angle and Angle Only with Angle on Original Learning

Group	Mean Errors	S.D.*	t**	Parameter Tests	
				z***	-2ln(λ)****
Angle	19.50	10.51	--	--	--
Large Angle	8.35	9.07	-3.60	2.58	8.75
Angle Only	2.60	2.06	-7.07	4.50	36.45

* Standard Deviation

** $t_{.05} > 2.04$ with $df = 30$

*** $z_{.05} = 1.96$

**** $-2 \ln(\lambda) = \text{Chi}^2_{.05} = 3.80, 1 \text{ df}$

Both groups showed nearly perfect transfer to the Angle problem. In each group, only three Ss made errors on transfer. Transfer was so very positive that statistical tests comparing Large Angle and Angle Only with Angle were unnecessary. After Ss reached learning criterion, reduction in the relevant angle size difference or introduction of new irrelevant dimensions had a very small disruptive effect on transfer. Transfer data for Large Angle and Angle Only are summarized and compared with original learning performance on Angle in Table 5.2.

Table 5.2. Transfer to the Angle Problem by Large Angle and Angle Only Compared with Original Learning of Angle

Group	Mean Errors	S.D.
Angle	19.50	10.51
Large Angle	0.20	0.62
Angle Only	0.40	1.35

Color as a Constant Emphasizer

The role of color as a constant emphasize on acquisition of the angle concept was studied by comparing Red Angle and Green Angle with Angle on original learning. Transfer of the constant emphasize groups to the Angle problem tested the effect of removing a stimulus which could not be the basis for solution. Red Angle and Green Angle are compared with Angle on original learning in Table 5.3 and on transfer in Table 5.4.

The effect of color on the angles during original learning was not pronounced, but transfer to the Angle problem was perfect for Red Angle and nearly perfect for Green Angle. Since the transfer was perfect or nearly perfect, removal of color, which was not a basis for a correct strategy but an emphasize, had no disruptive effect.

Table 5.3. Comparison of Red Angle and Green Angle with Angle on Original Learning

Group	Mean Errors	S. D.	t	Parameter Tests	
				z	-2 ln(λ)
Angle	19.50	10.51	--	--	--
Red Angle	12.45	11.79	-2.00	1.85	4.37
Green Angle	19.40	13.01	-0.03	0.00	0.00

Table 5.4. Transfer to the Angle Problem by Red Angle and Green Angle Compared with Original Learning of Angle

Group	Mean Errors	S. D.
Angle	19.50	10.51
Red Angle	0.00	0.00
Green Angle	0.05	0.00

On original learning Red Angle made one-third fewer errors than Angle, but the difference was of borderline significance in two of the three statistical tests. Green Angle made the about same number of errors as Angle, indicating that the green color had no emphasizing effect.

As a further control on transfer, 14 Ss in Angle who solved the original learning problem were given an additional 10 trials and made no errors on this "hard-to-hard" transfer.

Color as an Irrelevant Dimension and an Emphasizer

In Angle + Angle Color Irrelevant, color played two rôles which lead to opposite effects. If the irrelevant rôle predominated, the group would learn more slowly than Angle. If the emphasize rôle predominated, faster learning would occur. The performance of Angle + Angle Color Irrelevant is summarized and compared with Angle in Table 5.5.

Table 5.5. Comparison of Angle + Angle Color Irrelevant with Angle on Original Learning and Transfer

Group	Condition	Mean Errors	S.D.	t	Parameter Tests	
					z	-2 ln (λ)
Angle	Original Learning	19.50	10.51	--	--	--
Angle + Angle Color Irrelevant	Original Learning	14.65	12.99	-1.31	1.26	1.77
Angle + Angle Color Irrelevant	Transfer	0.00	0.00	unnecessary		

When color was both an emphasizee and an irrelevant dimension, learning was slightly facilitated. Although Angle + Angle Color Irrelevant made fewer errors than Angle on original learning, the difference was not significant. The slight facilitation is an indication, however, that the emphasizee rôle predominated over the retarding effects of the added irrelevant dimension. Since color was irrelevant and could not be the basis for a correct strategy its removal had no effect as Angle + Angle Color Irrelevant showed perfect transfer to the Angle problem.

Color as an Emphasizer or a Counter-emphasizer
and a Redundant Relevant Dimension

Angle + Angle Color and Angle + Flower Color, which had problems involving color as a relevant and redundant cue during original learning, are compared with the original learning of Angle in Table 5.6 and on transfer in Table 5.7.

Table 5.6. Comparison of Angle + Angle Color and Angle + Flower Color with Angle on Original Learning

Group	Mean Errors	S. D.	t	Parameter Tests	
				z	-2 ln(λ)
Angle	19.50	10.51	--	--	--
Angle + Angle Color	3.45	5.31	-6.10	4.44	30.85
Angle + Flower Color	2.40	3.59	-6.90	4.52	40.20

Table 5.7. Transfer to the Angle Problem by Angle + Angle Color and Angle + Flower Color Compared with Original Learning of Angle *

Group	Mean Errors	S.D.	t
Angle	19.50	10.51	--
Angle + Angle Color	16.45	14.45	-0.76
Angle + Flower Color	23.45	12.14	1.13

* Normal distribution and likelihood ratio parameter tests were not made, since with the possibility of transfer, error distributions would no longer be geometric.

Both groups learned faster than Angle which had only the angle dimension relevant. All Ss solved the original learning problem and were transferred to the Angle problem in both groups. Angle + Angle Color, in which the angle was emphasized during original learning, made slightly fewer errors on transfer than Angle. Angle + Flower Color, in which the angle was counter-emphasized during original learning, made slightly more errors on transfer than Angle. Transfer for both groups did not differ significantly from the original learning by Angle.

Color Versus Angle as a Salient Cue

Color was found to be more salient as a cue than angle. Both Angle Color Control and Flower Color Control, which had fixed 45° angles, made significantly fewer errors than Angle. See Table 5.8.

Table 5.8. Comparison of Angle Color Control and Flower Color Control with Angle on Original Learning

Group	Mean Errors	S. D.	t	Parameter Tests	
				z	-2 ln(λ)
Angle	19.50	10.51	--	--	--
Angle Color Control	4.05	8.49	-6.13	3.88	20.22
Flower Color Control	3.40	5.31	-6.12	4.07	28.94

Additivity of Strategies

Additivity of angle and color strategies is suggested since the redundant relevant cue groups made fewer errors than their color controls which had only color relevant, or the Angle group, which had only the angle relevant. Angle + Angle Color made 3.40 mean errors compared with 4.05 for Angle Color Control and Angle + Flower Color averaged 2.40 errors compared with 3.45 for Flower Color Control. Angle made 19.50 mean errors.

Test for Transfer of an Observing Response

The slight positive transfer of Angle + Angle Color and the slight negative transfer of Angle + Flower Color to the Angle problem, although not significant, could be interpreted as supporting a strict interpretation of the observing response theory of Wyckoff (1952). The interpretation is that Ss learned to look at that part of the pattern which was colored during original learning. Transfer of an observing response to the angle dimension would lead to positive transfer in Angle + Angle Color, for Ss would have presumably learned to look at the angles during original learning. Transfer of an observing response to the flowers in Angle +

Flower Color would lead to negative transfer on the Angle problem.

The possibility of transfer of an observing response to the place which was colored during original learning was tested by comparing transfer performance of Angle Color Control and Flower Color Control on the Angle problem with the original learning of Angle. The comparisons are given in Table 5.9.

Table 5.9. Transfer to the Angle Problem by Angle Color Control and Flower Color Control Compared with Original Learning of Angle

Group	Mean Errors	S. D.	t	Parameter Tests	
				z	-2ln(λ)
Angle	19.50	10.51	--	--	--
Angle Color Control	20.10	13.86	0.15	-0.01	0.00
Flower Color Control	19.75	14.01	0.06	-0.01	0.00

Transfer of Angle Color Control and Flower Color Control to the Angle problem did not differ significantly from original learning by Angle. These results indicate that there was no important transfer of an observing response to the place which was colored during original learning.

Other Evidence for an Emphasizer Effect

One way of evaluating the emphasizer and counter-emphasizer effects in Angle + Angle Color and Angle + Flower Color is to compare the number of Ss in each group who show perfect transfer to the Angle problem.

Perfect transfer indicates that an S probably solved on the angle dimension during original learning.

In transfer from Angle + Angle Color, five Ss made no errors, whereas, from Angle + Flower Color, only one S made no errors. The difference is borderline significance ($z = 1.82$, $P < .10$) and is attributed to the role of color. In Angle + Angle Color, color was spatially contiguous to the angles. In Angle + Flower Color, the color was spatially separated from the angles. In Angle + Angle Color, the color apparently increased the probability of Ss attending to and solving on the basis of the angle dimension, but in Angle + Flower Color, color had the opposite effect.

When the five Ss who made no errors on the Angle problem are removed from Angle + Angle Color, the mean errors for the remaining fifteen Ss was 21.93, not far from the 20.10 of Angle Color Control.

Efficiency of Training

Total efficiency comparisons of all the groups cannot be adequately evaluated since a number of Ss failed to solve within the time limit on original learning. The existence of non-solvers on transfer also precludes a trials to criterion measure or a total error measure for use in efficiency comparisons.

Three groups, Angle Only, Angle + Angle Color and Angle + Flower Color, did have all Ss solve original learning problems at about the same rate. In Angle Only, training Ss on the relevant angle without irrelevant dimensions present and then transferring them to the Angle problem proved to be a highly efficient training procedure. Angle Only made a total of 3.00 mean errors on original learning and transfer, with all Ss solving the Angle Problem. Compared with the performance of Angle, the difference was highly significant ($t = 6.88$).

The redundant cue groups, Angle + Angle Color and Angle + Flower Color, did not perform significantly better on transfer than the Angle group in original learning. Training on these redundant-cue Easy problems was apparently wasted. The apparent inefficiency of this redundant-cue program is difficult to evaluate, because the program gave the Ss more opportunities to make errors. However, the redundant-cue-to-hard procedure cannot be rated more efficient than training on the hard problem alone.

Despite the presence of two non-solvers, Large Angle, which made 8.35 mean errors on original learning and only 0.20 mean errors on transfer, may be regarded as more efficient than Angle, which had six non-solvers and made 19.50 mean errors. This result of efficiency in transfer along a stimulus continuum in a concept identification task is in accord with previous findings on this training program (Lawrence, 1952; Baker and Osgood, 1954; Restle, 1955).

Comparison of Statistical Tests

In all comparisons that were made, the probability values of the t- and z- tests and $-2 \ln(\lambda)$ were about equal. These results suggest that the tests are about the same power.

Summary of Experimental Findings

The major experimental findings reported in this chapter are:

1. Doubling the angle size difference during original learning led to faster acquisition and nearly perfect transfer of the angle concept.
2. Removing the irrelevant dimensions during original learning led to faster learning and nearly perfect transfer of the angle concept.
3. The constant emphasize effects were not strong. When red appeared on the angles, acquisition was faster. When green was used,

no emphasize effect was shown. Transfer from the constant color emphasize groups to the Angle problem was perfect or nearly perfect.

4. Adding color which was an irrelevant dimension and an emphasize led to slightly faster learning. Transfer to the Angle problem was perfect. Removing color, when not a basis for a correct strategy, had no disruptive effect on transfer.

5. Adding color as a redundant relevant cue led to faster learning but transfer was either (a) slightly positive when the color had been an emphasize during original learning or (b) slightly negative when color had been a counter-emphasize during original learning.

6. Color was found to be a more salient cue than the angle.

7. Two procedures appeared to be most efficient:

- a. training Ss on the relevant angle cue in the absence of irrelevant dimensions and then transferring them to the hard Angle problem and
- b. training Ss with a large difference between the angle sizes and then transferring them to the hard Angle problem.

CHAPTER VI

DETAILED PREDICTIONS

According to the Strategy Selection Theory, the probability of solving the problem on any trial is c , the proportion of correct strategies. Correct and wrong strategies depend upon stimulus dimensions which vary during training and are correlated with reward. Irrelevant strategies depend upon stimulus dimensions which vary but are uncorrelated with reward. Other factors such as the outcome of previous trials, incidental events, etc. also determine the number of irrelevant strategies in the problem.

The strategies in the present experiment arise from the flower patterns. Several problems had patterns where the angle dimension was varied and relevant. Some of the strategies arising from the angle dimension depend upon the contrast of the small and large angles. These strategies are relevant when that dimension is made relevant, as when the 30° angles are in class A and the 60° angles are in class B. Such strategies from the angle are called A . Strategies from the flower color are called C_f and from angle color are called C_a . All other irrelevant strategies in the problem, from all sources, are assembled into one set, I .

If a dimension is fixed (does not vary), then strategies depending upon the contrast between values of the dimension do not exist in the problem, though other strategies arising from them are presumably present. Fixing a dimension by making it one-valued removes just the contrast set of strategies but does not affect the irrelevant strategies in the set I (cf. Trabasso, 1960).

The following sets of strategies recur, in various combinations, in more than one problem: Angle (A), Color on Angle (C_a), Color on Flower (C_f), and the irrelevant or background set, (I). In addition, two emphasize effects (multipliers), r and g , for red and green emphasizeers respectively, are used. Theoretically, these six numbers are sufficient to calculate the probability of solving, c , for eight groups by use of Eqs. 2.6 to 2.10 of Chapter II. These same parameters, as employed in Eqs. 2.11 to 2.14 of Chapter II, yield quantitative predictions of transfer.

The measures of the sets A , C_a , and C_f will be estimated relative to the measure of I . For example, to estimate $m(A)$, we use the Angle problem in which the angle dimension is relevant, color does not appear, no emphasizeer was used, and the set I are irrelevant. Since evidence was found for a strong relationship between the proportions of wrong and correct strategies (Chapter IV), the assumption made in Chapter II that wrong and correct strategies from any dimension are equal in measure, i.e., that $m(A) = m(A^*)$, appears justified, and shall be made. In all problems to be considered, the dimensions of flower shape, leaf shape and leaf position are irrelevant, and strategies based on these dimensions are incorporated into the irrelevant set I . Then from Eq. 2.6,

$$\begin{aligned} c(\text{Angle}) &= m(A) / [m(A) + m(A^*) + m(I)] \\ &= m(A) / [2m(A) + m(I)]. \end{aligned} \quad (6.1)$$

In Chapter IV, Table 4.1, an estimate of c for the angle group is given as .035. Putting this estimate in for $c(\text{Angle})$ in Eq. 6.1, we find

$$m(A) = .038 m(I). \quad (6.2)$$

This gives the measure of angle strategies relative to the measure of irrelevant strategies.

Formulas for c in five groups which are used to estimate $m(A)$, $m(C_a)$, $m(C_f)$, r , and g are shown in summary form in Table 6.1.

The formulas are derived using equations from Chapter II, as referred to in the table. Expressions proportional to $m(I)$, analogous to Eq. 6.2, are shown in Table 6.2.

Table 6.1. Formulas of c for Experimental Groups used in Predictions

Group	c	Definition of c	Theoretical Equation
Angle	.035	$m(A)/[2m(A) + m(I)]$	2.6
Red Angle	.067	$r.m(A)/[2.r.m(A) + m(I)]$	2.9
Green Angle	.035	$g.m(A)/[2.g.m(A) + m(I)]$	2.9
Angle Color Control	.190	$m(C_a)/[2m(C_a) + m(I)]$	2.6
Flower Color Control	.218	$m(C_f)/[2m(C_f) + m(I)]$	2.6

Table 6.2. Estimates of Sets of Strategies Used in Predictions

Set of Strategies	Measure	Proportional Estimate
Angle	$m(A)$	$.038m(I)$
Red Angle	$r.m(A)$	$.077m(I)$
Green Angle	$g.m(A)$	$.038m(I)$
Angle Color	$m(C_a)$	$.306m(I)$
Flower Color	$m(C_f)$	$.386m(I)$

Additivity of Strategies in Original Learning

Strategies are said to be additive when it is shown that Ss can learn a problem with two relevant sets of strategies (either of which can be used to solve the problem) more rapidly than a problem with only one set of relevant strategies. The Strategy Selection Theory asserts that the rate of learning (c) depends directly on the proportion of correct strategies, so that from additivity of strategies, one can predict additivity of learning rates. Provided that the sets of strategies are disjoint, these predictions constitute a test of the correctness of the theory.

In the problem of Angle + Angle Color, the emphasized angle and the color of the angles are relevant and redundant dimensions. All others are irrelevant. When the angle is red, by Eq. 2.8 and Eq. 2.9,

$$c(\text{Red Angle} + \text{Angle Color}) = \frac{r \cdot m(A) + m(C_a)}{[2r \cdot m(A) + 2m(C_a) + m(I)]}$$

and when the angle is green,

$$c(\text{Green Angle} + \text{Angle Color}) = \frac{g \cdot m(A) + m(C_a)}{[2g \cdot m(A) + 2m(C_a) + m(I)]}$$

Since red and green angles appear equally often, the mean

$$c(\text{Angle} + \text{Angle Color}) = \frac{1}{2} \left[\frac{r \cdot m(A) + m(C_a)}{2r \cdot m(A) + 2m(C_a) + m(I)} \right] + \frac{1}{2} \quad (6.3)$$

$$\left[\frac{g \cdot m(A) + m(C_a)}{2g \cdot m(A) + 2m(C_a) + m(I)} \right] .$$

To predict $c(\text{Angle} + \text{Angle Color})$, appropriate estimates of the sets of strategies (from Table 6.2) were substituted into Eq. 6.3.

The predicted $c(\text{Angle} + \text{Angle Color}) = .210$, which is close to the observed value of .225. This prediction is not rejected by the likelihood ratio test ($-2 \ln(\lambda) = 0.09$ $P > .05$). This test treats the prediction as a fixed parameter, i. e., tests the hypothesis that $c = c_0$.

Converting the prediction to mean errors by Eq. 2.2, $\bar{T} = (1-c)/c = (1-.210)/(.210) = 3.76$ mean errors predicted where 3.45 were observed. Even if this prediction is taken as a fixed value, the difference between the predicted and observed is not significant at the .05 level ($t = 0.26$).

In the problem of Angle + Flower Color, the angle and the color of the flowers are relevant and redundant dimensions. All others are irrelevant. From Eq. 2.8,

$$c(\text{Angle} + \text{Flower Color}) = \frac{[m(A) + m(C_f)]}{[2m(A) + 2m(C_f) + m(I)]} \quad (6.4)$$

The predicted $c(\text{Angle} + \text{Flower Color}) = .229$, somewhat lower than the observed .294, but the difference is not significant at the .05 level ($-2 \ln(\lambda) = 1.52$).

Predicted mean errors are $(1-.229)/(.229) = 3.37$, compared with 2.40 observed, a result within sampling variability ($t = 1.46$, $P > .05$).¹

Additivity of Irrelevant Strategies

The theory states that the learning rate (c) depends upon the relative amounts of relevant and irrelevant strategies, from which it follows that increasing the number of irrelevant strategies should retard learning.

¹To follow the theoretical argument exactly, one must say that when the flower is colored, this not only adds relevant color cues but also emphasizes the shape of the flowers. Since the shape of the flower is irrelevant, the result is that one component of I has been emphasized. Unfortunately, this possibility was not foreseen at the time the experiment was designed, and no group has introduced which would permit estimating $m(F)$, the measure of strategies associated with the shape of the flower. Hence, even if it is assumed that the emphasis effects, r and g , would have the same effect on flowers as they have on angle, it is not possible to predict the increase in I . The calculations reported here are predicted on the assumption that the additional measure of irrelevant strategies, through emphasis of flower shape, is negligibly small. The observed discrepancy is opposite to that expected from emphasis of flower shape.

In the problem of Angle + Angle Color Irrelevant, the angle dimension was relevant and emphasized, red and green angles appearing about equally often. Angle color was an added irrelevant dimension and all others were irrelevant. When the angles are red, by Eq. 2.6,

$$c(\text{Red Angle}) = r.m(A)/[2r.m(A) + m(C_a) + m(I)]$$

and when the angles are green,

$$c(\text{Green Angle}) = g.m(A)/[2g.m(A) + m(C_a) + m(I)].$$

Since red and green angles appear equally often, the mean

$$\begin{aligned} c(\text{Angle} + \text{Angle Color Irrelevant}) = \\ = \frac{1}{2} \left[\frac{r.m(A)}{2r.m(A) + m(C_a) + m(I)} \right] + \frac{1}{2} \left[\frac{g.m(A)}{2g.m(A) + m(C_a) + m(I)} \right]. \end{aligned} \quad (6.5)$$

The predicted $c(\text{Angle} + \text{Angle Color Irrelevant}) = .040$, lower but not significantly different from the observed value of $.055$ ($-2 \ln(\lambda) = 1.54$, $P > .05$).

Converting to predicted mean errors, $\bar{T} = (1 - .040)/(.040) = 24.00$ where only 14.65 were observed. Taking the prediction as a fixed value, the difference is significant at the .05 level ($t = 3.22$).²

Transfer

Ss who were transferred to the Angle problem had first met a learning criterion on the original learning problem. The probability of

²The conversion formula for predicted mean errors assumes that learning is complete. This condition was not satisfied since the group was stopped at the end of 64 trials, and 3 of the 20 Ss were still making errors. If the same conversion formula were applied to the maximum-likelihood estimate of .055, $\bar{T} = 17.48$, which is still less than the prediction. The likelihood ratio test is not as stringent as the t-test since the t-test does not take into account the sampling variance of the parameter estimates used in the prediction.

solution on problem 2 (transfer) after having solved problem 1 (original learning) is given in Eq. 2.11 (Chapter II).

Consider those original learning problems where the angle dimension is the only relevant dimension, namely Angle, Red Angle, Green Angle, and Angle + Angle Color Irrelevant. If color is present and either constant or irrelevant, then the \underline{S} cannot come to depend upon it as a basis for a correct strategy. Since the angle dimension is relevant in both problems 1 and 2 and neither color nor any other strategy can be correct, $C_1 = C_2$. Eq. 2.11 then becomes unity and all groups should show perfect transfer.

This prediction is confirmed. Angle, Red Angle, Green Angle and Angle + Angle Color Irrelevant showed perfect or nearly perfect transfer. A total of 74 \underline{S} s made only 1 error in transfer.

It is possible for \underline{S} to have reached criterion by chance and then commit errors on transfer. In original learning, the probability of ten in a row correct is $c + (1-c)(\frac{1}{2})^{10}$. Using Angle as an example, with $\bar{c} = .035$, the guessing probability is

$(1-c)(\frac{1}{2})^{10} \cong .001$. The proportion of \underline{S} s who guess to those who solve is $(1-c)(\frac{1}{2})^{10} / c$, or in Angle, $.001/.035 = 1/35$. This result means that one might expect one \underline{S} in thirty-five to solve the original learning problem by chance and commit errors on transfer. This did not occur in the present study. (One \underline{S} in Green Angle made one error on transfer).

In the control problems of Angle Color Control and Flower Color Control, the C_1 strategies arose from the color dimension. Since the C_2 strategies were based on the angle, the set $(C_1 \cap C_2) = 0$. Then, Eq. 2.11 is also zero. The expectation is that these groups would perform on the Angle problem much like S s in Angle. The expectation is confirmed since no evidence for any kind of transfer was observed in these groups.

Predictions of transfer for the redundant-cue groups is by Eq. 2.13. For Angle + Angle Color, the conditional probability of solving on the angle during original learning is

$$P(A/A \cup C_a) = \frac{1}{2} \left[\frac{r.m(A)}{r.m(A) + m(C_a)} \right] + \frac{1}{2} \left[\frac{g.m(A)}{g.m(A) + m(C_a)} \right] \quad (6.6)$$

The predicted proportion of Ss who solve on the angle during original learning is .155, and the expected number is $20(.155) = 3.10$. The remaining 16.90 Ss, who have theoretically solved on angle color, should show no transfer to the Angle problem since C_a is not contained in the A strategies of problem 2. Using the estimate of c from Angle Color Control, $.034(16.90) = 0.57$ of these Ss are expected to solve the Angle problem without any errors. Together, $3.10 + 0.57 = 3.67$ Ss are predicted to make zero errors in transfer to the Angle problem. Five Ss in Angle + Angle Color did show perfect transfer.

To predict mean errors on transfer for Angle + Angle Color, the expectation is that 3.10 Ss make no errors since they solved on the angle during original learning and that the remaining 16.90 Ss would perform like their controls in Angle Color Control. Since Angle Color Control averaged 20.10 errors, the predicted mean errors for Angle + Angle Color on transfer is

$$(3.10)0 + \frac{16.90(20.10)}{20} = 16.98, \text{ very close to the } 16.45 \text{ observed.}$$

Taking the prediction as a fixed value, the difference is not significant at the .05 level ($t = 0.16$).

For Angle + Flower Color, by Eq. 2.13, the conditional probability of solving on the angle during original learning is

$$P(A/A \cup C_f) = m(A)/[m(A) + m(C_f)]. \quad (6.7)$$

The predicted proportion of Ss who solve on the angle during original learning is .09, and the expected number is $20(.09) = 1.80$. The remaining 18.20 Ss, who have theoretically solved on flower color, should show no transfer to the Angle problem since C_f is not contained in the A strategies of problem 2. Using the estimate of c from Flower Color Control, $.034(18.20) = 0.62$ of these Ss are expected to solve the Angle problem without any errors. Together, $1.80 + 0.67 = 2.42$ Ss are predicted to make zero errors in transfer to the Angle problem. One S in Angle + Flower Color did show perfect transfer.

To predict mean errors on transfer for Angle + Flower Color, the expectation is that 1.80 Ss make no errors since they solved on the angle during original learning and that the remaining 18.20 Ss would perform like their controls in Flower Color Control. Since Flower Color Control averaged 19.75 errors, the predicted mean errors for Angle + Flower Color is

$$(1.80)0 + \frac{18.20 (19.75)}{20} = 17.97, \text{ lower than the } 23.40 \text{ observed.}$$

Taking the prediction as a fixed value, the difference is of borderline significance ($t = 2.02$, $P = .05$).

Predicted Cumulative Error Distributions on Transfer

For Angle Color Control and Flower Color Control, the predicted rate of learning is .035, the value from group Angle. All Ss in these two color control groups are expected to perform like Ss in group Angle since C_a and C_f strategies are not contained in the A strategies of the Angle problem. From Eq. 4.4 (Chapter IV), the predicted cumulative distribution of the proportion of Ss making n or fewer errors, for solvers, is

$$p(n) = 1 - (1 - .035)^{n+1}.$$

Angle contained 6 non-solvers so that 6 non-solvers are expected on transfer from Angle Color Control and Flower Color Control, respectively. Six Ss failed to reach criterion in both these control groups on transfer.

Assuming the binomial distribution with mean = 32 and variance = 16 for the 6 non-solvers in each group, the cumulative normal approximation for the 6 non-solvers was added to the cumulative geometric, $1 - (1 - .035)^{n+1}$, to form the predicted distribution of error scores. Figure 6.1 (C and D) shows the predicted and obtained distributions for Angle Color Control and Flower Color Control. The maximum discrepancy for Angle Color Control was .130 and for Flower Color Control, .146, both of which were non-significant at the .05 level by the Kolmogorov-Smirnov one-sample test (maximum discrepancy allowed = .320).

For Angle + Angle Color, 3.10 Ss are expected to show perfect transfer, as they theoretically solved on the angle during original learning. The learning rate for the 16.90 remaining Ss is .034 from Angle Color Control. The expected number of non-solvers is taken as proportional to the number of Ss who failed to solve in transfer from Angle Color Control. Since $6/20 = .30$, the expectation is that $.30(16.90) = 5.07$ Ss would not solve, whereas 6 were observed in Angle + Angle Color's transfer to the Angle problem.

The predicted cumulative proportion of Ss making n or fewer errors was obtained by

1. using the cumulative geometric (Eq. 4.4), $p(n) = 1 - (1 - .034)^{n+1}$, for the 16.90 Ss who theoretically solved their original learning problem on the angle color,

2. adding to this, a cumulative normal approximation to the binomial distribution of errors for the 5.07 expected non-solvers, with mean = 32 and variance = 16, and

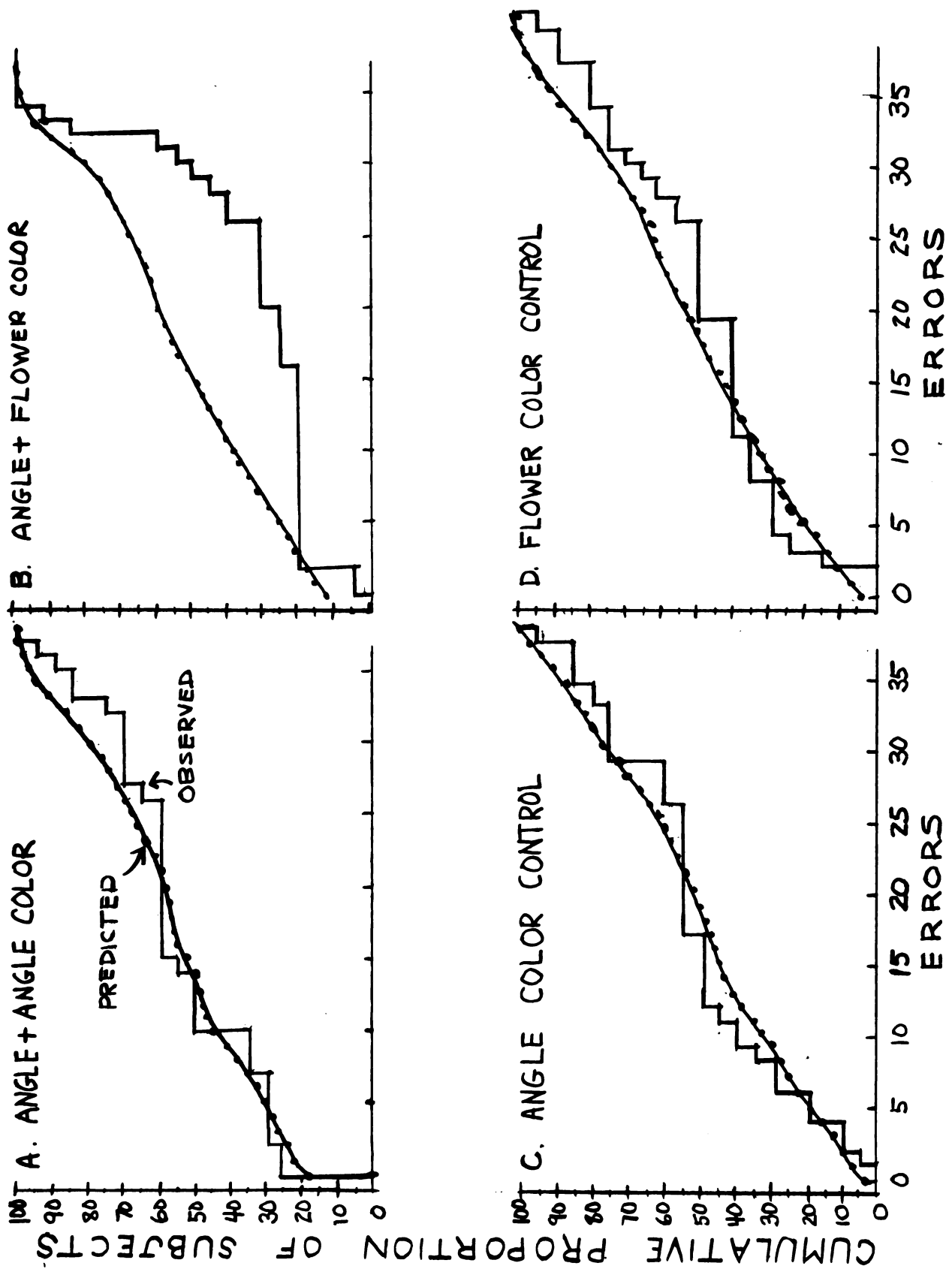


Fig. 6.1. Predicted and observed cumulative distributions of the proportion of Ss making n or fewer errors. Predicted distributions are computed from equations and parameter estimates given in the text.

3. adding to this composite distribution, the 3.10 Ss who were expected to make no errors on transfer since they theoretically solved on the angle during original learning.

The resulting predicted distribution is compared with the observed in Figure 6.1(A).

The fit to the observed distribution for Angle + Angle Color on transfer was quite close. The maximum discrepancy was .144 and is not significant at the .05 level.

For Angle + Flower Color, 1.80 Ss are expected to show perfect transfer as they theoretically solved on the angle during original learning. The learning rate for the 18.20 remaining Ss is .034 from Flower Color Control. The expected number of non-solvers is taken as proportional to the number of Ss who failed to solve in transfer from Flower Color Control. Since $6/20 = .30$, the expectation is that $.30 (18.20) = 5.46$ Ss would not solve, whereas 10 non-solvers were observed in Angle + Flower Color's transfer to the Angle problem.

The predicted cumulative proportion of Ss making n or fewer errors was obtained by

1. using the cumulative geometric (Eq. 4.4), $p(n) = 1 - (1 - .034)^{n+1}$, for the 18.20 Ss who theoretically solved their original learning problem on the flower color,

2. adding to this, a cumulative normal approximation to the binomial distribution of errors for the 5.46 non-solvers, with mean = 32 and variance = 16, and

3. adding to this composite distribution, the 1.80 Ss who were expected to make no errors on transfer since they theoretically solved on the angle during original learning.

The resulting predicted distribution is compared with the observed in Figure 6.1(B).

The fit to the observed distribution for Angle + Flower Color on transfer was poor. The maximum discrepancy was .357, larger than the allowed .320 and the prediction is rejected at the .05 level.

Some Further Questions on Emphasizer Effects

In the predictions of the rate of learning for Angle + Angle Color and Angle + Angle Color Irrelevant, an assumption was made that the color served as a red emphasize on half the trials and as a green emphasize on the other half. Since green had no emphasize effect when it appeared alone (Chapter V), the assumption amounted to saying that emphasis occurred on only the red trials. Another possibility is to consider that red and green contrast from trial to trial has an emphasis effect. Stimulus change has been noted as a variable influencing the S's attention, so that color contrast as an emphasize is not an unreasonable assumption (Berlyne, 1951).

Suppose that the emphasis effect is equal to that of the red emphasize in Red Angle. For Angle + Angle Color Irrelevant, Eq. 6.5 now is

$$c(\text{Angle} + \text{Angle Color Irrelevant}) = r.m(A)/[2r.m(A) + m(C_a) + m(I)]. \quad (6.8)$$

The predicted $c(\text{Angle} + \text{Angle Color Irrelevant})$ is .053, a value very close to the .055 observed ($-2 \ln(\lambda) = 0.07, P > .05$).

Converting to mean errors, $\bar{T} = (1-.053)/(.053) = 17.87$, which is not significantly different from the 14.65 observed. For complete learning, $\bar{T} = 17.48$ for Angle + Angle Color Irrelevant when the maximum-likelihood observed estimate of .055 is used. This being the case, the prediction is quite close.

If no emphasize effect were assumed, Eq. 6.5 becomes

$$c(\text{Angle} + \text{Angle Color Irrelevant}) = m(A)/[2m(A) + m(C_a) + m(I)]. \quad (6.9)$$

The predicted $c(\text{Angle} + \text{Angle Color Irrelevant})$ is .027 significantly lower than the .055 observed ($-2 \ln(\lambda) = 7.09$, $p < .05$).

Converting to mean errors, $\bar{T} = (1 - .027)/(.027) = 36.04$, which is rejected at the .01 level ($t = 7.38$). When no emphasize effect is assumed, the prediction is for about twice as many errors than were observed.

These analyses, given additivity of irrelevant strategies, show that an emphasize effect was operating in Angle + Angle Color Irrelevant and suggest that the effect was stronger than would arise if red and green trials had separate effects.

Applying the same reasoning to Angle + Angle Color, when the red and green contrast is assumed equal to the constant red emphasize effect, Eq. 6.3 now becomes,

$$c(\text{Angle} + \text{Angle Color}) = \frac{[r.m(A) + m(C_a)]}{[2r.m(A) + 2m(C_a) + m(I)]} \quad (6.10)$$

The predicted $c(\text{Angle} + \text{Angle Color})$ is .217, very close to the .225 observed ($-2 \ln(\lambda) = 0.02$), and not significantly different at the .05 level.

Predicted mean errors are $\bar{T} = (1 - .217)/(.217) = 3.61$, somewhat higher but not significantly different from the 3.45 observed ($t = 0.13$, $p > .05$). This prediction is closer than the one where red and green were assumed to have separate effects.

If no emphasize effect is assumed, Eq. 6.5 now becomes,

$$c(\text{Angle} + \text{Angle Color}) = \frac{[m(A) + m(C_a)]}{[2m(A) + 2m(C_a) + m(I)]} \quad (6.11)$$

Predicted $c(\text{Angle} + \text{Angle Color})$ is .204, not significantly different from the observed .225 ($-2 \ln(\lambda) = 0.23$, $p > .05$).

Predicted mean errors are $\bar{T} = (1 - .204)/(.204) = 3.90$ which is not significantly different from the 3.45 observed ($t = 0.38$, $p > .05$). The assumption of no emphasize effect cannot be rejected in Angle + Angle Color.

In this chapter, the Strategy Selection Theory was tested by predictions of the rate of learning during original learning and the degree of transfer. Additivity of relevant strategies was accurately predicted for two redundant and relevant cue groups. The predictions took account of emphasize effects. Additivity of irrelevant strategies was also predicted accurately, and in this case, it was necessary to take account of the emphasize effect to obtain an acceptable prediction. The degree of transfer was predicted in three ways: (1) number of Ss showing perfect transfer, (2) mean errors in transfer, and (3) cumulative distributions of error scores. All predictions were based on parameters which had been estimated from original learning data and independent groups. Predictions were accurate for seven of eight groups.

CHAPTER VII

DISCUSSION

Lawrence (1952) found that it is more efficient to train an S for n trials on an easy problem and then transfer him to a hard problem containing the same relevant dimension than to train from the beginning on the hard problem. The present experiments replicated Lawrence's result, and suggest that it depends upon stimulus emphasis.

If the relevant dimension is emphasized but not changed to make a problem easy, learning is facilitated and transfer is perfect. This produces an extreme form of the Lawrence effect. In the "transfer on a continuum experiment," the relevant dimension is emphasized and somewhat changed. Learning is facilitated and there is a small disruption in transfer. The net effect is to make the easy-to-hard program somewhat more efficient than the hard program alone. If the problem is made easy by introducing a redundant relevant cue which does not emphasize the final test cue, learning is facilitated but transfer is very slight, and no net gain in efficiency is realized. The experimental results on emphazier groups, difference in discriminanda (large Angle) group and the redundant-cue groups agree with the earlier results on these three kinds of easy-to-hard experiments as reviewed in Chapter I.

In the present study, color (red versus green) was more salient as a cue than the angle (30° versus 60°) or the emphasized (red colored) angle dimension. A distinction should be drawn between the color's saliency as a cue and its saliency as an emphazier. The weight of a cue may be reduced by reducing the discriminable differences between the values of the dimension, but the saliency of the dimension as an

emphasizer may not be impaired. For example, the color dimension might consist of a light red versus a dark red. The cue value of the color might be reduced since discrimination between the values is difficult, but the color would still "stand out" on the pattern and preserve the emphasize function. An extreme case of this is given in Red Angle, where there is no discriminable difference between the colors on the angles and an emphasize effect was observed.

The question arises as to whether the cues are hard to distinguish because they are (1) like the background (embedded) or (2) like one another (similar). An emphasize, like color, would seem to have its maximum effect when the cue to be learned is embedded but not similar to other cues. A reconsideration of the present problem suggests that the flower patterns, with fairly distinct but similar angles, were not optimal configurations for finding a large emphasize effect with color. The Angle problem was difficult because the two stimuli (angles) are physically similar and other irrelevant stimuli were present. However, the angles already stood out somewhat on the pattern and were separate from the other irrelevant dimensions, so that further emphasis was not of marked benefit. In Large Angle, where the difference between the discriminanda (angles) is doubled, the effect is more marked. Hull's (1920) embedded radicals, which were distinct from one another, apparently provided more opportunity for an emphasis effect of the red color.

The above discussion suggests that solving the problem requires

1. attending to the cue and
2. given that the cue is attended to, discriminating one cue from another.

Let the probability of attending to the cue be $P(A)$ and the probability of discriminating, given that it is attended to, be $P(D/A)$. Then the probability of discriminating is

$$P(D) = P(A) \cdot P(D/A).$$

If the cue is not attended to, then it cannot be discriminated, so that

$$P(D/\bar{A}) = 0.$$

The emphasize effect is on $P(A)$. In the case of Red Angle, the red color did not change the difference between the angles, it, then, does not presumably affect $P(D/A)$. The multiplier effect of the emphasize is on $P(A)$ while $P(D/A)$ remains constant.

One other interpretation would be that the emphasize adds a constant amount to the cue. This is, the angle would gain a constant amount due to the color. However, in the present study, the powerful color only increased $m(A)$ from .038 in Angle to .077 in Red Angle, whereas the angle color had a measure, $m(C_a)$, of .306 in Angle Color Control. Addition would give a prediction of a larger and wrong order of magnitude than observed, but a multiplying effect can give a value in line with the results. This result is taken as justification of the assumption that the measure of the set of strategies which arise from the emphasized dimension was multiplied by a constant larger than one. The addition of elements which are "more of the same" but "indistinguishable" from existing elements in the psychological field is referred to by Restle (1961c) as a case of "homogeneous classes" of stimuli. This is what North (1959) apparently was referring to by calling the filled-in triangles "enrichment of cues."

The precise rôle of an emphasize remains to be more fully investigated. In the present study, each problem was constructed to demonstrate some kind of emphasize effect. In three groups, the rôle was dual; in Angle + Angle Color, color was an emphasize and a relevant dimension; in Angle + Flower Color, color was a counter-emphasize and a relevant dimension; and in Angle + Angle Color Irrelevant, color was an emphasize and an irrelevant dimension. In each case, the rôles were experimentally confounded. By use of the Strategy Selection Theory, estimation of the

sets of strategies was made possible and predictions on the learning rate were accurate by assuming each rôle to be independent.

This assumption of independence may be open to question. It is possible for the rôles to interact. In Angle + Angle Color, the rôles might compete. The S might attend to color as a cue or to the colored angle as a cue. In Angle + Flower Color, the rôles might co-operate. Color was placed on a part of the pattern away from the angle (i. e., color was a counter-emphasizer) and served as a cue. S's attention might be diverted away from the angle and he might be more likely to use color as the basis for a correct strategy. In Angle + Angle Color Irrelevant, the rôles might compete, as they lead to opposite effects on the rate of learning. Since the group learned somewhat faster than Angle, the emphasizer rôle apparently predominated.

If the rôles are not independent, as assumed, then there remains the problem of finding a way to detect an interaction. Given the variability of the data and estimates, the predictions were not precise enough to show an interaction effect. If the color and angle cues were about equal in strength and problems constructed as in the present study, then a failure to make accurate predictions might indicate an interaction.

To study the distribution of attention over the parts of the stimulus pattern in a concept formation task, the use of a constant emphasizer seems promising. The constant emphasizer avoids some of the confounding of the rôles of the emphasizer but does not lead to an evaluation of the separate rôles. Suppose a problem were constructed of suitable complexity and contained two spatially separated relevant cues. The cues might be about equal in strength. A transfer of training design, similar to that in the present study, might be used to test effects on training. During original learning, one of the cues is emphasized by a constant and salient color. Transfer is to a problem with one of the cues and the color removed. If the color was on the cue which is removed, an

example of counter-emphasis with respect to the retained relevant cue is demonstrated. If the retained relevant cue on transfer was colored during original learning, an example of emphasis is demonstrated. Weights of the sets of strategies arising from each cue, emphasized or not emphasized, could be obtained from suitable control groups and then the amount of transfer predicted. Transfer tests are used to estimate the direction of the S's attention to cues and the degree of emphasis or counter-emphasis.

CHAPTER VIII

SUMMARY

A transfer-of-training design known as "easy-to-hard" transfer was used to study: (1) the rôle of attention, and (2) efficiency in concept formation. The degree of efficiency was hypothesized to depend upon stimulus emphasis and relationships of the original learning (easy) and transfer (hard) problems.

The stimuli were complex flower patterns and the correct responses (two-choice) depended upon one or two aspects of the pattern. The relevant dimension of the hard problem was the angle of the leaves to the stem of the flower.

Nine groups, of 20 Ss each, worked on different original learning problems and were all transferred to the same hard problem after criterion in original learning. A tenth group had the hard problem as its original learning problem and served as a control. All comparisons of acquisition and transfer were relative to this control group.

In two problems emphasis of the relevant angle dimension was achieved by either (1) doubling the difference between discriminanda, or (2) removing irrelevant cues during original learning. Both groups were highly efficient: they learned rapidly and showed nearly perfect transfer.

Color on the angle of the leaves to the stem constituted an "emphasizer." When a constant color was used on all trials, the effect was not strong; red had a detectable effect, but green did not facilitate learning at all. When color varied from trial to trial in a third problem, and was an irrelevant dimension, the net effect was slight facilitation of

learning. In these three groups, color could not serve as the basis for a correct strategy and transfer was perfect.

Two problems had color added as a redundant and relevant dimension during original learning. Both problems were learned faster than problems with only one dimension relevant, an example of "additivity of strategies." In one problem, the color was also an emphaser and transfer to the hard problem was somewhat positive. In a second problem, color was a counter-emphasizer, appearing over the flowers during original learning, and transfer to the hard problem was slightly negative.

Two control problems had color relevant and the angle dimension fixed. Color was found to be more salient as a cue than the angle. There was no evidence for transfer of an "observing response" to the angle in these groups.

The stochastic properties of the data were consistent with the expectations of the Strategy Selection Theory. Analyses of Ss' performances before criterion indicated that errors occurred at random with probability near one-half, constant and independent of how long S was in the pre-solution phase. Fitted theoretical error distributions yielded good approximations in eleven of twelve cases. There was some evidence that Ss use "wrong" as well as "irrelevant" strategies in the pre-solution phase. Since wrong strategies depend upon the same cues as correct strategies, it was predicted that estimates of the measure of wrong strategies would be about the same as estimates of correct strategies. This quantitative prediction was verified. Wrong strategies were detected and their measure correlated with the measure of correct strategies. Intercorrelations between practice, original learning and transfer problems indicated no stable individual differences.

By taking account of stimulus emphasis, and using the Strategy Selection Theory, the additivity of relevant strategies and additivity of irrelevant strategies were accurately predicted. The degree of transfer

was predicted in three ways: (1) number of Ss showing perfect transfer, (2) mean errors in transfer, and (3) cumulative distributions of error scores in transfer. All predictions were based on parameters which had been estimated from original learning data and independent groups. Seven of eight predictions on transfer were accurate.

Efficiency in concept learning was discussed in relation to the present and other findings. The question of the precise rôle of a stimulus emphasize was examined and further investigations on emphasizees suggested.

REFERENCES

- Archer, E. J., Bourne, L. E., and Brown, F. G. Concept identification as a function of irrelevant information and instructions. J. exp. Psychol., 1955, 49, 153-164.
- Baker, R. A. and Osgood, S. W. Discrimination transfer along a pitch continuum. J. exp. Psychol., 1954, 48, 241-246.
- Berlyne, D. E. Attention, perception and behavior theory. Psychol. Rev., 1951, 58, 137-146.
- Berlyne, D. E. Conflict, arousal and curiosity. New York: McGraw Hill, 1960.
- Blazek, N. C. and Harlow, H. F. Persistence of performance differences on discriminations of varying difficulty. J. comp. physiol. Psychol., 1955, 48, 86-89.
- Bourne, L. E., Jr. and Restle, F. Mathematical theory of concept identification. Psychol. Rev., 1959, 66, 278-296.
- Bower, G. H. Properties of the one-element model as applied to paired-associate learning. Tech. Rep. 31, Contr. Nonr 225 (17), Inst. for Math. Stud. in the Soc. Sci., Stanford Univ., 1960.
- Broadbent, D. E. Perception and Communication. London, New York: Pergamon Press. 1958.
- Bush, R. R. and Mosteller, F. A. A model for stimulus generalization and discrimination. Psychol. Rev., 1951, 58, 413-423.
- English, H. B. and English, A. C. A Comprehensive Dictionary of Psychological and Psychoanalytical Terms. New York: Longmans, Green. 1958.
- Eninger, M. U. Habit summation in a selective learning problem. J. comp. physiol. Psychol., 1952, 45, 511-516.
- Estes, W. K. Learning. In Ann. Rev. Psychol., 1956, 7, 1-38.

- Estes, W. K. The statistical approach to learning theory. In Koch, S. (Ed.), Psychology: A study of a science. Vol. 2. New York: McGraw-Hill. 1959, 380-491.
- Feller, W. Introduction to probability theory and its applications. (1st ed.). New York: Wiley. 1950.
- Gibson, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. Psychol. Rev., 1940, 47, 196-229.
- Guthrie, E. R. The psychology of learning. New York: Harper, 1935.
- Hammer, M. The role of irrelevant stimuli in human discrimination learning. J. exp. Psychol., 1955, 50, 47-50.
- Hara, K. and Warren, J. M. Stimulus additivity and dominance in discrimination performance by cats. J. comp. physiol. Psychol., 1961, 54, 86-90.
- Harlow, H. F. Studies in discrimination learning by monkeys: III. Factors influencing the facility of solution of discrimination problems by rhesus monkeys. J. gen. Psychol., 1945, 32, 216-227.
- Harlow, H. F. Learning set and error factor theory, in Koch, S. (Ed.) Psychology: A study of a science. Vol. 2, McGraw-Hill, New York. 1959, 492-537.
- Harlow, H. F. and Hicks, L. H. Discrimination learning theory: uniprocess vs. duoprocess. Psychol. Rev., 1957, 64, 104-109.
- Hayes, K. J. The backward curve: A Method for the study of learning. Psychol. Rev., 1953, 60, 269-275.
- Heidbreder, E. The attainment of concepts: II. The problem. J. gen. Psychol., 1946, 35, 191-223.
- House, B. J. and Zeaman, D. Transfer of a discrimination from objects to patterns. J. exp. Psychol., 1960, 59, 298-302.
- Hovland, C. I. A set of flower designs for experiments in concept formation. Amer. J. Psychol., 1953, 66, 140-142.

- Hughes, C. L. and North, A. J. Effect of introducing a partial correlation between a critical one and a previously irrelevant cue. J. comp. physiol. Psychol., 1959, 52, 126-128.
- Hull, C. L. Simple qualitative discrimination learning. Psychol. Rev. 1950, 57, 303-313.
- Hull, C. L. Quantitative aspects of the evolution of concepts. Psychol. Monogr., 1920, whole no. 123.
- James, W. The principles of psychology. Vol. 1, 1890. Dover Publications, Inc. (paperback), 1950.
- Kemeny, J. G., Snell, J. L. and Thompson, G. L. Introduction to finite mathematics. Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1957.
- Kendler, T. S. Concept formation. In Ann. Rev. Psychol., 1960, 12 447-472.
- Krechevsky, I. Hypotheses in rats. Psychol. Rev., 1932, 39, 516-532.
- Krechevsky, I. A study of the continuity of the problem-solving process. Psychol. Rev., 1938, 45, 107-133.
- Kurtz, K. H. Discrimination of complex stimuli: the relationship of training and test stimuli in transfer of discrimination. J. exp. Psychol., 1955, 50, 283-292.
- Lashley, K. S. Brain mechanisms and intelligence. Un. Chicago Press: 1929.
- Lashley, K. S. The mechanism of vision. XV. Preliminary studies of the rat's capacity for detail vision. J. gen. Psychol., 1938, 18, 123-193.
- Lashley, K. S. and Wade, M. The Pavlovian theory of generalization. Psychol. Rev., 1946, 53, 72-87.
- Lawrence, D. H. Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. J. exp. Psychol., 1949, 39, 770-784.
- Lawrence, D. H. Acquired distinctiveness of cues: II. Selective association in a constant stimulus situation. J. exp. Psychol., 1950, 40, 175-188.

- Lawrence, D. H. The application of generalization gradients to the transfer of a discrimination. J. gen. Psychol., 1955, 52, 37-48.
- Lawrence, D. H. The transfer of a discrimination along a continuum. J. comp. physiol. Psychol., 1952, 45, 511-516.
- LaBerge, D. L. and Smith, A. Selective sampling in discrimination learning. J. exp. Psychol., 1957, 54, 423-430.
- Moon, L. E. and Harlow, H. F. Analysis of oddity learning by rhesus monkeys. J. comp. physiol. Psychol., 1955, 48, 188-194.
- Murdock, J. The distinctiveness of stimuli. Psychol. Rev., 1960, 67, 16-31.
- North, A. J. Acquired distinctiveness of form stimuli. J. comp. physiol. Psychol., 1959, 52, 339-341.
- Prokasy, W. F., Jr. The acquisition of observing responses in the absence of differential external reinforcement. J. comp. physiol. Psychol., 1956, 59, 131-134.
- Restle, F. Additivity of cues and transfer in discrimination of consonant clusters. J. exp. Psychol., 1959, 57, 9-14.
- Restle, F. Discrimination of cues in mazes: A resolution of the "place-vs.-response" question. Psychol. Rev., 1957, 64, 217-228.
- Restle, F. Psychology of judgment and choice. New York: Wiley, 1961c.
- Restle, F. The selection of strategies in cue learning. Psychol. Rev. 1961a (in press).
- Restle, F. Statistical methods for a theory of cue learning. Psychometrika, 1961b, 26, 291-306.
- Restle, F. A theory of discrimination learning. Psychol. Rev., 1955, 62, 11-19.
- Restle, F. Toward a quantitative description of learning set data. Psychol. Rev. 1958, 65, 77-91.
- Schoeffler, M. S. Probability of response to compounds of discriminable stimuli. J. exp. Psychol., 1954, 48, 323-329.

- Spence, K. W. The differential response in animals to stimuli varying within a single dimension. Psychol. Rev., 1937, 44, 430-444.
- Spence, K. W. Continuous versus non-continuous interpretations of discrimination learning. Psychol. Rev., 1940, 47, 271-288.
- Thorndike, E. L. The psychology of learning. New York: Teachers College. 1914.
- Trabasso, T. R. Additivity of cues in the discrimination learning of letter patterns. J. exp. Psychol., 1960, 60, 83-88.
- Walker, H. M. and Lev, J. Statistical inference. New York: Henry Holt. 1953.
- Warren, J. M. Additivity of cues in visual pattern discriminations by monkeys. J. comp. physiol. Psychol., 1953, 46, 484-486.
- Warren, J. M. Perceptual dominance in discrimination learning by monkeys. J. comp. physiol. Psychol., 1954, 47, 290-292.
- Woodworth, R. S. and Schlosberg, H. Experimental Psychology, New York: Henry Holt. 1954.
- Wyckoff, L. B., Jr. The role of observing responses in discrimination learning. Part I. Psychol. Rev., 1952, 59, 431-441.

ROOM USE ONLY

MAY 25 1964

~~NOV 17 1964~~

~~DEC 1 1964~~

DEC 15 1964

JAN 6 1965

JAN 20 1965

~~FEB 2 1965~~

~~FEB 15 1965~~

MAR 2 1965

~~MAR 15 1965~~

APR 1 1965

MAY 5 1965

May 5/9/65

7/2/67 gh.

MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 03143 1699