INFORMATION TO USERS

This material was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.

2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in "sectioning" the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.

4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from "photographs" if essential to the understanding of the dissertation. Silver prints of "photographs" may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

Xerox University Microfilms
300 North Zeeb Road
Ann Arbor, Michigan 48106

7813656

FRIED, LISBETH JANE SOSS
PERCEPTUAL LEARNING AND CLASSIFICATION WITH ILL-DEFINED CATEGORIES.

THE UNIVERSITY OF MICHIGAN, PH.D., 1978

University Microfilms International 300 N. ZEEB ROAD, ANN ARBOR, MI 48106

PERCEPTUAL LEARNING AND CLASSIFICATION WITH ILL-DEFINED CATEGORIES

by
Lisbeth Jane Soss Fried

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Psychology)
in The University of Michigan
1978

Doctoral Committee:

Professor David H. Krantz, Chairman
Assistant Professor Keith J. Holyoak
Professor J. E. Keith Smith
Professor Wilson P. Tanner, Jr. (deceased)
Professor Michael Woodroofe

TO

WILSON P. TANNER, JR.

ACKNOWLEDGMENTS

It is with gratitude that I acknowledge here the many people without whose aid and support this dissertation would not have been completed. First I would like to thank Spike Tanner who picked me up when I was down, and who provided direction and encouragement throughout the initial stages of the research. I only regret that I did not express this to him while he lived.

I would like to thank Mike Woodroofe who, in many hours of discussions, guided me in the formulation of the model, although I take full responsibility for all errors of commission and omission. I will also always be grateful to Keith Holyoak for asking if the model could be applied to prototype learning. That question led to a whole new fascinating area of research that I had not previously considered, as well as leading to the two final experiments. However, without Dave Krantz' vision of excellence, encouragement, determination and prodding this dissertation would never have achieved its final form. I am very grateful to him; he is always able to see the whole forest when I become lost among the trees.

I would also like to thank J. E. Keith Smith for many helpful consultations; and John Patterson who wrote the computer program which generated the visual patterns and which then displayed them to the subjects.

Thanks are also due to Lih-wen Huang who helped to specify the equations in the model, to Clayton Lewis who suggested the use of transparencies in the first two experiments, and to Diedre McKinney who ran the subjects in experiment three.

It is with loving gratitude that I especially acknowledge the unending support, encouragement and assistance of my husband, Michael, who patiently put up with me, my anxieties, and countless dinners out.

Thank you Carrie and Aaron too, for the many hours you put in at babysitters and at summer camps when you would have preferred your home.

TABLE OF CONTENTS

DEDICATION. ii

ACKNOWLEDGMENTS iii

LIST OF TABLES. vii

LIST OF FIGURES viii

CHAPTER

I A COMPARISON OF A PSYCHOPHYSICAL AND AN INFORMATION PROCESSING APPROACH TO THE PROBLEM OF PERCEPTUAL CATEGORIZATION WITH ILL-DEFINED CATEGORIES 1

The Problem of Recognition With Ill-Defined Categories. 1
The Nature and Structure of Ill-Defined Categories. 3
Category Learning. 5
Decision Rules for the Classification of Novel Stimuli. 14
Learning Without Feedback. 18
A Gaussian Ideal Learner 20
Statement of Purpose 22

II EXPERIMENT 1. 24

Specific Predictions and Hypotheses. 24
Method 30
Results. 39
Discussion of Results. 49

III EXPERIMENT 2. 53

Statement of Purpose 53
Method 55
Results. 56
Discussion of Results. 71

CHAPTER

IV EXPERIMENT 3. 74

Statement of Purpose 74
Method 75
Results. 80
Discussion 90

V EXPERIMENT 4. 93

Predictions and Hypotheses 93
Method 98
Results. 102
Discussion of Results. 111

APPENDIX. 118

REFERENCES. 120

LIST OF TABLES

TABLE

1 AVERAGE INCREASE IN EXPECTED VALUE OF CLASSIFICATION BOUNDARIES FOR EACH SUBJECT AND THE MODEL OVER ALL EQUAL VARIANCE CONDITIONS. . . . 42

2 NUMBER OF OBSERVATIONS TAKEN BY EACH SUBJECT AS A FUNCTION OF d' 45

3 PRESENCE OR ABSENCE OF CLASSIFICATION INVERSIONS UNDER COMBINED VARIANCE d' MANIPULATIONS 48

4 AVERAGE INCREASE IN EXPECTED VALUE OF CLASSIFICATION BOUNDARIES FOR EACH SUBJECT AND THE MODEL OVER ALL EQUAL VARIANCE CONDITIONS. . . . 59

5 MEAN NUMBER OF OBSERVATIONS AND MEAN STANDARD DEVIATION AMONG UPPER AND LOWER BOUNDARIES FOR EACH SUBJECT IN EACH SET OF EXPERIMENTAL CONDITIONS. 61

6 NUMBER OF STANDARD DEVIATIONS BETWEEN MIDPOINT OF SUBJECT'S UNCERTAIN REGION AND ACTUAL MIDPOINT BETWEEN THE DISTRIBUTIONS, FOR EQUAL PRIOR ODDS AND VARIANCES. 65

7 PRESENCE OR ABSENCE OF CLASSIFICATION INVERSIONS WITH UNEQUAL VARIANCES 70

8 MEAN NUMBER OF TRIALS TO CRITERION AND PERCENT CORRECT AT EACH DISTORTION LEVEL. 81

9 STIMULUS-RESPONSE CONTINGENCY TABLE FOR THE TRANSFER TASK 84

10 MEAN NUMBER TRIALS TO CRITERION AND PERCENT CORRECT AS A FUNCTION OF DISTORTION LEVEL . . . 103

11 STIMULUS-RESPONSE CONTINGENCY TABLE FOR EACH DISTORTION LEVEL AT TRANSFER. 105

LIST OF FIGURES

FIGURE

1 SIZE OF UNCERTAIN RANGE AS A FUNCTION OF PAYOFFS. 40

2 SIZE OF UNCERTAIN RANGE AS A FUNCTION OF d'. . . 44

3 CHANGES IN MIDPOINT OF UNCERTAIN RANGE AS A FUNCTION OF PRIOR PROBABILITIES. 46

4 SIZE OF UNCERTAIN RANGE AS A FUNCTION OF PAYOFFS. 57

5 SIZE OF UNCERTAIN RANGE AS A FUNCTION OF d', MEANS KNOWN. 63

6 SIZE OF UNCERTAIN RANGE AS A FUNCTION OF d', MEANS UNKNOWN. 64

7 MIDPOINT OF UNCERTAIN RANGE AS A FUNCTION OF PRIOR PROBABILITIES, MEANS KNOWN 67

8 MIDPOINT OF UNCERTAIN RANGE AS A FUNCTION OF PRIOR PROBABILITIES, MEANS UNKNOWN 68

9 THE TWO PROTOTYPES USED IN EXPERIMENTS 3 and 4 . 77

10 PERCENT OF PATTERNS CALLED "OTHER" AS A FUNCTION OF THE DISTANCE FROM THE TRUE PROTOTYPE. 86

11 MEAN PERCENT OF PATTERNS CALLED "OTHER AS A FUNCTION OF LIKELIHOOD RATIO 89

12 A DESCRIPTION OF THE DIFFERENCE IN PREDICTIONS REGARDING THE CLASSIFICATION OF PATTERNS a, b, AND c FOR LOW AND HIGH VARIANCE DISTRIBUTIONS WHEN THE CATEGORY "OTHER" IS AVAILABLE 95

13 PERCENT PATTERNS CALLED "OTHER" AS A FUNCTION OF DISTANCE FROM THE TRUE PROTOTYPE. 107

14 PERCENT CALLED "OTHER" AS A FUNCTION OF ABSOLUTE LOG LIKELIHOOD RATIO 109

CHAPTER I

A COMPARISON OF A PSYCHOPHYSICAL AND AN INFORMATION PROCESSING APPROACH TO THE PROBLEM OF PERCEPTUAL CATEGORIZATION WITH ILL-DEFINED CATEGORIES

The Problem of Recognition With Ill-Defined Categories

The problem of recognition is basic to a theory of perception and as such has been an integral part of both psychology and psychophysics. Neisser (1967) traced its long history in psychology: in Behaviorism it is known as the problem of stimulus generalization or stimulus equivalence; in Gestalt psychology it is the "Hoffding step." How is it, for example, that so many different visual configurations are all called "A" (Neisser, 1967)? How is it that only twelve color names suffice for over a million hue-brightness-saturation combinations (Bruner, et al., 1967)? While Reed (1973) states that the task of pattern recognition theories is "to account for our ability to treat different exemplars of a pattern as belonging to a common category" (p. 6), Bruner (1957) is more grandiose, stating that "all perceptual experience is necessarily the end product of a categorization process."

To Luce (1963) the problems of recognition and discrimination, along with the problems of detection and scaling, delineated the whole field of psychophysics. Here, however, the recognition problem is the problem of classification--which of the many possible alternatives is it? This view may be due to the influence of signal detectability theory (Tanner and Swets, 1954; Swets, Tanner and Birdsall, 1961) which introduced the idea that decision-making and perception are intimately related. Perceptual processes such as detection and recognition were shown to be dependent upon such environmental (or experimental) contingencies as the prior probabilities of an event and the payoffs for correct responses, as well as on the stimulus itself. Recognition was still the problem of discriminable events being categorized into equivalence classes, but no longer for purely perceptual reasons. Unfortunately, this approach to the problem of recognition has not been incorporated into many cognitive models, perhaps, because it was designed to account for choice behavior at asymptote, rather than to provide a model of perceptual learning.

Two classes of solutions for the problem of recognition provided by cognitive psychologists themselves were template-matching and feature analysis. Writing in 1967, however, Neisser felt that while these

theories were adequate in accounting for the recognition of alpha-numeric characters where clear-cut criteria of category membership exist, most such categories are not so well defined.

> Ill-defined categories are the rule, not the exception, in daily life. The visual distinctions between dogs and cats, or between beauty and ugliness, are ill-defined, just like the conceptual differences between creative science and hack work, or health and neurosis. So are the EEG patterns which indicate a particular stage of sleep, the x-ray shadows which suggest a tumor, the style of painting which identifies a Picasso, or the features which continue to characterize the face of a friend through the years (p. 58).

Theories which would account for classification decisions into these types of categories were only just being developed at the time of his writing.

The Nature and Structure of Ill-Defined Categories

During the past decade, however, a large body of literature appeared in both psychophysics and cognition which attempted to account for the learning of ill-defined categories and their boundaries. The foundation of both approaches lay in their assumptions regarding the nature and structure of categories. As is well known, the signal detectability theory approach assumes, in its most basic case, that each category is made up of a population of perceptual events for which

the likelihood of such events is normally distributed on the decision axis (Swets, Tanner and Birdsall, 1961; Green and Swets, 1966; Tanner and Swets, 1954; Luce, 1963). The degree of confusability among the categories is represented by the amount of overlap between the distributions.

Surprisingly, the theory of natural (rather than sensory) categories, as expressed by Rosch and her co-workers (Rosch, 1973; 1977; Rosch and Mervis, 1975), is remarkably similar. Natural categories were found to be continuous. Moreover, category members were found ordered according to the degree to which they were judged good exemplars of the category, with the most typical or focal exemplars clustered in the center, atypical exemplars arranged around the periphery and those in between forming a continuous gradient of category membership (Rosch and Mervis, 1975). This was found true of perceptual categories such as colors and forms as well as such semantic categories as games and mammals. In addition, others have found that observers reliably rate degree of membership in such descriptive categories as "small," "big," etc. and that membership in these categories is also found to form a continuous gradient from most to least typical (Hersh and Carmezza, 1976). Thus the membership function of natural

categories might be approximated by a normal distribution on a continuum, the most typical members representing the focal point, or the central tendency of the category. Degree of typicality and degree of membership decrease as a continuous function of the distance from the center of the category.

Category Learning

Psychophysical Approach

In spite of the apparent similarity, noted above, in the basic assumptions regarding the nature and structure of categories, there are sharp differences between psychophysicists and the more cognitive psychologists. Signal detectability theory assumes that it is not the perceptual events themselves which are necessarily normally distributed, but rather the likelihood of the event given the category, which is distributed normally. The observer in signal detectability theory is viewed as sampling an observation fron one of two normally distributed populations of likelihoods and acting in a way which is mathematically equivalent to the computation of likelihood ratio (Green and Swets, 1966; Swets, Tanner and Birdsall, 1961). Since computation of likelihood ratio presupposes knowledge of the distribution parameters in question, these parameters

are assumed known in advance. Moreover, since these distribution parameters are also assumed to be constant, pre-asymptotic variability in performance is presumed (by current investigators) to result solely from shifts in the decision criteria with feedback and not to shifts in estimates of the distribution parameters themselves (Dorfman and Biderman, 1971; Dorfman, Saslow and Simpson, 1975; Kac, 1962; and Thomas, 1973, 1975). These were assumed fixed and known to the observer at the start.

Accordingly, models of the learning process which these researchers developed were models of criterion shift as a result of feedback, basically extensions of Kac' (1962) model for learning signal detection. As in classical signal detectability theory Kac assumes two distributions completely known to the observer, but, in addition, he assumed an arbitrarily chosen likelihood ratio decision criterion which he called 'threshold'. In this model if the response is correct on any given trial the 'threshold' (criterion) is retained for the next observation. If in error the 'threshold' is corrected; that is, if signal is present and an error is made the 'threshold' is lowered a specified amount, while if the signal is absent and an error is made the 'threshold' is raised that same amount. For normally distributed observations of equal variance, with symmetric payoffs

and prior odds, the process becomes stationary at the optimal likelihood ratio criterion midway between the two distributions. This model is quite vulnerable to empirical testing since it is completely deterministic, the probability of a criterion shift being 1.0 after an error and 0.0 after a correct response; and is completely symmetric, the amount of shift being constant whether a miss or a false alarm occurs. Moreover, the direction of shift is always away from the mean of the distribution being sampled. Larkin (1971), however, found criterion shifts after only 30-50 percent of the error trials instead of after all of them, while 5-30 percent of all criterion shifts were after correct responses when there should have been none. Thomas (1973), employing a theorem of Norman (1972), showed that this pure error correction model of Kac' implies probability matching at asymptote. Larkin did not find probability matching; which argues against the pure error correction mode. The lack of probability matching may have resulted from the large number of shifts after correct responses.

Recent investigators (Dorfman and Biderman, 1971; Dorfman, Saslow and Simpson, 1975; and Thomas, 1973) have found it fruitful to generalize Kac' model and to test concomitant results. In the general model of Dorfman, et al. (1971) the possibility of a shift

after correct responses as well as after errors is recognized. The probability and amount of each type of shift are left as parameters to be estimated from the data. The pure error correction model of Kac' was found to give a relatively poor fit to the data (Dorfman, et al., 1975), but it was found to be superior to two other special cases of the model--one in which no shifts were assumed after error, but only after correct responses; and the other, in which the amount of shift after an error or after a correct response were assumed constant but not necessarily equal to each other. The general model gave a significantly better fit to the data than either of the special models, of course, since it had more degrees of freedom. Moreover, shifts in criteria after correct responses were estimated to be quite small and non-existant at asymptote. The authors concluded that an approximate error correction model with very small changes in criterion after correct responses and with pure error correcting at asymptote would be the most reasonable description of their data. It is interesting to note that a static criterion was not apparent even at asymptote.

Kubovy and Healy (1977) offer a simple model of an Ideal Learner in a signal detection task which takes a different approach to the learning process than that taken by previous investigators. In

their model the distribution means are not assumed known to the observer at the start and it is the task of the Ideal Learner to form his decision criterion by first estimating these means. In their experimental setting the observer is given feedback after each response so he is always aware of the source of each observation. Since their stimuli were numerical (heights of men and women in centimeters), the Ideal Learner easily keeps a separate running tally of the averages of the male and female heights and uses them as his best estimates of the means of the two distributions. Only equal prior probabilities, equal variances and symmetric payoffs were used in their experiment so that the optimal single decision cutoff is always the midpoint between the two estimated means.

Their model gave a better fit to their data than the pure error correction model of Kac, but shifts were not always in the predicted direction (i.e., toward the mean of all the observations sampled), and occasionally their subjects appeared to them to be acting irrationally by shifting up after low observations and down after high ones. The Gambler's Fallacy was cited as a possible explanation for these discrepancies.

Since only equal prior probabilities, equal variances and symmetric payoffs were used in their

experiment, the optimal decision criterion was always the running average of all the observations sampled. Because of this, the experimenters were unfortunately prevented from achieving an adequate test of their "Ideal Learner" model. The Ideal Learner is expected to keep separate estimates of distribution means, but this is actually wasted effort. Indeed, observers could have different and erroneous conceptions of these population means; but as long as their cutoff remained their best estimate of the running average of all the observations, they would maximize their expected value. Thus their experiment is actually another test of a decision criterion model, albeit one not dependent on feedback; and one cannot determine from it whether observers do indeed attempt to estimate distribution parameters or if they in fact merely concentrate on learning an optimal decision criterion.

The Pattern Recognition Approach

The pattern recognition literature offers an entirely different conception of what is being learned during category learning than is held by those operating from a signal detection point of view. Called schema or prototype theory, it was first set forth in 1932 by Bartlett (1932) who maintained that the common elements, or central tendency, of a series of perceptual events

was somehow abstracted and stored. Oldfield (1954) also believed that since the individual exemplars of a category could not all be remembered, economy of storage was achieved by filtering out and storing only the common elements, i.e., the central tendency of the category or the 'schema'.

A great deal of experimental evidence found support for this notion. In a task involving random visual forms Attneave (1957) discovered that observers trained on a class of stimuli were likely to learn three characteristics of the class: 1) its central tendency, 2) the relevant dimensions of difference, and 3) the degree of dispersion on these dimensions. Using random dot patterns Posner and Keele (1968, 1970) found that following an initial learning stage, the category prototypes (representing the central tendency of the categories) were classified more accurately and with faster reaction times than other novel category exemplars even though they too, had not been seen before. Posner and Keele concluded that the prototypes had to have been abstracted during the learning process for this to occur. This pattern of results was found to exist even after a week's delay, supporting Oldfield's theory of economy of storage.

Following an initial stage of learning a series of geometric designs, Franks and Bransford (1971) asked subjects to state their confidence during a transfer task that members of a second series of patterns were 'old' or 'new'. Unknown to the observer all the patterns in the transfer task were actually new. The patterns which were rated 'old' with the highest degree of confidence were the previously unseen category prototypes. Other novel patterns were rated 'old' only according to the degree of their physical similarity to the prototypes. Since all the patterns in the transfer task were actually new, a process of prototype abstraction was inferred.

Peterson, Meagher, Chait and Gillie (1973) also found that observers abstract a central prototype except when variability within the categories was so great that "category names were learned in a paired associate manner" (p. 386).

Homa and his colleagues (Homa, Cross, Cornell, Goldman and Schwartz, 1973; Homa and Chambliss, 1975; Homa and Vosburgh, 1976; and Goldman and Homa, 1977) also found in numerous experiments that correct classification of novel category prototypes occurs during transfer and, therefore, must be learned during the observation process. They found percent correct classifications of novel prototypes increased with increases in

the number of exemplars within a category and decreased with the number of categories to be learned. These results indicate that the central tendency of a category may somehow be abstracted during category learning.

Unexplained results indicate, however, that a theory based on abstraction of the central tendency alone is inadequate. Posner and Keele (1968; 1970) found that the greater the degree of variability among the exemplars within a category the greater the number of confusions and the greater the number of trials to criterion. Even more surprising to the authors was the finding that the group trained with a moderate amount of category variability actually made fewer errors upon transfer than did the group trained with a low amount of category variability--even though the latter group learned more quickly and with fewer errors. The low variability group was found more likely to mis-label a pattern as random when it was actually a member of the category than was the moderate variability group. The authors had expected the opposite results, reasoning that those trained on the low variability category would obtain a more accurate perception of the prototype and therefore have less difficulty in sorting novel items. The authors concluded that observers were learning something about category variability; but it was

not suggested how knowledge of category variability would produce these results.

Besides the unexplained effect of category variability, similarity between the prototypes was also producing unexpected results. Reed (1972) found percent of correct classifications of a previously unseen prototype decreased unexpectedly with the degree of similarity between them. Reed had predicted that category prototypes would always be correctly identified if they were in fact correctly abstracted, but he suggests that "the greater similarity of the two prototypes . . . made it difficult for Ss to use a prototype decision rule, causing them to use a decision rule requiring very little integration of information" (p. 391). Reed thus abandons the theory of prototype abstraction in this situation of similar prototypes. Goldman and Homa (1977) also found percent correct classifications of novel category prototypes to decrease with an increase in category overlap, again demonstrating that category prototypes are not always correctly identified.

Decision Rules for the Classification of Novel Stimuli

The effects of category variability, of prototype similarity, and of category overlap described in

the previous section were unpredicted and unexplained only because they contradict the decision rule currently a cornerstone of prototype abstraction theory. The decision model assumed by pattern recognition theorists was set forth explicably by Reed (1972, 1973). Called the "prototype model," it is based on the work of Shepard (1962a; 1962b) and Kruskal (1964) which provided a method of converting similarity measures into measures of psychological distance. The prototype decision rule is to "'decide a pattern X_j belongs to Category 1 if the distance from it to the Category 1 prototype is less than or equal to the distance from it to the Category 2 prototype. Decide x_j belongs to Category 2 otherwise'" (Reed, 1972, p. 387). The prototype is defined here as the central tendency of the category. Found to be the model which best described his data against several competitors, Reed determined that "the predominent strategy [among his subjects] was to form an abstract image or prototype to represent each category and to classify test patterns on the basis of their similarity to the two prototypes" (1972, p. 401). The descriptive validity of this decision rule has not been questioned, either implicitly or explicitly, in any of the studies reviewed thus far.

This distance from the prototype model cannot explain the results cited above, however, as it is readily apparent that the prototype is always closest to itself and, hence should always be correctly classified if it is indeed correctly abstracted. This prediction should not be affected by category variability, category overlap or even degree of prototype similarity, as long as the prototypes are distinct.

Moreover, if distance from the prototype is the only relevant variable, two stimuli which are the same distances from the prototypes should always be classified into the same category independent of category variability. That this was not the case was shown in the work of Posner and Keele (1968; 1970) as discussed above. The same pattern was classified as random by the low variability group and as a member of the true category by the moderate variability group even though both groups were trained with exactly the same prototypes. Although the authors did conclude that subjects had learned the variability of the category they did not suggest how the observer's rule had been modified to make use of this information.

The decision rule for the classification of novel stimuli developed by signal detectability theory, on the other hand (Swets, Tanner and Birdsall, 1961;

Green and Swets, 1966), is quite different. It is not based on distance from the central tendency, but rather the observer is assumed to compute an optimal likelihood ratio decision criterion based on prior probabilities and payoffs and to classify each observation according to whether its likelihood ratio is greater or lesser than this predetermined optimal criterion. The likelihood ratio criterion is viewed as optimal in that it maximizes the observer's expected value (Green and Swets, 1966). As pointed out earlier, current theorists maintain that this optimal criterion is not predetermined but learned during the observation process as a result of feedback.

A likelihood ratio decision rule, predicts the apparently anomalous results found in the category learning literature reviewed above, since the likelihood of an observation is a function of category variability, and the ratio of the likelihoods is a function of both category overlap and prototype similarity. An optimal likelihood ratio criterion would also take into account the prior probabilities of the categories themselves as well as the penalties for errors--all factors which are irrelevant to a strict distance from the prototype model.

The signal detectability theory approach differs markedly from the approach of information processing

models of category learning. While both points of view assume (either implicitly or explicitly) that the categories can be described as normal, overlapping distributions, they disagree completely regarding what is learned as well as regarding the nature of the decision rule. Current signal detectability theorists assume that an optimal likelihood ratio decision criterion is learned as a function of feedback and that a likelihood ratio decision rule is used. They also base their theories on the assumption that the distribution parameters of the observation process are known in advance. On the other hand, cognitive psychologists theorize that a category prototype is abstracted during the learning process and that decisions are based on the city-block or Euclidean distance to this prototype.

Learning Without Feedback

Although the psychophysical and the cognitive approach differ in their theories of what is being learned and what decision rule is employed, all of the models have depended on feedback for the learning process to occur. A model of an ideal learner learning the central tendency and variability of a category without benefit of feedback has not been developed, although several non-optimal models of criterion shift in a

non-feedback situation have been proposed (Schoeffler, 1965; Thomas, 1977). That such a model is necessary is attested to by the fact that previous researchers have found performance in feedback and no feedback conditions to be nearly equivalent. In a signal detection task Tanner, Rauk and Atkinson (1970) found hit and false alarm rates comparable in both the feedback and no feedback conditions. Carterette, Friedman and Wyman (1966) found no significant difference in percent correct between the feedback and no feedback conditions in either a hard or an easy discrimination task except that percent correct was higher in the no feedback condition for the easy discrimination task. A 1964 study by Levine, Leitenberg and Richter also showed learning to be faster in the no feedback condition.

Moreover, the ability to engage in perceptual learning without requiring feedback is at the heart of the theory of perceptual learning set forth by Gibson (1953, 1969; Gibson and Gibson, 1955).

> Perceptual learning refers to an increase in the ability to extract information from the environment as a result of experience and practice with stimulation coming from it. . . . Perceptual learning is self-regulating, in the sense that modification occurs without the necessity of external reinforcement. (Gibson, 1969; p. 3-4).

A general model of an ideal learner should therefore account for category learning under conditions

of no feedback as well as under feedback, unless it can be demonstrated that entirely different types of learning occur in the two situations.

A Gaussian Ideal Learner

The model to be developed, a Gaussian Ideal Learner (GIL), is an attempt at such a general model of category learning. It attempts to account for the learning of ill-defined categories independent of the particular modality used. Moreover, it is designed to account for category learning without external feedback.

GIL is a model for the learning of normally distributed ill-defined categories. It makes the basic assumption of signal detectability that categories can be represented as overlapping populations of stimulus events normally distributed on a continuum. This is consistent with the findings of Rosch and her colleagues (Rosch, 1973, 1977; Rosch and Mervis, 1975) that natural categories are analog in structure, overlapping, and can be ordered in degree of typicality. It makes the further assumption not made in signal detectability theory that it is the perceptual events themselves which are normally distributed rather than their likelihoods. Although this is admittedly restrictive it permits the assumption that the categories themselves and hence the category parameters are actually objective phenomena and can be discovered by

a rational observer. Signal detectability theorists, in not wishing to restrict applications of the theory to normally distributed categories, were content to assume that only the likelihoods were normally distributed. This assumption seems too restrictive since, for reasons explained above, it requires the parameters of these distributions to be known to the observer at the beginning of the observation process. Future research can perhaps develop a model where weaker assumptions regarding the shape of the distributions are made and in which aspects of the shape itself are learned during the observation process.

GIL assumes that observers sample observations randomly from a mixture of exemplars from two unknown, overlapping normally distributed categories. These observations are then used by the observer to estimate the mean (prototype) and the variance of each category. If the proportion of exemplars from each is not known, that is assumed to be estimated also. (Details of a possible estimation process are outlined in the appendix.)

The observer is assumed to use his current opinion about these category parameters, along with his payoffs for correct and incorrect responses, to form an optimal decision criterion. This criterion is assumed to be based on likelihood ratio, the ratio of the likelihood that the observation is a member of one category to the likelihood that it is a member of the other. As stated earlier, a

decision criterion based on likelihood ratio has the advantage of taking into account the distance between the central tendencies of the two categories as well as the variability of each, an advantage which a decision rule based purely on distance to the prototype does not have. This decision criterion is then used to classify the observation. Estimates of each category mean and variance, as well as of the proportion of each category in the mixture, is modified after each observation when necessary. Since each such modification requires a subsequent revision of the decision criterion, preasymptotic variability in performance is assumed to result from changes in opinion about category parameters and not merely from shifts in the decision criteria with feedback. Moreover, this entire estimation-reestimation process is assumed to occur without benefit of feedback. Classifications are assumed to be poor in the beginning, but to improve with each observation until an asymptote is reached.

Statement of Purpose

It is the purpose of this research to test the dynamic cutoff models of signal detection theory as well as the prototype abstraction models of cognitive psychology and to compare them with GIL, a model of category learning which does not depend on feedback. It is hypothesized that category learning consists of learning

both the central tendency and the degree of variability of the category (the two sufficient statistics of a normal distribution) and that these should not be considered to be known in advance of the observation process even in traditional psychophysical experiments. Second, it is hypothesized that estimates of both of these distribution parameters is used by the observer to develop likelihood ratio decision criteria--and that it is these changes in his parameter estimates which cause his criteria to fluctuate. Third, it is predicted that novel stimuli will be classified according to a likelihood ratio decision rule and not according to their linear distance from the prototype. It is also predicted that ill-defined categories can be learned without external feedback.

CHAPTER II

EXPERIMENT 1

Specific Predictions and Hypotheses

To test the descriptive applicability of GIL it was necessary first to develop a task which would permit changes in the observer's estimates of the parameters of the underlying distributions to be interpreted. Accordingly, a payoff matrix was developed which included the response category: "Uncertain" in addition to the two traditional alternatives of "Category 1" and "Category 2." The three alternatives force the selection of two decision cutoffs. Even though a fine of one cent was levied for each response in the uncertain category, large negative payoffs for errors and small positive rewards for correct responses make the 'uncertain' response optimal for posterior probabilities often far in excess of absolute uncertainty of 50-50. Indeed, payoffs could be manipulated so that posterior probabilities of 90 percent could still elicit optimal 'uncertain' responses. Thus each payoff matrix has an

optimal minimal posterior probability associated with it such that at that posterior probability or at greater ones a classification should be risked.

For a specific payoff matrix, then, and for equal variances and equal prior odds it can be shown that the number of standard deviations comprising the subject's uncertain area should vary only with his perceived distance between the means of the distributions. If d' is perceived to be small the number of scores falling in the uncertain range will necessarily be large because the optimal minimum posterior probability necessary will only be achieved with scores located at the extremes of the distributions. As perceived d' becomes larger this uncertain range should necessarily become smaller as intervening observations will have larger likelihood ratios associated with them and therefore larger posterior probabilities. The larger perceived d', then, the fewer should be the number of observations labeled uncertain while the smaller d', the greater should be the number of observations labeled uncertain.

Uncertain regions which vary with actual d' in the predicted direction will therefore indicate that the observer has learned the relative distances between the means. Since observers will not receive external

reinforcement they must learn the appropriate decision criteria through the decomposition process alone. The Kubovy and Healy (1977) Ideal Learner model specifies one can learn a single optimal decision criterion without feedback and without knowledge of the underlying distribution means by merely averaging the values of the observations. However, two optimal decision criteria cannot be discovered in this manner; the underlying distributions themselves must be learned.

Additional predictions are also testable through the use of this three alternative payoff matrix. A second prediction, of course, is that the decision-maker will maximize his expected value; i.e., his use of the "uncertain" category should vary directly with the severity of the penalty for incorrect responses. In other words, for matrices with large penalties, optimal posterior odds of being correct may need to be as high as 4 or 5 to 1 before a classification should be made; while, for matrices with lesser penalties for errors, posterior odds of 2:1 may be optimal. Thus there should be a wider range of scores labeled "uncertain" when there are stiff penalties for errors than when penalties for error are slight. Uncertain regions which vary appropriately with payoffs will indicate the observer is maximizing his expected value.

Thirdly, it is predicted that decisions will reflect sensitivity to the prior odds of the distributions being sampled. In the case of equal variance distributions sensitivity to prior odds will, for optimal decision-makers, result in the midpoint of the uncertain regions being shifted toward the mean of the less likely distribution, the amount of shift being a function of the extent of deviance from prior odds of 1:1.

Fourthly, it is predicted that decisions will reflect sensitivity to differences in the variances of the distributions being sampled. It is hypothesized that if the variance of one distribution is much larger than that of the other, and if the means of the two distributions are close enough together to warrant it, inversions will occur in the classifications made. That is, high and low scores will be classified as having come from one distrubution, while intervening scores will be classified as having come from the other. (A decision rule based only on distance to the prototype [mean] predicts no such inversion.) Instead of only one uncertain region, two separate uncertain regions will necessarily result. Although the existence of ROC curves dropping below the negative diagonal can be used to infer unequal variances in signal detection experiments (Green and Swets, 1966, p. 63), sensitivity to differences

in variances among objective distributions has not previously been examined.

A fifth prediction, and perhaps the most important, is that learning will occur, i.e., decisions will improve with the number of observations made even though feedback will not be used. The expected value of decisions is predicted to increase from initial decisions to final ones. Although it is suggested above how a simple averaging technique will work is some situations without feedback when only one decision criterion is required, it will not work when two decision criteria are needed. If an increase in expected value occurs here, without feedback, the basic hypothesis of the Dynamic Cutoff Models--that learning occurs through a shift in decision criterion after an error--will not be adequate to explain the results.

A test of all these predictions is facilitated by the use of numerical stimuli and by the use of verbally reported decision criteria. In agreement with the reasoning of Kubovy and Healy (1977), Kubovy, Rapoport and Tversky (1971), and many others (Hammerton, 1970; Lee & Janke, 1964; and Ward, 1973), stimuli were integers sampled randomly from a discrete approximation to a

normal probability density function constructed over the set of positive integers from zero to 999. Unlike the stimuli used in signal detection tasks or recognition memory experiments these stimuli are familiar, recognizable and not confusable. Moreover, being objective both the experimenter and the research participant see the same stimulus; and, more importantly, the parameters of the distributions can be independently manipulated.

A second advantage of numerical stimuli is that their use permits the research participants to report verbally their criterion boundaries, rather than requiring the experimenter to derive them from an examination of hundreds of decisions. Since Kubovy and Healy (1977) found no essential difference in decision criteria between the group verbally reporting their decision cutoffs and those of a second group whose cutoffs were derived from decisions, the research participants in this experiment were allowed to state their decision criteria after each response.

The concern that the cutoffs subjects report may not be the ones they actually use classifying the stimuli is obviated in this situation, since subjects knew that they were being paid only according to their decision cuttoffs--not according to

the decisions themselves. Upon achieving asymptote, each subject's boundaries were used to classify an additional fifty observations. They were then paid according to how well these fifty scores were classified--not according to the result of any of their previous decisions. To insure that observers made their best decision cutoffs as they went along, however, there was a chance (one in 50) on each trial that the session would stop and the subject's current decision boundaries be used to classify his next fifty scores.

Method

Experimental Design

Five experimental variables were manipulated:

1. Number of unknown parameters. This was varied from zero, the training condition when all parameters were known, to two when the two means were unknown, to four when the two means and the two variances were unknown, to five when all parameters were unknown.

2. d', the separation between the means of the standardized stimulus distributions. There were five levels of d': .5, 1.5, 2.0, 3.0, and 5.0.

3. Prior probability, P_o, of drawing a score from the distribution of successful applicants. There

were four levels of prior probability, .5, .333, .25 and .125.

4. Payoff matrices. Payoff matrices were manipulated to require an increasingly stronger minimum posterior probability, P_1, of being correct before it was optimal to classify the observation and risk an error. There were four levels of P_1: .67, .75, .83 and .90.

5. Variance ratio. There were three levels of the ratio between the variances of the two distributions. These were: 1:1, 1:4, 1:64.

Since a complete factorial block design would imply 960 separate conditions only conditions representing main effects were employed; conditions representing interactions were not used. The prototypical condition employed 2 unknowns (the two means), d' = 2.0, P_o = .5, P_1 = .75 and variance ratios of 1:1. In all other experimental conditions irrelevant variables were held constant at those values, while each relevant variable was manipulated separately. The three levels of variance ratio and two of the levels of d', .5 and 3.0, were manipulated in a separate complete factorial design making six additional conditions. There were 21 conditions in all.

Subjects

Three University of Michigan undergraduate students (two males and one female) from the Human Performance Center paid subject pool participated in the first experiment. Although one subject was a "professional" subject and had been in many psychological experiments and although another subject had had a slight familiarity with normal distributions, all were totally naive with respect to this task and to the type of problem they were called upon to solve.

Apparatus

An IBM 1800 computer with three television screen terminals, each in a semi-isolated booth, displayed stimuli and questions to the subjects, and collected responses from them. "Pass," "Fail," "Uncertain" and later "Very Good" and "Poor" responses were communicated via pushbuttons. Additional responses were typed on terminal keyboards situated between the tubes and the buttons. A programming system was used which allowed the three subjects to be run simultaneously and independently.

Stimuli

Stimuli were positive integers from zero to 999 representing scores on a battery of Civil Service Tests.

For each of the twenty-one separate conditions two distributions of scores were constructed such that each score was an approximation to a normally distributed random number with a specified mean and standard deviation. An approximation to normally distributed random numbers were derived from a sequence of uniform random numbers by the formula:

$$Y = \frac{\left[\sum_{i=1}^{k} x_i\right] - \frac{k}{2}}{\sqrt{k/12}}$$

where x_i is a uniformly distributed random number,

$$0 < x_i < 1,$$

and k is the number of values x_i to be used. Y approaches a true normal distribution as k approaches infinity. For programming purposes k was chosen as 12 to reduce execution time: the equation thus becomes:

$$Y = \left[\sum_{i=1}^{12} x_i\right] - 6.0.$$

The adjustment for the required mean and standard deviation is then

$$Y' = Y * S + M$$

where Y' is the required normally distributed random number, S is the standard deviation selected and M is the selected mean.

Y' was truncated at its integer value for display to the subjects.

S and M were chosen so that no score would fall below zero or above 999.

Procedure

Subjects were brought into the experimental room prior to the initial session and informed that on each trial they would be shown a number representing the score of a particular job applicant on a Civil Service Test. They were informed that it would be their task to grade the score "pass," if they felt the applicant would eventually turn out to be successful on the job, or to grade it "fail" if they felt the applicant would eventually turn out to be a failure. They were also encouraged to grade the score "uncertain" if they really were uncertain about the eventual success of the applicant. It was explained that the distributions of test scores were overlapping which meant that it was possible for someone to get a high score on the test and turn out to be a failure on the job, while another person might get a low score on the test but turn out

to be a success on the job. They were told that on the average, however, a high test score would indicate eventual job success.

After grading the score they were asked to set their pass/fail decision boundaries, that is, they typed in the lowest score they would have given a pass to and the highest score they would have failed.

The operation of the payoff matrices was explained and subjects were informed that they would not be paid according to how well they classified the individual scores as they went along, but on how well their final decision boundaries did at classifying an additional 50 scores drawn at random from the same two distributions. They were told that the losses and gains resulting from the classification of these 50 scores would be added to or subtracted from their $2.00 hourly wage. Participants were also informed of the random device which could stop them during the session and cause their current pass/fail boundaries to be used to classify them. They were told that

therefore they should be as careful as they could each time they set their boundaries, as they may be stopped at that point. There is every evidence that observers took this seriously--their decision cutoffs represented their best guess at the moment. Subjects were not told exactly how many sessions they would be given, except that it was between twenty and thirty, so that an extra session could be given if a subject were stopped. In this way results were always based on self-terminated sessions.

Besides the written instructions, research participants were also shown various examples of overlapping normal distributions and were given an intuitive understanding of the roles the separation between the distributions and the spread within them played in making their classification task easier or more difficult. Subjects were questioned until the experimenter felt they understood these concepts. Subjects were able to verbalize the fact that unless the distributions were very far apart errorless responding was impossible.

Each subject was run in one condition per day, five days a week, for up to six weeks. The order of the conditions was randomized for each subject.

At the start of each experimental session, subjects received a form stating the job and test title of

the day, a verbal description of the parameters to be known at the start, and the payoff matrix for that session. They were also given a transparency of each distribution. The use of transparencies enabled the subjects to overlap the distributions to the degree they felt was appropriate.

Each subject then sat at his own booth and each engaged in the following dialogue with the machine:

Computer: "Welcome and thank you for coming. Are you ready to begin?"

Subjects: "Yes."

C: "Here is the first score: _____ How do you grade this score?"

S: "Uncertain" (or "pass" or "fail")

C: "What do you think is the lowest score you would have given a pass to?"

S: _____ (Subject types in his passing boundary on his keyboard)

C: "Can you imagine a score above ____ (the previous typed passing boundary) which you would fail?"

S: "No" (A yes response here would indicate to the machine that the subject felt he was dealing with unequal variances. Additional questions were then asked to elicit boundaries for the two uncertain ranges.)

C: "What do you think is the highest score you would have failed?"

S: ____ (Subject types in his failing boundary).

C: "You are uncertain about scores between ____ (his passing boundary) and _____ (his failing boundary)." This is the only feedback the subject receives.

C: "Here is your next score _____."

This cycle continues until the subject types identical boundaries for five trials in a row. At which point the following dialogue occurs.

C: "Do you think you might see some scores that will make you change your pass/fail boundaries?"

S: "No"(If yes, the next score is displayed and the cycle repeats.)

C. "Remember you are being paid according to your permanent boundaries. If you are willing to let these be your permanent boundaries we can stop, otherwise we can go on. Shall we go on?"

S: "No"(A yes response initiates the display of the next score.)

C: "OK. We can stop now. Thank you very much for all your effort. You will now be scored on an additional 50 scores classified according to your final boundaries."

C: "The number of scores graded correctly according to your final boundaries is ____.

The number graded uncertain is ____.

You have won ____ cents. "

The subject was then paid his $2.00 fee for his time plus (or minus) his winnings.

Results

The first prediction of the model was that observers' decision criteria would be sensitive to differences in payoff matrices. The usefulness of this type of payoff scheme rests on the validity of this assumption. The model states that the steeper the penalty for error relative to the reward for a correct response the larger should be the size of the uncertain area. This prediction was supported. As can be seen in Figure 1, subjects did increase the size of their uncertain area as a function of increases in penalty for error. MK's large deviance from the predicted value occurred when he suddenly verbalized the realization that a penalty of $1.00 for each error meant a possible loss of $50.00. He set his uncertain area so large that no scores could possibly fall outside of it.

It was also necessary to determine if improvement in performance occurred before performing additional

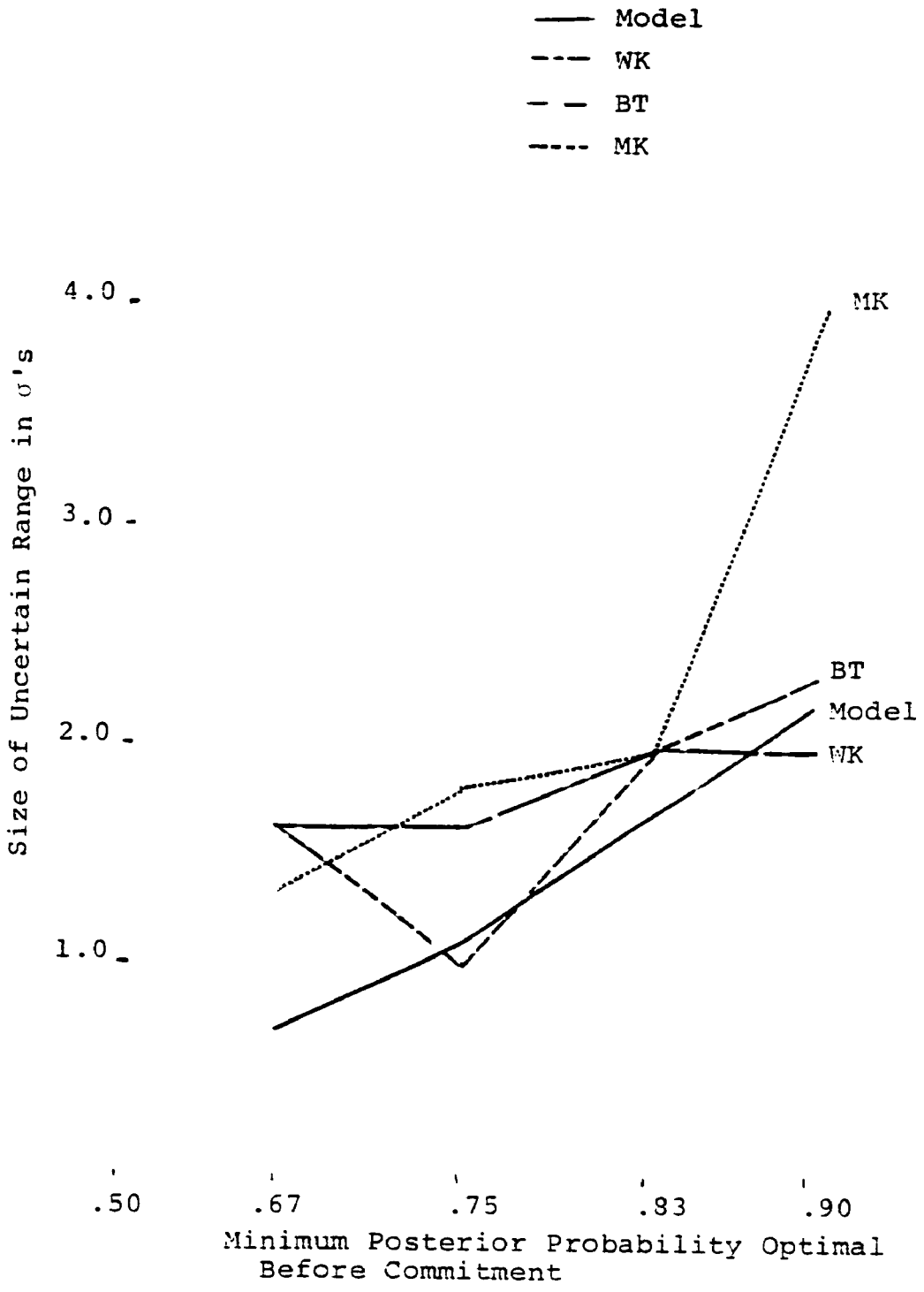


FIG. 1: Size of Uncertain Range as a Function of Payoffs

analyses. If none occurred the experimental situation would have to have been abandoned as being too difficult for untrained observers. Improvement was measured by the increase in expected value of final decision boundaries over initial ones. Increase in expected value was chosen rather than increase in percent correct since observers were paid only according to their decision criteria and not according to how well they classified the socres they received. Expected value was computed by measuring area under the normal density function as delineated by the decision boundaries. The model is assumed to classify all socres "Uncertain" after the first observation and to have learned the true parameter values at asymptote. The expected values attributed to the model in the following discussion is computed based on these simplifying assumptions.

As can be seen from Table 1 there is a pronounced increase in expected value for all three subjects over all equal variance conditions. Indeed out of the 48 non-training equal variance conditions (16 condictions x 3 subjects) only nine showed a decrease in expected value. Of these nine conditions, three were in conditions of very small d' (.5) while four were found in conditions of unequal prior odds. Thus while the final expected value of the subjects did not approach that of an optimal model assumed to have learned the true means, taking all conditions together there is a marked improvement in

TABLE 1

AVERAGE INCREASE IN EXPECTED VALUE OF CLASSIFICATION BOUNDARIES FOR EACH SUBJECT AND THE MODEL OVER ALL EQUAL VARIANCE CONDITIONS

Number of Unknown Parameters	Subject	Expected Value of Boundaries After One Observation	Expected Value of Boundaries at Asymptote	Net Increase
Two	BT	- .60	- .03	+ .51
	WK	- 1.03	- .19	+ .84
	MK	- .53	+ .09	+ .62
	Model	- .50	+ .23	+ .73
Four and Five	BT	- 1.007	+ .275	+1.352
	WK	- 1.024	+ .63	+1.65
	MK	- .0910	+ .2972	+ .3882
	Model	- .50	+ .374	+ .874

expected value occurring even without feedback.

Given the existence of improvement in expected value, what was being learned? If the subjects were learning the parameters of the underlying distributions, as predicted, then they must have been learning the differences between the means as well, and their decisions should reflect this knowledge. The model predicts that uncertain ranges be large with small d' and small with large d'.

Figure 2 shows the size of subjects' uncertain ranges as a function of d'. Only MK follows the behavior of the asymptotic model over all five experimental conditions, although WK does follow it in the three conditions in the middle portion of the curve. It could be surmised from this data that both BT and WK were unaware of differences in d' across experimental conditions, but Table 2 suggests a different interpretation.

Table 2 shows the number of observations taken prior to asymptote for each subject for each level of d'. Both BT and WK obtain large amount of data with small d' and small amounts with d' large, and may therefore indicate that these subjects were aware of differences in d' but were not able to or did not know how to make their decision boundaries appropriately reflect this knowledge.

The model also predicts that differences in prior probabilities would be detected by the observer if

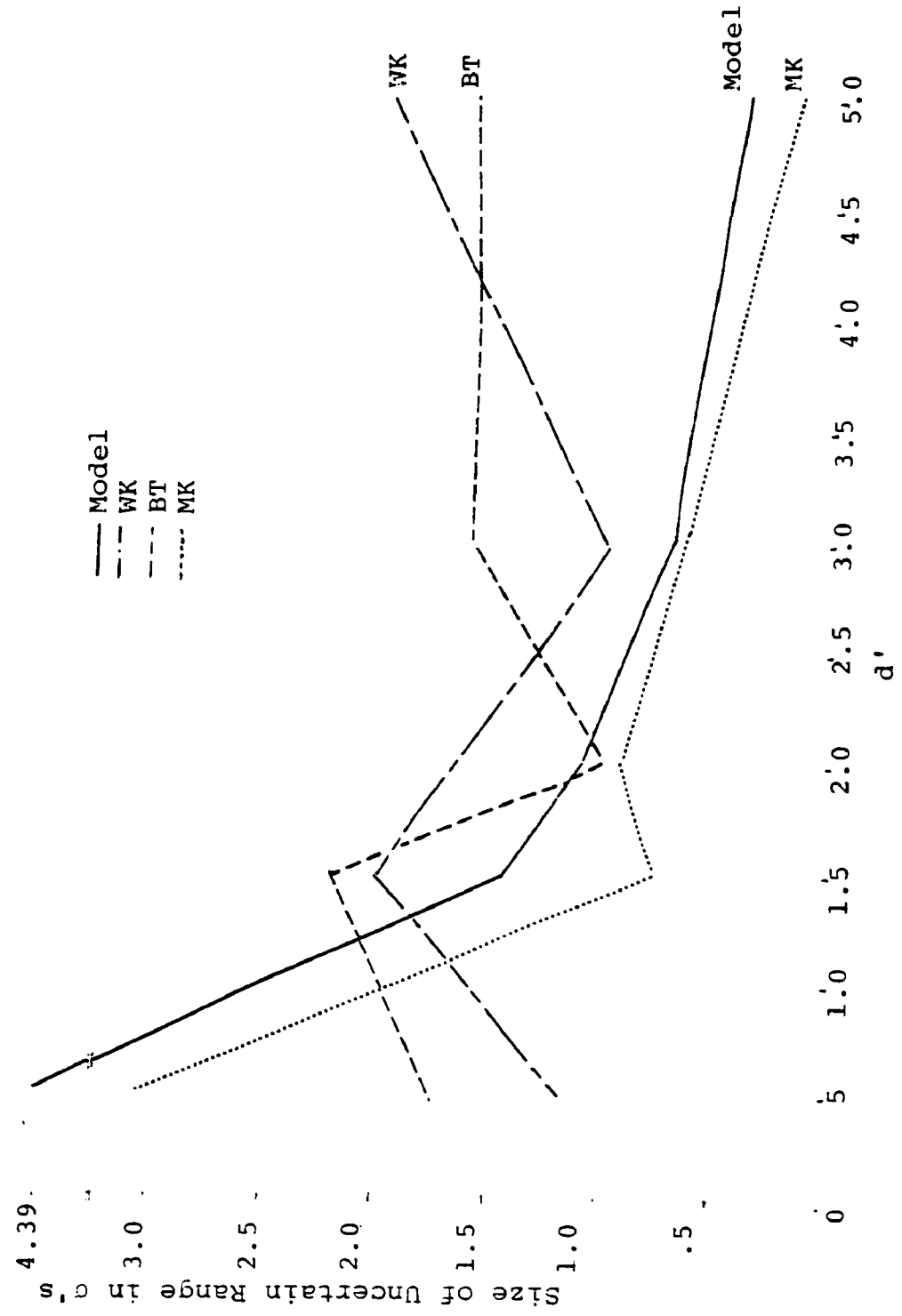


FIG. 2: Size of Uncertain Range as a Function of d'

TABLE 2

NUMBER OF OBSERVATIONS TAKEN BY EACH SUBJECT AS A FUNCTION OF d'

	Subjects		
d'	BT	WK	MK
.5	24	33	28
1.5	20	24	48
2.0	24	27	29
3.0	9	24	42
5.0	9	19	29

unknown, and that the observer would react to these differences appropriately. For all else held constant and for equal variances, the midpoint of the subjects' uncertain range should shift toward the mean of the least likely distribution and should be more pronounced as prior probabilities become more extreme. Before ascertaining if differences in prior probabilities could be detected it was necessary to determine if they would be reacted to appropriately even if known. Accordingly prior probabilities were varied but made known to the subjects. Only the two means were unknown in these conditions.

As seen in Figure 3 only one subject imitated the model and shifted the midpoint of her uncertain range

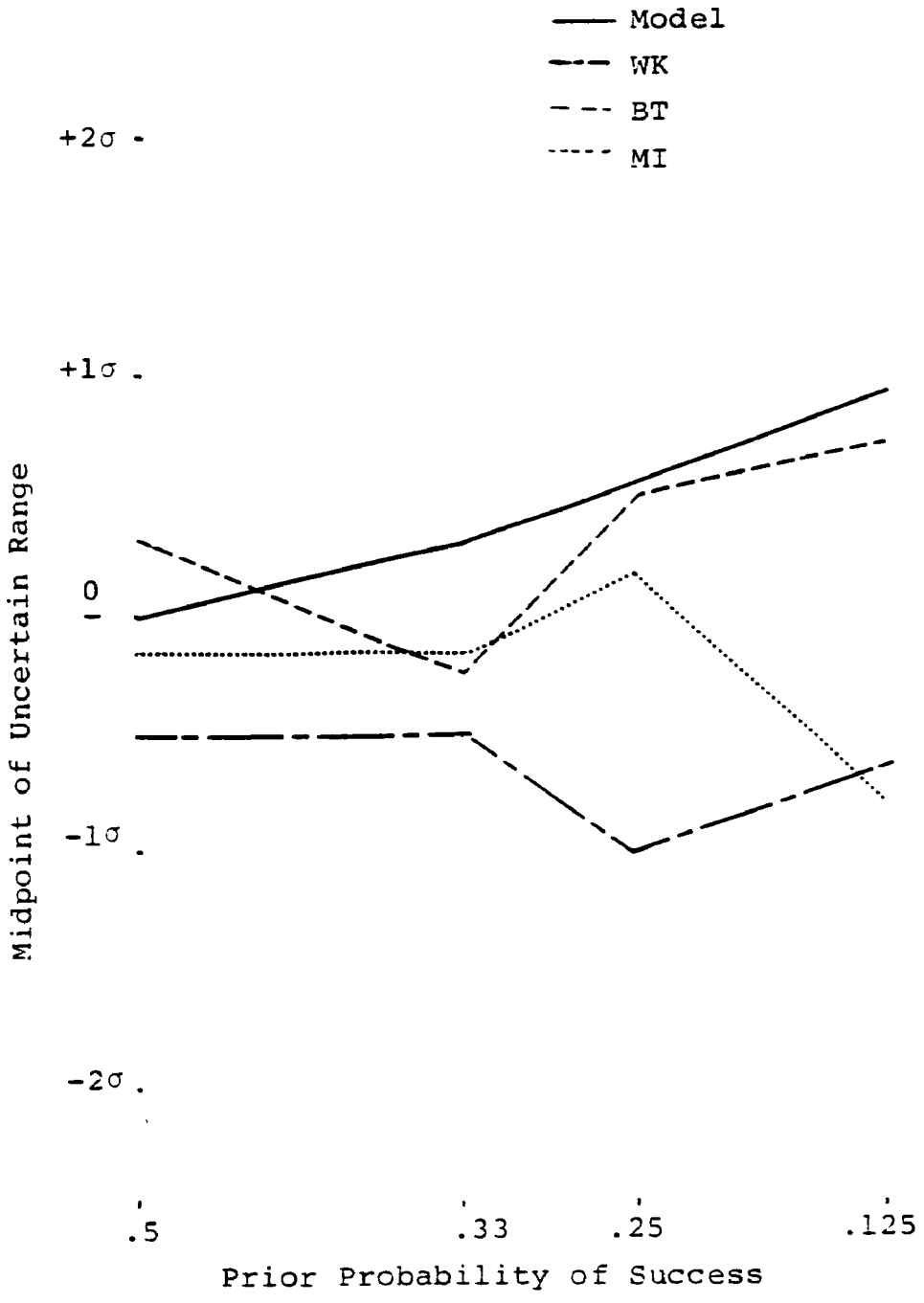


FIG. 3: Changes in Midpoint of Uncertain Range as a Function of Prior Probabilities

toward the mean of the least likely distribution. However, the model dictates that the size of this region be kept constant, while this subject accompanied each shift in midpoint by a corresponding increase in size. Only the upper boundary point was shifted to accomodate the unlikely population; the lower boundary remained constant. MK also accompanied his one and only shift (at prior odds of 3:1) by an increase in size of his uncertain region. Only WK did not alter his behavior at all with respect to differences in prior odds.

A fourth prediction of the model is that decisions would be sensitive to differences in variance ratios. The existence of unequal variances should force an inversion of classification with respect to the decision variable. That is, very high scores and low scores would be called fail, while only moderately high scores would be called pass. Two levels of variance ratios, 1:4 and 1:64, and two levels of d' were compared in a complete factorial design.

Table 3 shows that for variance ratios of 1:4 there were no inversions. However, as can be seen in the table, for an extreme variance ratio of 1:64 and small d', two of the three subjects did invert their

TABLE 3

PRESENCE OR ABSENCE OF CLASSIFICATION INVERSIONS
UNDER COMBINED VARIANCE d' MANIPULATIONS

	Variance Condition					
	$\sigma_x = \sigma_y$		$\sigma_x = 1/2\ \sigma_y$		$\sigma_x = 1/8\ \sigma_y$	
Subject	d' = .5	d' = 3.0	d' = .5	d' = 3.0	d' = .5	d' = 3.0
WK	no	no	no	no	no	no
BT	no	no	no	no	yes	no
MK	no	no	no	no	yes	no
Model	no	no	yes	yes	yes	yes

classifications on the decision axis. This was indicated by the setting of two separate uncertain areas, one on either side of the mean of the narrow distribution. Thus observers do alter their decision criteria with respect to differences in variance ratios when these differences are known to them. It is important to realize that inversions could not occur unless observers were following a likelihood ratio decision rule. A decision rule based solely on distance to the mean predicts no inversions.

Discussion of Results

The large improvement in expected value of final decision criteria over initial ones indicates that learning occurred in this task even though the observers did not receive feedback. The questions of interest then become: 1) what exactly is being learned, and 2) is the model (GIL) a valid description of the learning process.

Success in this unique two-boundary paradigm requires that the parameters of the two distributions be learned and that a likelihood ratio rule be employed. The fact that observers alter their decision rule appropriately with changes in payoff matrices when all else is held constant indicates that for a given set of

distribution parameters observers are employing a likelihood ratio decision rule (Green and Swets, 1966).

Variation in distribution parameters leads to inadequate variation in decision criteria, however. This is especially salient in conditions in which prior odds were varied since the value of the prior odds was always known to the observer. It may be that human observers are incapable of correctly utilizing information regarding asymmetric prior probabilities. Tversky & Kahneman (1974), for example, found prior probabilities were attended to only when no additional information was given. Healy and Kubovy (personal communication) also found that their research participants did not change their decision cutoffs with changes in prior probability. It cannot be determined here if observers misinterpreted the underlying distribution means or if they merely failed to respond to variation in prior odds or both.

Separation of the learning process from the decision component is even more difficult in situations of varying d'. Optimal responding would have indicated that both the parameters of the subdistributions were learned and that appropriate decision rules were employed. Again, a failure to change decision criteria

with changes in d' may be due to an inability to select appropriate decision boundaries or it may be due to an inability of observers to adequately estimate the parameters of the separate distributions from a combined sample, or both. Lewin (1975) has shown that observers can perform the opposite task with a fair degree of accuracy; that is, observers can estimate the combined mean of two samples when each is drawn from a separate population. This accuracy in estimation was exhibited even when the samples differed greatly in size and in variance. Whether observers can perform the complementary task of estimating means of two subpopulations from a combined sample has not been determined. One of the three observers in this experiment obviously did estimate the distribution parameters; and the fact that the other two observers varied the number of observations they obtained inversely with d' indicates that these two observers might also have learned the distribution parameters but responded inappropriately. A second indication that learning may have occurred is that the two observers who responded appropriately to differences in variance ratios showed they were able to distinguish large d' from small d' in those situations.

These results do not permit the unequivocal conclusion, however, the observers in a signal detection or similar task are necessarily decomposing the underlying distributions. It may be that observers attempt to learn the underlying parameters when they sample information but because they don't sample enough data, or perhaps because the task is simply too difficult, their final estimates are inaccurate, and faulty decision criteria result. On the other hand, it may be that observers sample information only with the goal of learning appropriate decision criteria without concerning themselves with the value of the underlying parameters. Faulty decision criteria could be caused by much the same reasons in this case as well. The problem is one of separating the reasons for sampling information from observers' ability to respond appropriately.

CHAPTER III

EXPERIMENT 2

Statement of Purpose

It seemed that it may be possible to distinguish between the two purposes for sampling, i.e., learning a decision criterion or learning distribution parameters, without requiring that optimal decision criteria be exhibited. The purpose of the second experiment then was to determine the goal of the observer in sampling information. If an observer's purpose in sampling is to learn the values of the underlying distribution parameters, and if he were to be given those values in advance, then he would no longer need to sample data--he would be able to set his decision cutoffs in advance. Preasymptotic variability in decision criteria would then be significantly reduced. On the other hand, if his goal in sampling data is to learn an optimal decision rule, with the values of the underlying parameters being irrelevant, as previous authors have suggested, knowledge of these parameters would not affect the number of observations sampled nor the degree of variability in his preasymptotic decision criteria.

It was decided then to repeat the entire first experiment adding eight conditions in which all five parameters would be known to the observer in advance. These eight conditions would include five of varying d' and three of varying prior odds. Since it was hypothesized that observers sample data primarily in order to learn the parameters of the underlying distributions, it was predicted that, in those conditions in which the parameters were known, significantly fewer observations would be obtained and the amount of variability in decision criteria prior to asymptote would be significantly reduced. It was also predicted that if the values of the parameters were indeed relevant to the observers, expected value would be higher in conditions when these values were known than when they were not. Failure to find these results would indicate that the dynamic cutoff models reviewed above could not be rejected and that observers sample in order to learn an optimal decision cutoff.

The addition of these eight conditions may also help in distinguishing the learning process from the responding process. It may be that observers are capable of estimating underlying distribution parameters, but fail to make their decision criteria adequately reflect these estimates. If observers still

fail to vary their decision cutoffs with d' when d' is known then this hypothesis cannot be rejected. However, if decision criteria were adjusted appropriately when d' is known and not adjusted when it must be learned, it can then be safely assumed that learning did not occur.

Method

Experimental Design

The experimental design was as in Experiment 1 except that eight new conditions were added. These had all five parameters known to the observers. In the first five conditions $P_o = .5$, $P_1 = .75$, and $\sigma_x^2 = \sigma_y^2$ were held constant and only d' was varied. The five levels of d' were the same as in Experiment 1: .5, 1.5, 2.0, 3.0, and 5.0. In the last four conditions d' = 2.0, $P_1 = .75$ and $\sigma_x^2 = \sigma_y^2$ were all held constant. Only prior odds were varied. The four levels of prior odds were also the same as in Experiment 1. They were 1:1, 2:1, 3:1 and 7:1 against sampling a score from the successful distribution of applicants.

Subjects

Subjects were three male graduate students from the Human Performance Center paid subject pool at The University of Michigan. One was specializing in English,

one in Law and one in Public Administration. None had any background in statistics and none had familiarity with the type of problem solving expected of them.

Apparatus, Stimuli, Procedure, and Instructions to Subjects

These were all identical to those in Experiment 1. Order of conditions was re-randomized for presentation to each subject.

Results

Before examining the predictions specific to this second experiment it was first necessary to determine if the results of the first were replicated. Failure to replicate would place the validity of the experimental paradigm in doubt. Of primary concern was whether these observers were also sensitive to the role of the payoff matrix in determining decision boundaries.

As can be seen from Figure 4 and from a comparison of it with Figure 1 the results of the first experiment were completely replicated here--even to the existence of the single overcautious research participant. Observers are indeed sensitive to differences in payoff matrices and vary their decision criteria in the same direction that an optimal observer would. This again demonstrates the viability of the payoff matrix.

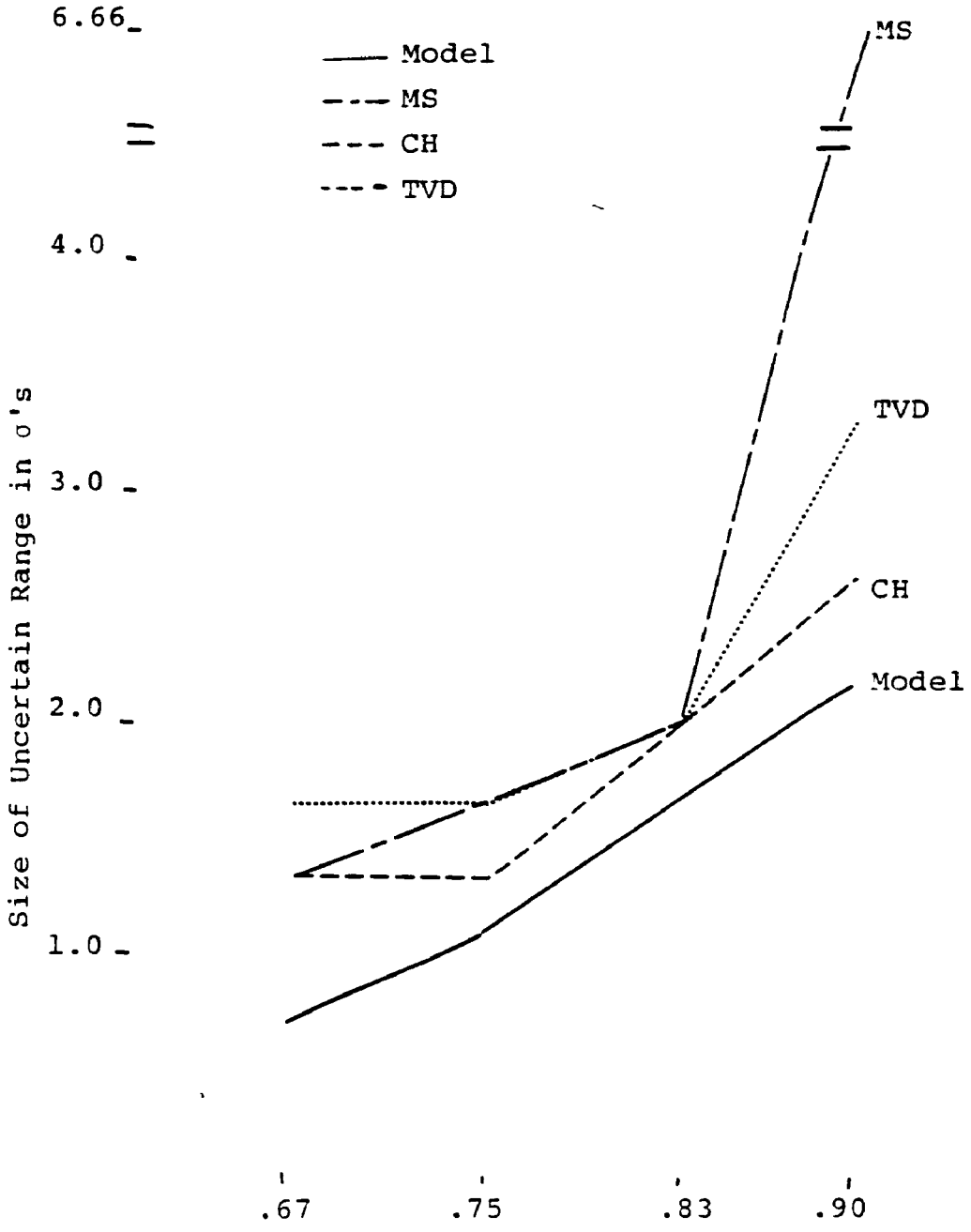


FIG. 4: Size of Uncertain Range as a Function of Payoffs

The second question was whether observers in this experiment also improve in performance from initial boundary decisions to final ones, without feedback.

Table 4 shows the large increases in expected value for each subject over all equal variance conditions when the distribution parameters are not known in advance. These results replicate those of the first experiment; significant amounts of learning also occurred in experiment 2.

When the distribution parameters are known in advance, however, no such increases in expected value can be found. It can be seen that except for one observer, expected value does not change appreciably from initial decision boundaries to final ones. Learning apparently did not occur in two observers in these conditions. Moreover, it can be seen that expected value resulting from initial boundary decisions in conditions when the parameters were known are greater on the average than the expected values resulting from final boundary decisions when the means were unknown in advance. Indeed, 16 of the 24 pairs (3 x 8) showed the same or larger expected value at the start of the observation process when parameters were known than occurred at asymptote when they were not. Seven of the eight pairs remaining were among the nine conditions of small d' (.5 and 1.5) while one had extreme prior odds of 7:1.

TABLE 4

AVERAGE INCREASE IN EXPECTED VALUE OF CLASSIFICATION BOUNDARIES FOR EACH SUBJECT AND THE MODEL OVER ALL EQUAL VARIANCE CONDITIONS

Subject	No. Unknown Parameters	EV After 1st Obs.	EV at Asymptote	Net Change
MS	0	+ .0417	+ .2588	+ .217
CH		+ .0375*	+ .1239	+ .086
TVD		+ .1606	+ .2080	+ .0474
Mean		+ .0799	+ .1696	+ .117
Model		+ .2300	+ .2300	0
MS	2	- .306	- .03	+ .276
CH		- .9747	+ .0285	+1.0032
TVD		- .6796	+ .0549	+ .7345
Mean		- .6534	+ .0178	+ .6712
Model		- .50	+ .2300	+ .7300
MS	4	- .2121	+ .2195	+ .4316
CH		- .3400	+ .0836	+ .4236
TVD		- .5779	+ .3028	+ .8807
Mean		- .3767	+ .202	+ .5786
Model		- .50	+ .37	+ .87

*Does not include condition number 29, in which prior odds were 7:1

If increases in expected value are taken as a measure of learning then, these data suggest that learning does not occur when the distribution parameters are known in advance and, moreover, that observers are close to asymptote at the beginning of the observation process when all the parameters are known.

The notion that the observer who knows the distribution parameters in advance is already at asymptote is further supported by a comparison of the degree of variability in decision boundaries as a function of number of unknown parameters. The standard deviation around the mean upper decision boundary as well as that around the mean lower decision boundary were taken as measures of criterion variability. These standard deviations were then averaged over all equal variance conditions. It can be seen in Table 5 that the amount of preasymptotic variability in performance is significantly ($p \leq .01$) reduced when the distribution parameters are known. These data suggest that observers are able to set their decision criteria in advance when the parameter values are known. Indeed two of the observers, TVD and CH, did set them in advance in half of the sixteen conditions in which the parameters were known and in two more conditions they varied only their upper decision boundaries, keeping

TABLE 5

MEAN NUMBER OF OBSERVATIONS AND MEAN STANDARD DEVIATION AMONG UPPER AND LOWER BOUNDARIES FOR EACH SUBJECT IN EACH SET OF EXPERIMENTAL CONDITIONS

Conditions in which d' varied	Means Unknown			All Parameters Known		
	Av # obs	Av σu	Av σl	Av # obs	Av σu	Av σl
MS	22.8	58.29	50.76	11.4	13.06	7.04
CH	18.4	31.33	48.07	12.4	2.08	15.35
TVD	11.2	40.06	76.21	6.8	3.40	3.53
MEAN	17.467	43.226	58.347	10.2	6.18	8.64
P_o varied						
MS	24.75	25.48	34.73	15.5	2.14	4.09
CH	17.75	40.33	27.47	15.25	8.83	18.65
TVD	14.6	28.40	55.70	6.25	4.13	2.89
MEAN	19.033	31.403	39.30	12.333	5.033	8.54
combined means						
MS	22.75	48.1	47.389	12.87	9.234	7.48
CH	17.625	38.138	47.337	13.62	4.417	15.142
TVD	14.125	38.29	67.702	6.75	4.202	3.65
MEAN	18.167	41.509*	54.143*	11.08	5.95*	8.76*

*These differences are significant at $p \leq .01$.

their lower ones constant. Although not completely eliminating it, MS also greatly decreased the amount of variability in his decision boundaries as well. Thus when parameter values are known variability in preasymptotic decision criteria is significantly reduced.

The next question is, does knowledge of parameters enable decision boundaries to fluctuate appropriately with changes in d' and prior odds?

As can be seen from Figures 5 and 6 only MS can be said to have altered the size of his uncertain region appropriately with changes in d' as a result of his knowing the distribution parameters. Knowing the parameters did not help CH or TVD at all in determining the size of their uncertain ranges.

The model states that for equal prior odds and equal variances when d' is varied only the size of the observer's uncertain range should vary with it. The midpoint of the uncertain range should remain fixed exactly at the point midway between the two distributions. Knowing the means of the distributions may not indicate to the lay observer how wide his uncertain area should be, but it evidently does tell him exactly where to place the midpoint. As can be seen in Table 6,

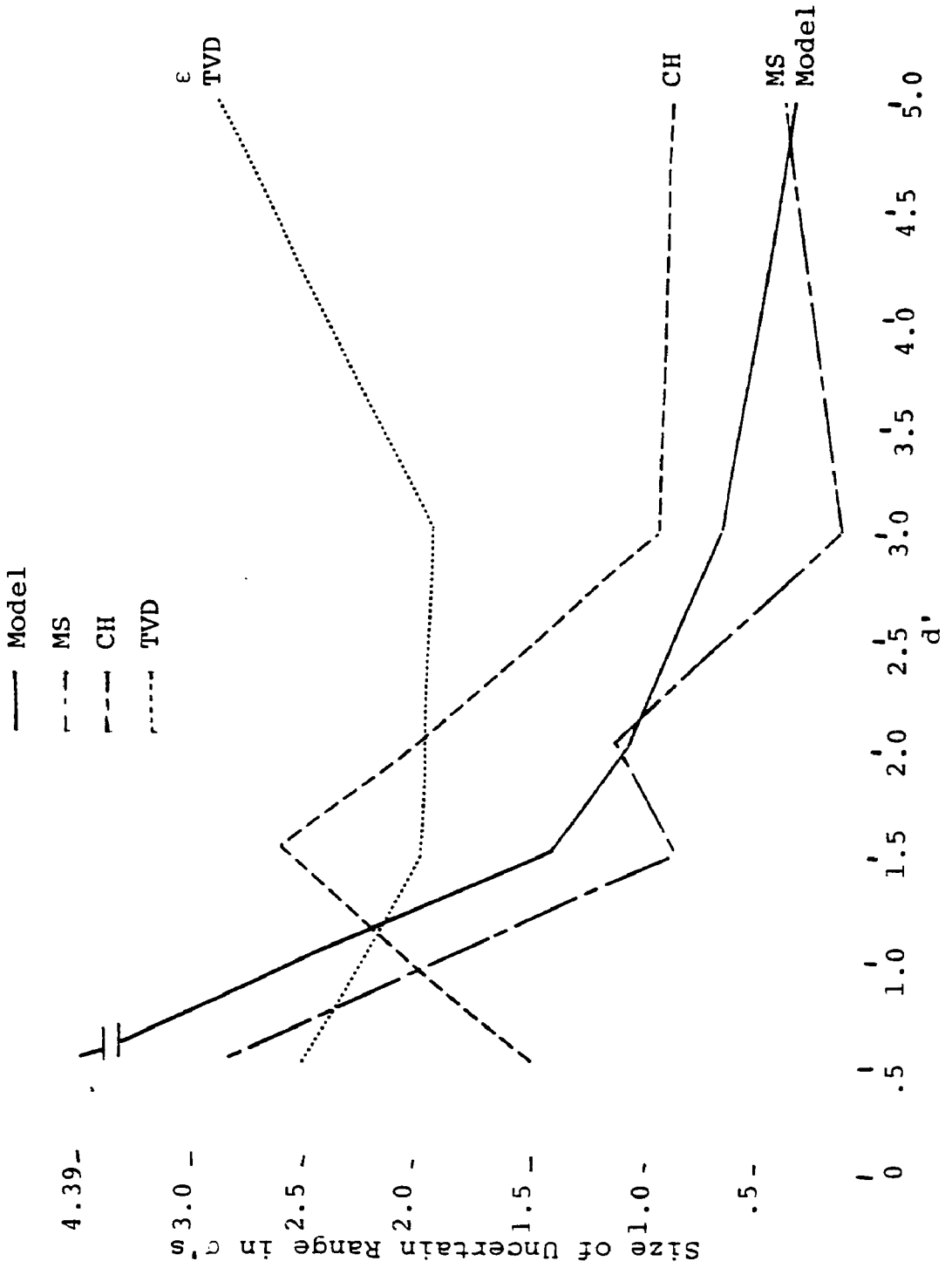


FIG. 5: Size of Uncertain Range as a Function of d', Means Known

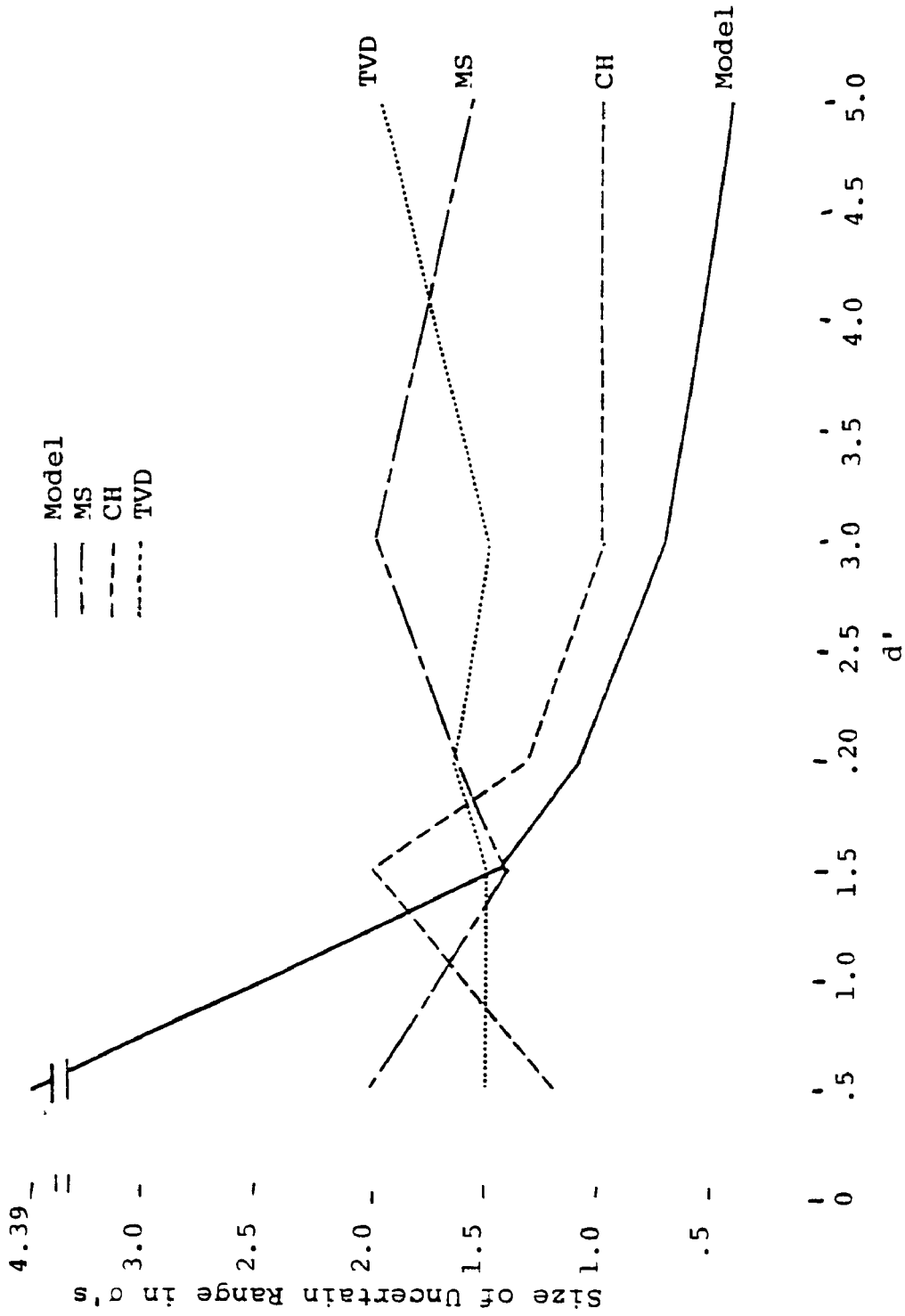


FIG. 6: Size of Uncertain Range as a Function of d', Means Unknown

TABLE 6

NUMBER OF STANDARD DEVIATIONS BETWEEN MIDPOINT OF SUBJECT'S UNCERTAIN REGION AND ACTUAL MIDPOINT BETWEEN THE DISTRIBUTIONS, FOR EQUAL PRIOR ODDS AND VARIANCES

	Means Unknown			Means Known		
	Subject			Subject		
d'	MS	CH	TVD	MS	CH	TVD
.50	-.333	-.35	-1.0	0	0	0
1.50	-.35	-.25	+ .50	0	-.65	-.05
2.00	+.167	+.33	+ .333	+.08	0	0
3.00	-.333	0	+ .75	0	0	0
5.00	+.30	-.50	+ .5	+.85	0	0
Average Absolute Deviation	.297	.286	.617	.186	.13	.01

of the 15 (5 x 3) conditions in which d' was varied, equal prior odds obtained, and the means were known, eleven had the midpoint of the observers' uncertain region placed exactly midway between the distribution means, and in another two conditions the deviation from this was by only trivial amounts. Thus thirteen of these fifteen uncertain regions were optimally centered when distribution means were known; while only one out of the fifteen uncertain regions was appropriately centered when these means were unknown. Thus it can be seen that in conditions of equal prior odds the variability that remains in preasymptotic decision criteria in this experimental paradigm, when distribution parameters are known, is entirely due to ambiguity regarding the appropriate size of the uncertain region. There is no ambiguity at all regarding its correct location.

The opposite result is predicted when only prior odds are varied and all else remains constant. Here the model predicts that the size of the uncertain region will remain constant while the midpoint moves toward the mean of the least likely distribution. A comparison of Figures 7 and 8 with Figure 3 will show that when parameter values are unknown the results of the second experiment completely replicate those of the first. None of the observers shift the midpoint of their uncertain region appropriately, and in fact

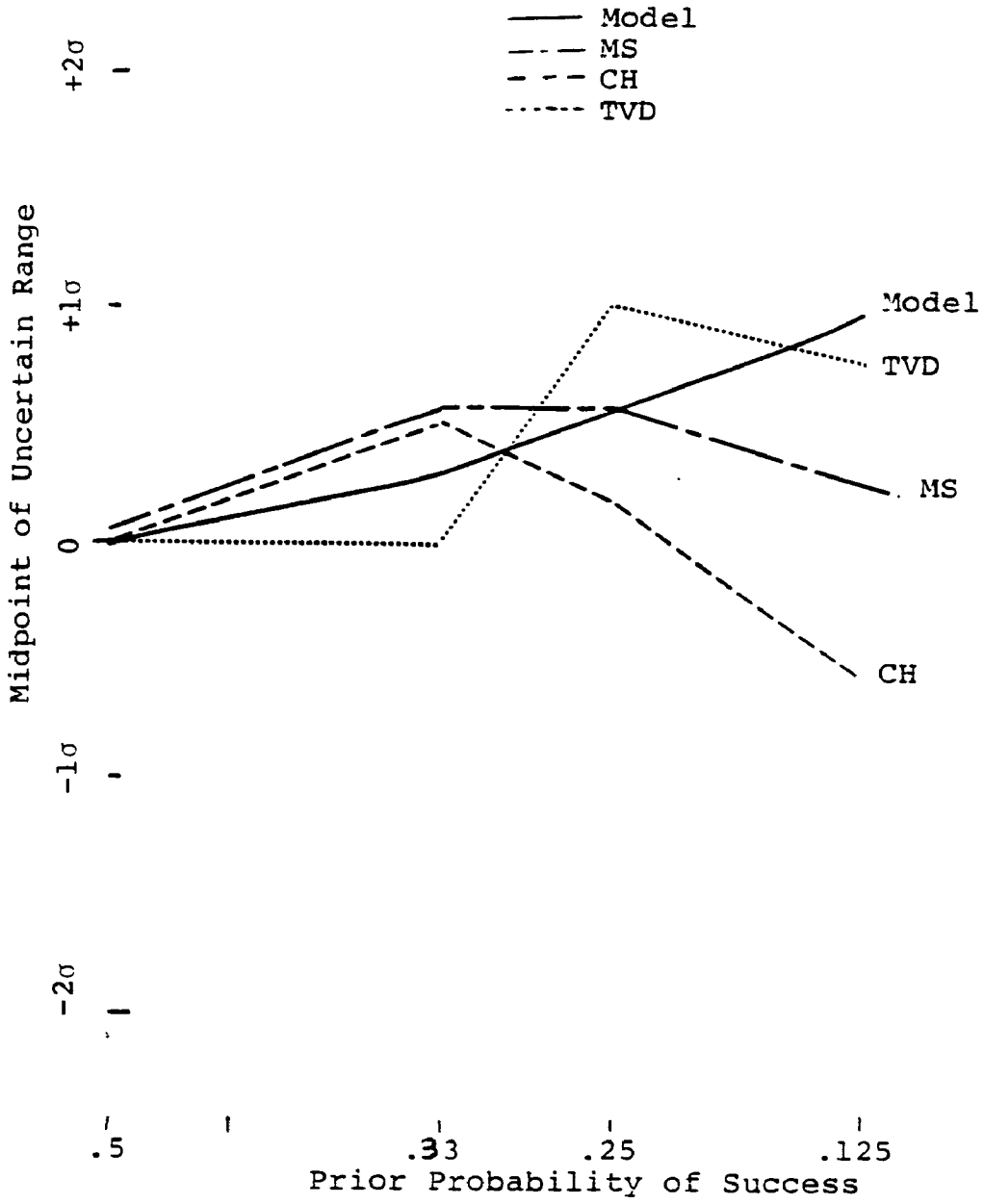


FIG. 7: Midpoint of Uncertain Range as a Function of Prior Probabilities, Means Known

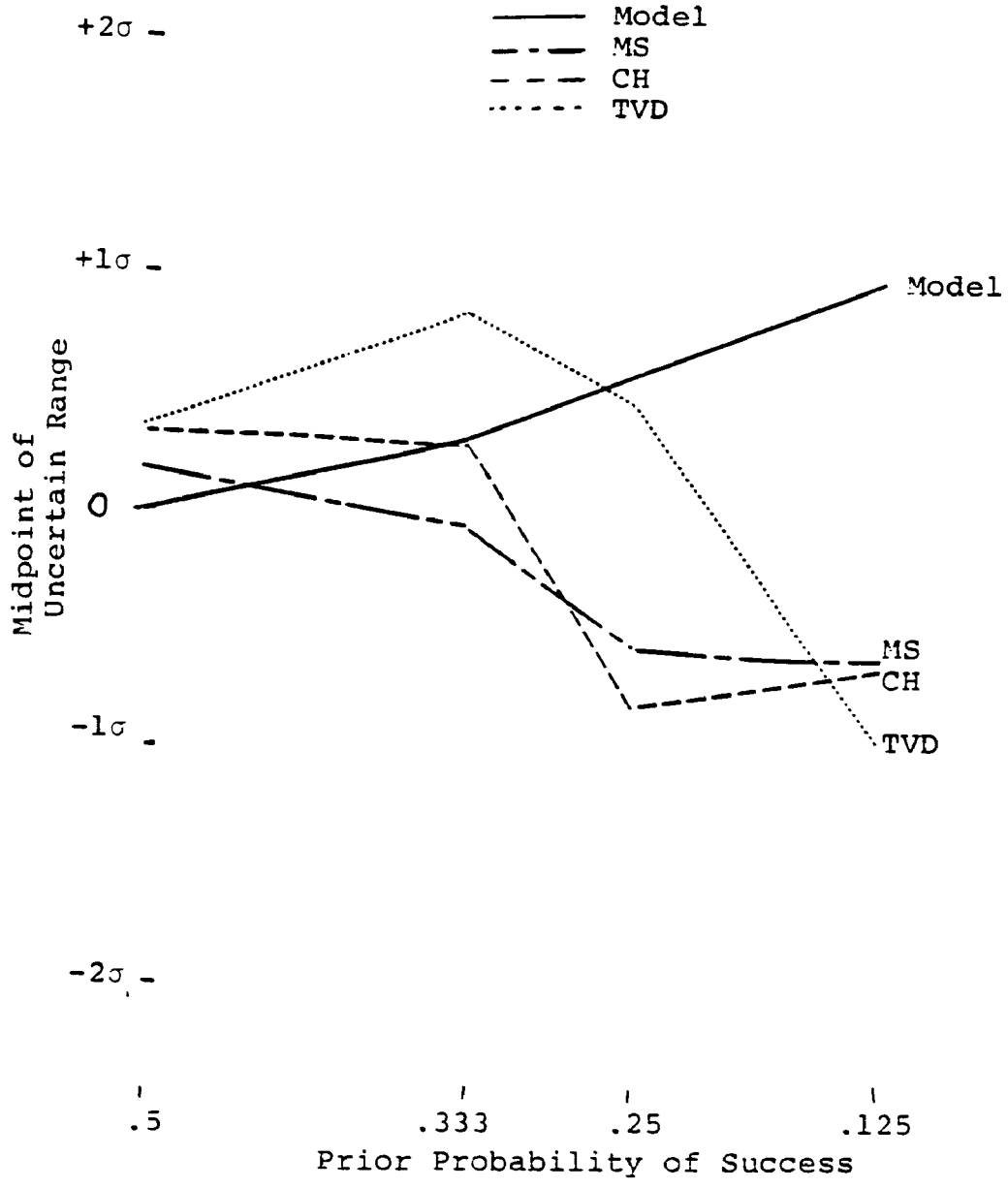


FIG. 8: Midpoint of Uncertain Range as a Function of Prior Probabilities, Means Unknown

shifted them inappropriately toward the mean of the more likely distribution when prior probabilities were extreme. With all but one of the observers, however, knowing parameter values prevented this inappropriate shift with extreme priors. In fact, a major portion of the improvement in expected value resulting when parameters are known occurred when prior odds was varied. This was due to the drastic effect a shift toward the more likely distribution had on expected value when the means of the distributions were unknown.

In addition to the above, the results of the first experiment were also replicated when variance ratios were unequal.

Table 7 shows that two of the three observers in this experiment inverted their classifications on the decision axis when d' was small and variance ratios extreme as in Experiment 1. These two observers also inverted their classifications when d' was large as was appropriate, and one observer inverted at large d' when variance ratios were only 1:4. Thus again two observers followed a likelihood ratio decision rule, in the direction of an optimal observe, while one followed the suboptimal distance rule for category learning suggested by Reed (1972, 1973).

TABLE 7

PRESENCE OR ABSENCE OF CLASSIFICATION INVERSIONS
WITH UNEQUAL VARIANCES

	Unequal Variance Conditions					
	$\sigma_x = \sigma_y$		$\sigma_x = 1/2\ \sigma_y$		$\sigma_x = 1/8\ \sigma_y$	
Subject	d' = .5	d' = 3.0	d' = .5	d' = 3.0	d' = .5	d' = 3.0
#7	no	no	no	no	yes	yes
#8	no	no	no	yes	yes	yes
#9	no	no	no	no	no	no
MODEL	no	no	yes	yes	yes	yes

Discussion of Results

Observers in a numerical probabilistic classifiction task exhibited great improvement in performance when the underlying distribution parameters were not known--improvement which was absent when the parameter values were known in advance. Moreover, when parameter values were known the variability in preasymptotic decision criteria was radically reduced and asymptote was reached with far fewer observations. In many instances the observers were able to set their decision criteria in advance.

Furthermore, the expected value which resulted from initial boundary decisions when parameters were known equalled or exceeded the expected value of final decision cutoffs when parameter values were not known in advance. This was due to the fact that when equal prior probabilities obtained observers were able to place the midpoint of their uncertain ranges optimally between the two distribution means. Indeed, they probably would have selected this optimal cutoff as their single decision boundary had they been required to give only one. The fact that the optimal decision cutoff midway between the two means was not chosen in the Kubovy and Healy (1977) experiment or, for that matter, at the beginning of any signal detection

experiment, indicates that the means simply were not known in advance. They had to be learned during the observation process.

It can be concluded then that signal detection and pattern recognition experiments in which large amounts of preasymptotic variability is found and in which significant amounts of learning occurs resemble the conditions in this experiment in which distribution parameters are not known at the start and do not at all resemble the conditions in which they are known in advance. The existing model of signal detectability theory, which assumes that the likelihood ratio of even the first observation is computed with respect to the parameters of the two underlying distributions, must be re-evaluated in the light of this data and extended to provide a model of the way in which the parameters themselves are learned.

GIL is offered as a first approximation to this model of an ideal observer in the beginning of the observation process. GIL assumes that the observer samples information in order to estimate unknown parameter values and refines his estimates with each observation. It also assumes that the observer's decision criteria fluctuate optimally as a function of fluctuations in his current parameter estimates. The data

show that the subjects in this experiment did obtain information from the observations, even without feedback, since there was obvious improvement from first boundary decisions to later ones. That this improvement reflected improvement in parameter estimates is indicated by the fact that when parameter values were given in advance there was little fluctuation in decision boundaries and little improvement in expected value. However, even when parameter values were known in advance decision boundaries were not optimal. Therefore, the assumption of GIL that information is sampled to improve parameter estimates and that decision boundaries fluctuate appropriately as a function of fluctuations in these estimates is probably correct; but the assumption that they fluctuate optimally must be rejected.

CHAPTER IV

EXPERIMENT 3

Statement of Purpose

Results of the two previous experiments are consistent with the theory that parameter values are learned, but do not demonstrate that such learning actually occurred. Observers were never asked to state their estimates of population mean. Even if they had been able to give accurate estimates of these means, it would not have indicated that parameter abstraction occurs in other paradigms as well. The ability to estimate the means of two distributions of numbers, when given a sequential presentation of a sample from the mixture, is interesting in its own right and has many practical applications; but does not permit one to generalize to the process of parameter abstraction when perceptual or sensory stimuli are used.

Accordingly, a third experiment was designed employing random visual forms as stimuli rather than numbers. This experiment attempted to determine if the central tendency is abstracted from perceptual, ill-defined categories during category learning without feedback.

Moreover, it also attempted to test the hypothesis that category learning implies a process of parameter abstraction. Although GIL assumes a likelihood ratio decision rule, rather than the distance from the prototype rule prevalent in the prototype learning literature, for categories with equal variance, equal prior odds, and symmetric payoffs the decisions predicted by the two decision rules are equivalent. Therefore, if the central tendency, or prototype, is indeed abstracted, it is predicted that the closer a novel pattern is to the prototype the more likely it will be called a member of the category; patterns farthest away should be least likely to be placed in the category, while the probability of the membership of those in between these two extremes should form a continuous monotonic function of distance.

Method

Stimuli

Using the directions specified in Attneave and Arnoult (1956), two random patterns (the prototypes) were created and superimposed on a 10 x 10 grid of 100 cells. The cells which included the pattern in them were completely darkened, while the cells which comprised the background were left white. Two additional restrictions were made: 1) that each prototype be

composed of an approximately equal number of white and black cells, and 2) that half the cells in the first prototype would be the same as the corresponding cell in the second, while the other half of the cells of the first would differ from the corresponding cell in the second. These two restrictions required minor additional adjustments of the original patterns. Figure 9 show the two original prototypes.

Normal distributions around these two prototypes were approximated by creating probabilistic distortions of the patterns and using these distortions as the exemplars of each category. These distortions were generated by randomly "flipping" each of the 100 cells of the original prototype from white to black (or vice versa) with some pre-set probability, p. This led to two normal distributions with mean number of cells flipped being 100p and with variance equal to 100p(1-p). Since the prototypes are multi-dimensional and since a distortion can be in any direction away from the original, the original pattern can be conceived of as being in the center of the distribution with the distortions surrounding it. The degree of dispersion is, thus, created solely by the independent variable, p.

Stimuli were not repeated except by chance. The use of an essentially limitless sample size corrects

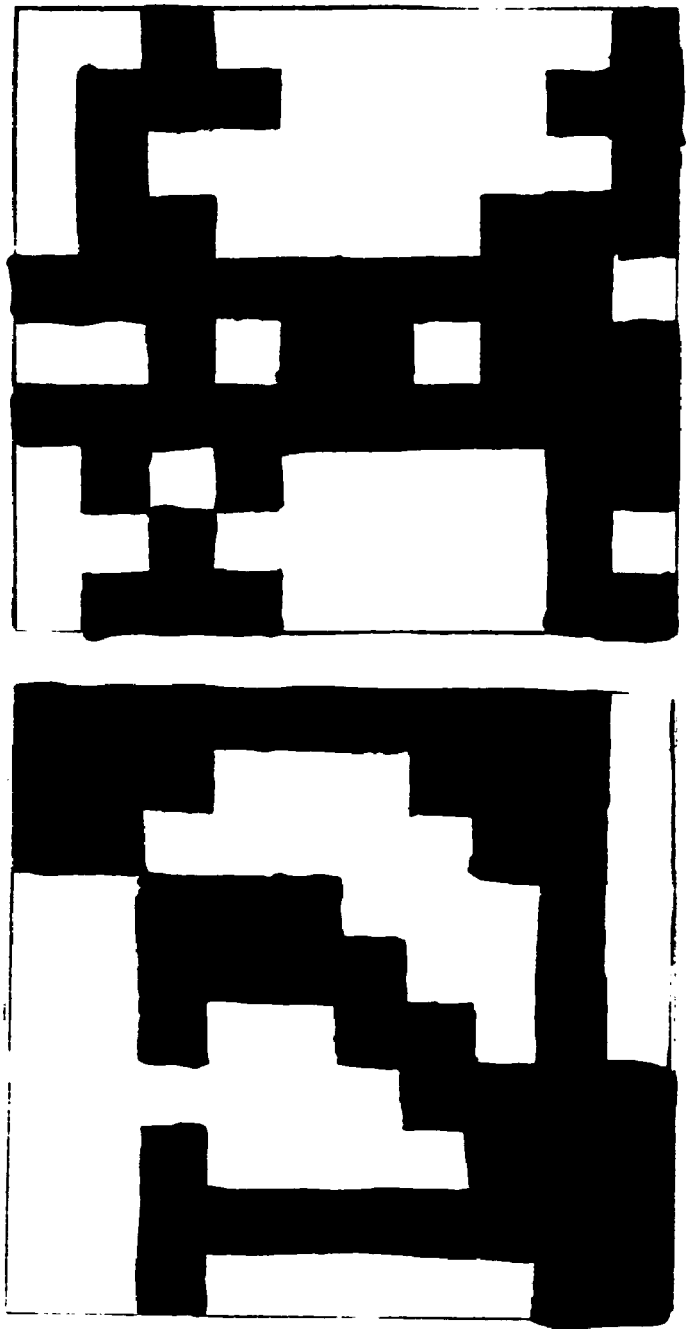


FIG. 9. The Two Prototypes Used in Experiments 3 and 4

a major methodological defect which was pointed out to the author by the late Wilson P. Tanner, Jr. (personal communication), a defect which occurs in most previous studies of prototype learning. Although it is always assumed in those studies that the prototype, from which the category exemplars have been derived, is somehow the central tendency of the category, this may fail, due to sampling error when only a small number of exemplars is used. Indeed most studies have used category sizes under ten, while Posner and Keele (1968) used only three exemplars in each of their categories.

Apparatus

The apparatus was the same as in the previous two experiments. The IBM 1800 was used to generate on-line random distortions of the two prototypes and display them to the subject on the television screen. Each subject received the distortions in his own separate random order. Classification decisions were communicated by the observer via pushbuttons placed in front of the screens.

Experimental Design

Subjects were run at three different values of p: .08, .10 and .12. Eight subjects were run at .08,

four subjects at .10 and four at .12. The same prototypes were used for all the groups. Each subject participated in a training and a transfer session. The training session lasted for 100 trials or until a criterion of 10 correct was achieved. The transfer task lasted for 100 trials. There was a five minute "stretch" between the two sessions.

Procedure

Observers were run individually in semi-isolated booths. They were told that in the first session they would see a series of abstract designs, each drawn by one of two different artists (Dave Smith and Keith Wilson). They were told they would have to discover which were the Smiths and which were the Wilsons, but that they would be able to do this since the Smiths would be similar to each other and different from the Wilsons and vice versa. They were told that the patterns would no doubt all appear the same to them at the beginning but that after a while they would be able to distinguish the Smiths from the Wilsons with little difficulty. They were also told that, in addition to the two dollars they had received for coming to the sessions they would be paid a penny for every correct classification and would lose a penny for every error.

Observers in this experiment did not receive feedback, neither did they see the original prototypes nor a defining sample. Because they did not receive feedback they were allowed to call the Smiths "Wilson" and the Wilsons "Smith" as long as they were consistent. Criterion was reached with ten wrong as well as with ten right.

The transfer task was the same as the training session except that here they had, in addition, a third alternative available--"Others." They were told that they would see both more of the Smiths and more of the Wilsons, but in addition they would see a great many patterns which were designed by other people. They were told to call these other patterns "Other." In actuality correct "Others" were random patterns; i.e., distortions of the original prototypes created with distortion probability, p,being .5. Again, all stimuli were generated on-line by the computer and were novel except for chance occurrence of a replication.

Results

As can be seen in Table 8, of the eight subjects trained on a .08 distortion level, six were judged as having learned the categories. That is, they either reached a criterion of 10 correct during the learning

TABLE 8

MEAN NUMBER OF TRIALS TO CRITERION AND PERCENT CORRECT AT EACH DISTORTION LEVEL

	Distortion Probability	Learning			Transfer
		No. Ss[a]	Mean No. Trials	Percent Correct	Percent Correct
	.08	6 (8)	62	71.8	76.8
	.10	3 (4)	70	71.4	78.3
	.12	2 (4)	40	79.3	75.0
MEAN	.10	11(16)[b]	60	72.8[a]	76.9
CHANCE				50.0	33.3

[a]Numbers in parentheses indicate total number trained on that distortion level

[b]This is the total, of course, rather than the mean

session or they were judged as responding significantly above chance at transfer--or both. Two of the eight did not reach criterion within 100 trials, but responded above chance during transfer (.66 and .83 correct, where chance is .33). Their data were retained for future analyses. Additional learning might have occurred during the transfer task even though no feedback was used and on one-third of the trials the distortions received were random. On the other hand, one subject reached criterion in 40 trials, but responded only slightly above chance at transfer (38 percent correct). He was judged as not having learned (or as having unlearned) the categories and his data were omitted during subsequent analyses. Only one of the eight observers neither reached criterion nor responded above chance at transfer. All of the four subjects trained at $p = .10$ learned the task, but unfortunately through an error one subject received only 50 trials at transfer instead of the usual 100 so his data could not be used.

At $p = .12$ only two of the four subjects were judged as having learned the categories.

Thus of the sixteen observers, twelve were judged to have learned to correctly classify these patterns even though they did not see either the

prototypes, nor a defining sample, nor did they receive feedback.

As can be seen in Table 8 there were no appreciable differences in either number of trials to criterion, percent correct during learning, or percent correct at transfer among the three groups, so their data were massed for the additional analyses.

Table 8 shows that almost 77 percent of the patterns were correctly identified at transfer even though all the patterns were completely new to the observer. Only the prototypes and the distortion level remained the same as in training. The location of these errors can be seen in Table 9. Cell entries are percent of the total trials falling into that cell. The data are over all 1100 trials (eleven subjects x 100 trials each). As can be seen from Table 9 the 77 percent correct responses appear to be evenly divided among the three alternatives, with the chance of giving a true "Other" only slightly favored. χ^2 is significant at $p \leq .01$. On the other hand, the errors do not appear to be evenly divided; the chance of incorrectly calling a pattern "Other" is greater than the chance of confusing the categories themselves. Gibson's assertion (Gibson, 1969) that perceptual learning can occur without external feedback is therefore supported by the data.

TABLE 9

STIMULUS-RESPONSE CONTINGENCY TABLE
FOR THE TRANSFER TASK

	Response I	Response II	Response "Other"
Category I	24.2	2.2	5.1
Category II	.8	23.9	7.9
Random	3.2	3.9	28.8

NOTE: Cell entries are percent of total
$\chi^2 = 981.28$

Learning without feedback is not restricted to numerical stimuli alone.

Given that observers were learning the categories, the next question is: What is it about the categories that was learned? If the observers were abstracting the prototypes and learning the central tendencies as both GIL and the prototype abstraction theories suggest, then (given equal variances, equal prior odds and symmetric payoffs) the probability that a pattern will be called "Other," i.e., not a member of the category, will increase with its distance from

the prototype. On the other hand, if observers were only learning a decision criterion during the learning session all the patterns on one side of this criterion should have the same probability of being called "Other," and the curve would be flat with respect to distance from the prototype.

Figure 10 shows the percent of patterns called "Other" as a function of the distance from the correct prototype. Distance here is measured in number of "flips," i.e., number of cells in the distorted pattern which distinguish it from the original. This corresponds to a city-block measure of distance (Reed, 1972; Shepard, 1962a, 1962b; Kruskal, 1964). Each point on the graph is the mean percentage, averaged over each subject's responses to patterns at that block of distances away. As can be seen from the figure, the mean percentage classified as "Other" increases almost linearly with the distance from the prototype. Only ten percent of the patterns, on the average, are called "Other" when only 0-4 cells are flipped, while 87 percent are called "Other" when from 40-44 cells distinguish it from the original. This proportion starts to decrease as the patterns start to approach the second prototype. The gap in the curve represents the gap between the patterns derived from a .12 distortion probability and

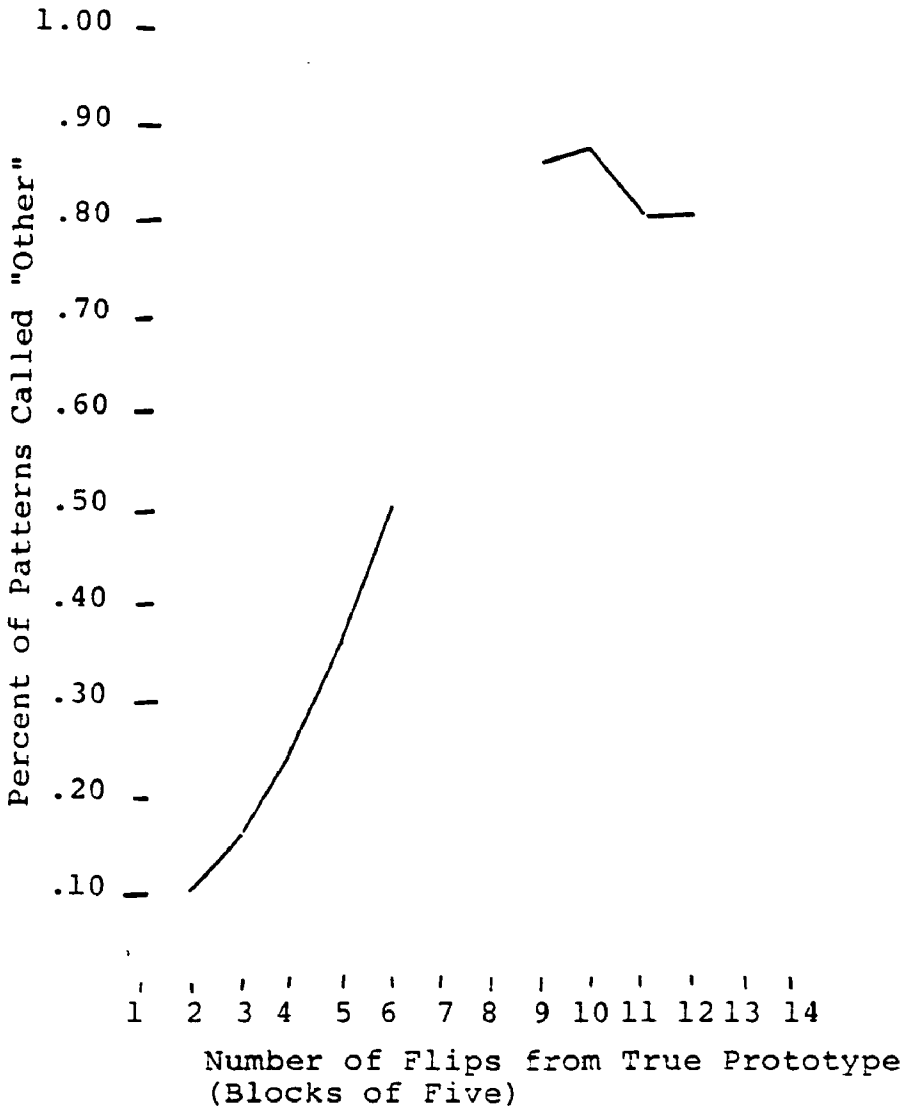


FIG. 10: Percent of Patterns Called "Other" as a Function of the Distance From the True Prototype

those resulting from a .50 distortion level. There simply were no patterns between 25 and 35 flips away from the originals.

These results indicate that these subjects did abstract a prototype from the data during the observation process even though feedback was not given, and they did not see either the prototypes themselves or a defining sample. Since the curve is obviously not flat with respect to distance from the prototype, more was learned than a decision criterion. The central prototype must have also been learned.

More generally, however, GIL predicts that classifications are based on likelihood ratio and not on linear distance from the prototype. A strict likelihood ratio decision rule is deterministic, not probabilistic: patterns are on one side or the other of the likelihood ratio criterion. Percent classified as "Other," when expressed as a function of likelihood ratio, is predicted to be a step function.

The likelihood ratio (LR) of a pattern is defined as

$$LR = \frac{P^{N} (1-P)^{T-N}}{P^{M} (1-P)^{T-M}}$$

where P is the distortion probability used to generate the pattern (.08, .10 or .12),

where T is the Total number of cells in the pattern (100 here),

where N is the number of cells which distinguish the pattern from Prototype I,

and where M is the number of cells which distinguish the pattern from Prototype II.

As can be seen from Figure 11 a step function of log likelihood ratio does describe the data fairly well. It appears that when expressed as a function of log likelihood ratio the patterns fall into two groups: those not called "Other" and those called "Other." It cannot be determined, however, if the function really is a step function and if the subjects were actually basing their decision on a strict likelihood ratio decision criterion since patterns with log likelihood ratios falling in the center of the curve were not generated. The curve could merely be linear with two large tails at the ends. However, observers are no doubt not altogether indifferent to likelihood ratio since high ratios in favor of a prototype generate a very low probability of calling the pattern "Other," while low ratios in favor of a prototype generate a high probability of being called "Other."

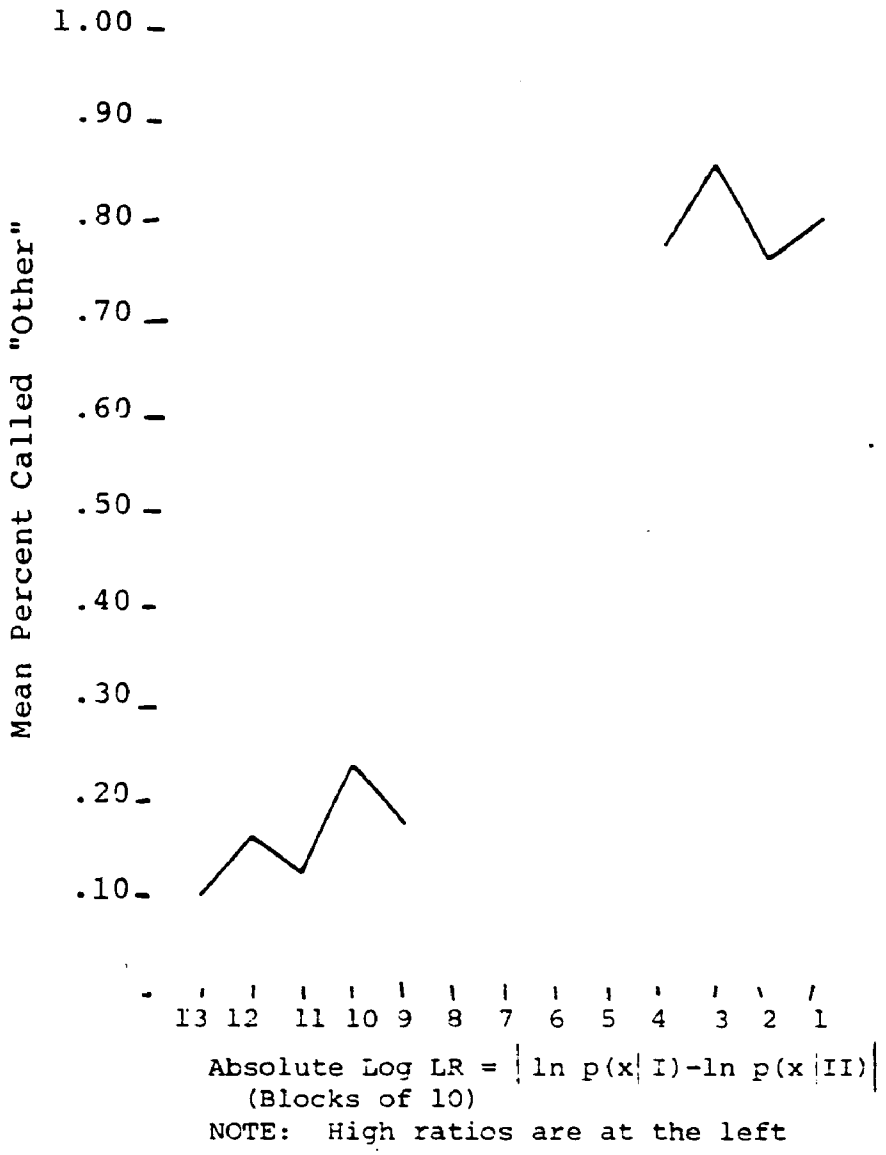


FIG. 11: Mean Percent of Patterns Called "Other" as a Function of Likelihood Ratio

This behavior would maximize the observer's expected value since low ratios should indicate near indifference for the two categories and a greater likelihood that "Other" is the correct response. However, since these classification data are also equally well described by the distance from the prototype decision rule, it is not possible to determine whether subjects are following a likelihood ratio or a distance from the prototype decision rule.

Discussion

These results do not permit a distinction between the likelihood ratio decision rule and the distance from the prototype decision rule since both describe the data. They do, however, permit a conclusion regarding what is learned in an ill-defined category learning task. These results indicate that the central tendency of the category, the category prototype, is learned. Classification decisions could not be based on a distance from the central prototype unless the prototype had been abstracted first. Neither could classification decisions be based on likelihood ratio if distribution parameters were not first learned, since there was no way of knowing them in advance. A model of category learning which is not

based on prototype abstraction or parameter estimation can not adequately explain these results. Moreover, a model which depends on feedback for learning to occur is also inadequate.

Two major problems remain unsolved, however. First, the question of whether observers use a likelihood ratio decision rule or a distance from the prototype decision rule remains unanswered. A strict interpretation of the distance model as described by Reed (1972, 1973) implies a deterministic rather than a probabilistic model. A novel stimulus is either closer to one prototype or to the other--unless they are equidistant--and should be classified into the category with the closest prototype with probability 1.00. The fact that the probability of being called "Other" increased nearly linearly with distance from the prototype provides strong evidence that a distance rule is operating; but that it is a probabilistic function of distance, not a deterministic one. A probabilistic modification of the original rule would be more consistent with the data. As described above, a deterministic likelihood ratio decision rule cannot be ruled out, however, since the same data when described as a function of log likelihood does produce a step function.

A second issue left unresolved is the nature of the boundaries of ill-defined categories. Rosch and her colleagues (Rosch, 1973, 1977; Rosch and Mervis, 1975) as well as Hersh and Carmezza (1976) describe ill-defined categories as lacking clear cut boundaries and as having a continuous membership function. Since patterns intermediate in distance from the prototypes and possessing intermediate likelihood ratios were missing from the data, it was not possible to determine if a natural boundary would be found to exist or if a continuous membership function would be maintained.

CHAPTER V

EXPERIMENT 4

Predictions and Hypotheses

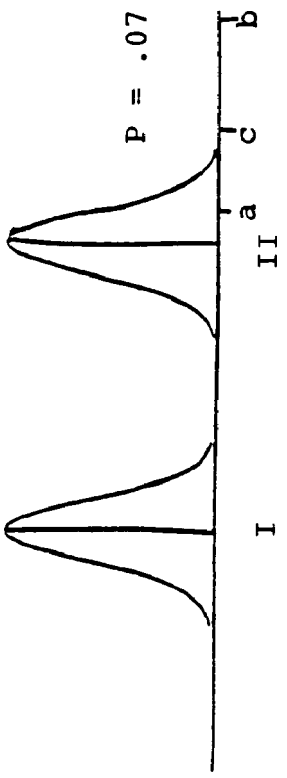
The Decision Rule

A fourth experiment was designed to shed some light on two problems left unresolved by the previous three studies. The first is the nature of the decision rule, and the second is the nature of the category boundary. Prototype theories adopt, either implicitly or explicitly (Reed, 1972, 1973; Posner and Keele, 1968, 1970) the "distance from the prototype" decision rule, which states that a novel item is classified into that category whose prototype is closest to it. The dynamic cutoff models, operating from a background of signal detectability theory, assert, on the other hand, that observers use an optimal likelihood ratio decision criterion. However, when equal prior odds, symmetric payoffs, and equal variances obtain, the optimal single likelihood ratio decision boundary is exactly midway between the two category means (for a further discussion see, e.g., chapter one above; Green and Swets, 1966; Swets, Tanner and Birdsall, 1961). Thus, if we

were to equate category prototypes with the distribution means, as is commonly done (Reed, 1972, 1973; Homa, et. al., 1973, 1975, 1976), it would be easily seen that in this basic case the two decision rules are identical. In cases other than this they can be distinguished, however. Manipulation of category variability (or variance), for example, may cause the optimal likelihood ratio decision criterion to deviate from center, since likelihood ratio is a function of category variability as well as of central tendency. A distance from the prototype decision criterion will remain constant, however, as it is independent of category variability. Novel patterns are classified only according to their distance from the prototype (either a Euclidean or a city-block metric) and so are not affected by changes in category variability so long as the prototypes remain the same.

The differences in predictions between the models which occur when variability is manipulated become more salient when the classification "Other" is available. Figure 12 shows the differences in classifications predicted by the two models.

In the upper portion of the figure there are two normal distributions with a low variance, which would occur with a distortion probability of p = .07



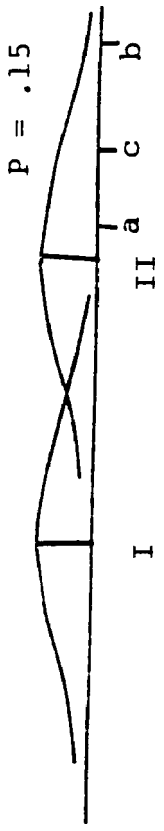


FIG. 12: A Description of the Difference in Preductions Regarding the Classification of Patterns a, b and c for Low and High Variance Distributions When the Category "Other" is Available

in the previous experiment. The lower portion of the figure shows two normal distributions with a larger variance which would occur if p = .15. It should be noted that both the distributions labeled 'I' have the same mean (or prototype) and both distributions labeled 'II' have the same mean (or prototype). The letters a, b, and c refer to stimuli lying on their respective points on the decision axis. The distance model predicts no difference in classifications between the low and the high variability groups for the three stimuli, since the patterns are the same distance to the prototype in both groups. Further, the model which describes category learning as purely criterion learning (Dorfman, et al., 1971, 1975; Kac, 1962) would also predict no difference between the two groups since the same decision criterion would have been learned by both groups during the training period.

Only a model in which the variances of the distributions, as well as the central tendencies, are learned during training and in which a likelihood ratio decision rule is used would predict a difference between the two groups. Such a model, GIL, predicts that both groups will classify pattern a as belonging to Category II and that both groups will call pattern b "Other." In addition GIL predicts that the two groups will differ in their classification of patterns

an intermediate distance from the prototype, namely patterns such as c. GIL predicts that such patterns will be called members of Category II by the high variability group but will be called "Other" by the low variability group. The distance of pattern c to prototype II is the same for both groups, but the likelihood that the item is generated by a distortion of prototype II should be viewed as being extremely low by the low variability group, but viewed as quite possible by the high variability group. If a difference is found between the groups in classification of intermediate distanced patterns, while no difference between them is found for near or far patterns it will indicate that observers learn the prototypes and the degree of variability when they learn to recognize novel members of a category and that they employ a likelihood ratio decision rule. It will also indicate that the 'distance from the prototype' decision rule and the theory that only decision criteria are learned during the observation process are inadequate.

The Category Boundary

Besides determining the decision rule, it was also possible to examine the nature of the category boundary. According to Rosch and her co-workers (Rosch,

1973, 1977; Rosch and Mervis, 1975) as well as Hersh and Carmezza (1976) category boundaries are fuzzy, that is, the membership function is a continuous, monotonic function of distance from the prototype. If the probability of saying "Other" is found to be an increasing function of distance, support will be given to this hypothesis. On the other hand, if a step function occurs, it will indicate that this type of ill-defined category is bounded; moreover, the location of the step will indicate the location of the boundary.

Method

Subjects

Subjects were twenty-two male undergraduate students of The University of Michigan. They were obtained from the Human Performance Paid Subject Pool. None of them had participated in the earlier experiments.

Stimuli, Apparatus and Procedure

The stimuli, apparatus and the procedure were the same as in the previous experiment. Stimuli were generated on-line by the computer as previously.

Experimental Design

The experimental design was the same as in the previous experiment except that here there were two separate groups of subjects. Ten observers were trained on the low variability task in which the distortion level, p, was .07; and twelve were trained on the high variability task in which p was .15. Both groups saw distortions of the same two prototypes, only the distortion level differed between the groups.

Because several subjects failed to learn in the previous experiment the instructions were modified slightly. Observers were told that there were two original abstract designs each created by a different artist, and that the computer would make a series of random distortions of them. Observers were told to sort distortions of one of the originals from distortions of the other. They were told they would be able to perform this task after some experience with the patterns because the distortions of one original would, in general, be more similar to each other than to the distortions of the other. Similarity was not defined.

Feedback was used during the training task as well. The reasons for using feedback were many, but none were felt to be crucial to the experimental design. First, it was felt that a valid criticism of the

previous experiments may be that category parameters may have to be estimated only when feedback is not given, but when feedback is given a decision criterion might be learned. In addition, the prototype learning studies reviewed above have also used feedback, so that in order to demonstrate the validity of the model in their experimental situations it was thought advisable to make our experimental setting as nearly like theirs as possible. It was also feared that many subjects would fail to learn the categories at a 15 percent distortion rate if feedback were not used. Finally, since the previous three experiments demonstrated that perceptual learning does occur without feedback it did not seem necessary to demonstrate it again.

Training lasted until a criterion of ten successive correct responses was achieved or until a maximum of 200 trials occurred.

Training was followed by the transfer task. Since the two decision rules being tested differed only with regard to patterns intermediate distances from the prototype, it was necessary to ensure that such patterns,

as well as patterns very near to and very far from the prototype, in fact appeared during transfer. This was accomplished by letting one-fourth of the stimuli at transfer be distortions of the prototype at p = .10, while one-fourth were distorted at p = .20, one-fourth at p = .30, and one-fourth at p = .40. In this way the number of 'flips' which separated the stimulus patterns from the original prototypes would run the gamut from 0 'flips' to about 60. (Patterns which have a great deal more than 50 percent of their cells 'flipped' begin to look like the original again. A 100 percent distortion, for example, looks like the original with figure-ground reversed.) The distribution of these flips approaches the normal as the number of patterns increases, with an expected mean number of flips of 25. An incidental result of this manipulation was that there were now no true randoms. All the patterns were actually distortions of the prototypes.

The instructions to the observers during transfer were the same as in the previous experiment except that, again, they were told that they were seeing computer generated distortions of the two original patterns along with an unknown number of patterns drawn by other people. As in the previous experiment there was no feedback in the transfer task. Also as in the

previous experiment, observers never saw the actual prototypes, nor a defining sample.

Results

Table 10 shows the number of observers who reached criterion at each distortion level. All ten of the observers trained at the .07 distortion level reached criterion while only nine of the twelve reached criterion at a distortion level of .15. Moreoever, those who learned in the high variability group averaged 77.78 trials to criterion, while the low variability group took only 24 trials before criterion was reached. This difference is significant at the .01 level. Both the proportion of observers who learned to criterion, and the number of trials they required shows that increasing category variability increases learning difficulty even though category prototypes remained the same for both groups.

Table 10 also shows the percent correct during both learning and transfer for the two groups. The pattern of correct responding which existed during training reversed itself upon transfer. During training the low variability group displayed a greater percent correct than the high variability group, while at transfer the high variability group displayed a greater

TABLE 10

MEAN NUMBER TRIALS TO CRITERION AND PERCENT CORRECT
AS A FUNCTION OF DISTORTION LEVEL

	Learning			Transfer
Distortion Probability	Number of Subjects*	Mean Trials to Criterion	Percent Correct	Percent Correct
.07	10 (10)	24.00	92.6	.356
.15	9 (12)	77.78**	66.9	.477

*Numbers in parentheses indicate total number trained on that distortion level

**$p \leq .01$

percent correct. This replicates the result of Posner and Keele (1968, 1970). It must be pointed out, however, that there were no correct "Others" as there were in the previous experiment. Here percent correct applies only to correct identification of the category of the pattern.

Table 11 gives the distribution of responses during the transfer task and may help to explain this interraction effect in percent correct. This table breaks down the "errors" into percent called "Other" and the true errors--the confusions. As predicted by GIL more patterns are called "Other" by the low variability group than by the high variability group. In addition, the low variability group exhibited fewer confusions between the categories than did the high variability group. A look back at Figure 12 illustrates why this result is predicted by GIL: there is simply more category overlap for the high variability group than for the low one.

A true test of the GIL model is in the percent of patterns called "Other" as a function of distance from the prototype. Both the dynamic cutoff learning models and the distance from the prototype classification models predict no difference between the groups. GIL predicts a difference only for patterns an intermediate distance from the prototype.

TABLE 11

STIMULUS-RESPONSE CONTINGENCY TABLE FOR EACH DISTORTION LEVEL AT TRANSFER

.07	Response I	Response II	Response "Other"	
Category I	17.4	2.4	29.0	$\chi^2 = 247.73$
Category II	1.9	18.4	30.9	
.15				
Category I	23.4	3.7	21.0	$\chi^2 = 269.24$
Category II	3.6	24.2	24.1	

NOTE: Cell entries are percent of total

Figure 13 displays the percent called "Other" as a function of the distance from the prototype. Distance is expressed in number of cells 'flipped', corresponding to a city-block metric. As in the previous experiment the solid line represents the low variability group while the hatched line represents the high variability group. As can be seen in the figure there is no difference between the two groups in their classification of patterns close to the prototype. Indeed, the response curve is identical for the two groups until a distance of twenty 'flips' is reached (block four). At a distance of from 20 to 25 'flips' (block five) the two groups separate in the direction predicted by GIL. Although not yet significant, the low variability group calls more of the patterns at this distance "Other" than does the high variability group. The difference between the two groups increases with the distance from the prototype. For patterns 25-35 'flips' away (blocks six and seven) the difference between the two groups is significant at the .01 level. Significantly more of these <u>intermediate distanced</u> patterns are called "Other" by the low group than by the high variability group. As the distance from the prototypes increases even more the difference between the two groups lessens. At a

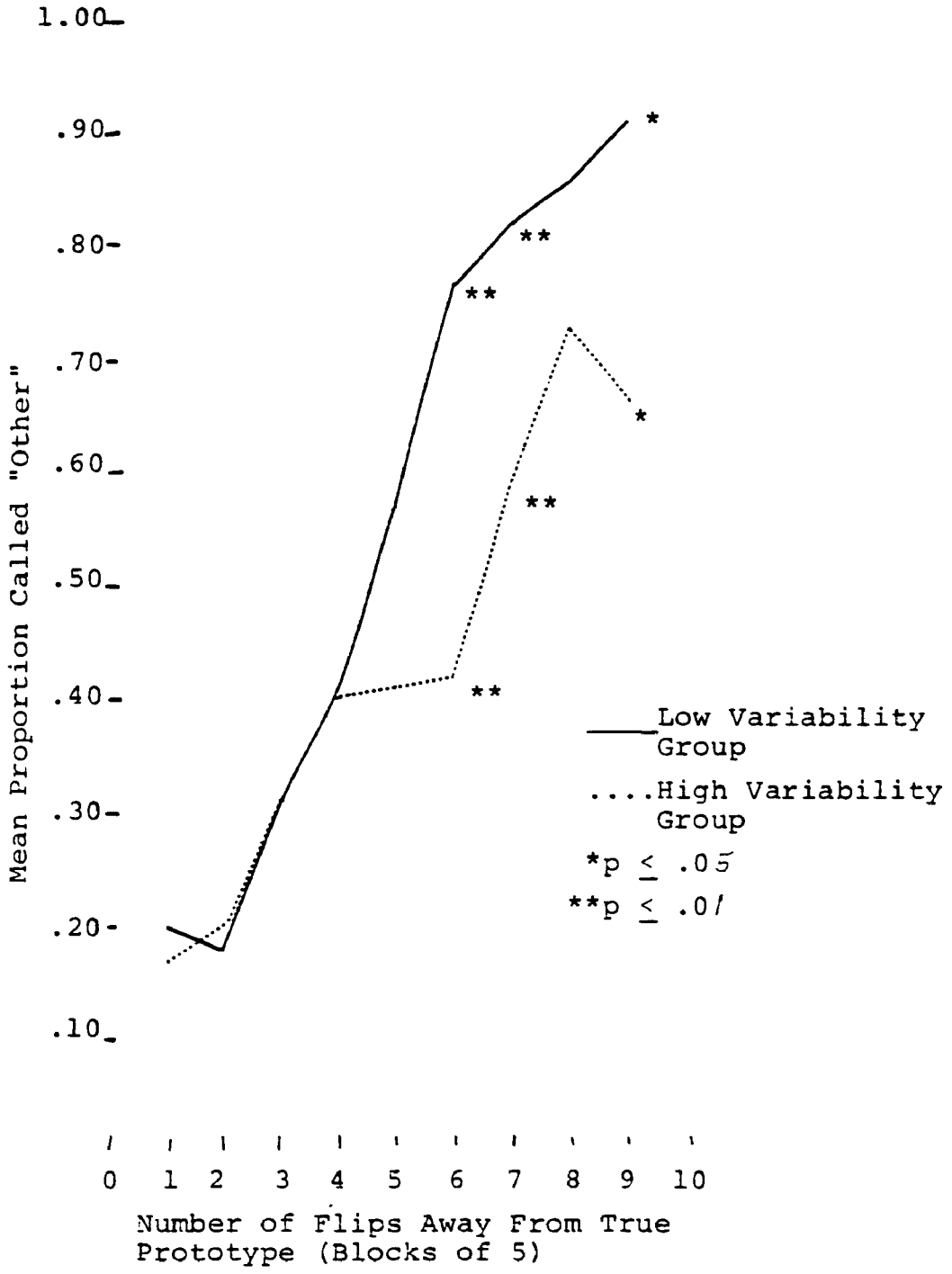


FIG. 13: Percent Patterns Called "Other" as a Function of Distance from the True Prototype

distance of 35-40 'flips' (block eight in the figure) away this difference is no longer significant, both groups call the patterns "Other" with the same high degree of frequency. This pattern of responding was not predicted by any of the model reviewed above, except GIL. Unexpectedly, the functions did not converge at distance far from the prototype. This no doubt results from the increase in confusion errors in that group. Patterns far from the original prototype are seen as belonging to the second, erroneous, category by the high variability group and as "Other" by the low variability group. This difference was significant at the .05 level. This pattern of responding demonstrates that the subjects in this experiment learned the category variability (variance) as well as the category prototype (mean) during the observation process, and that they employed a likelihood ratio decision criterion.

A striking difference between the two variability groups is apparent when the percent classified as "Other" is described as a function of log likelihood ratio, as shown in Figure 14. High ratios in

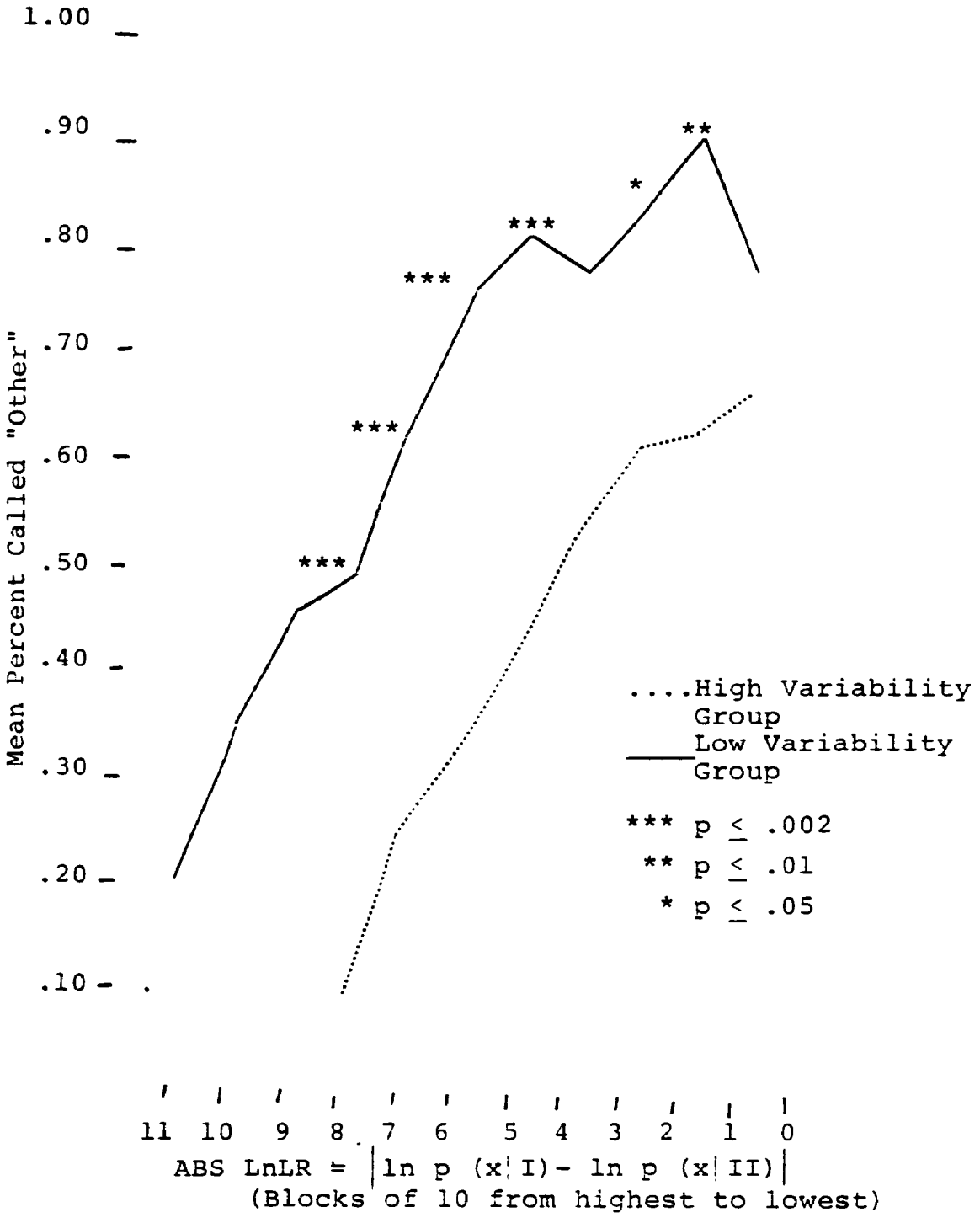


FIG. 14: Percent Called "Other" as a Function of Absolute Log Likelihood Ratio

favor of the true prototype are at the left, while log likelihood ratios approaching zero are at the right of the abscissae.

It should be pointed out that because of the method of calculating likelihood ratio (see Chapter IV) the very high likelihood ratios observed by the low variability group are simply not going to be perceived by the high variability group even though the patterns are the same for both. This can also be observed by a glance at Figure 12. The distortion level used in the determination of the likelihood ratio perceived by the observer at transfer is the distortion level he was trained on, not the actual distortion probability which was used to generate the pattern.

Considering only the likelihood ratios that both groups shared, however, it can be seen in Figure 14 that at every such level the low variability group was significantly more likely to call the pattern "Other" than was the high variability group. At likelihood ratios not close to one this difference was found significant at the .002 level. At likelihood ratios nearer to one separation was still highly significant. Only at ratios quite close to one did the separation become non-significant. Conversely, the low variability group required a significantly greater likelihood ratio

for each level of acceptance into the category than did the high variability group. Thus the low variability group had a stricter likelihood ratio membership criterion, even though payoffs and prior probabilities were the same for both groups. It can be seen in the figure that the membership function is a continuous, monotonic function of log likelihood ratio for both groups.

Neither Figures 13 or 14 are step functions. Rather these membership functions are continuous, with membership decreasing monotonically with distance from the prototype and with decreases in log likelihood ratio. This supports the findings of Rosch and her colleagues (Rosch, 1973, 1977; Rosch and Mervis, 1975) as well as those of Hersh and Carmazza (1976), that the membership function in ill-defined categories is not bounded, but fuzzy and continuous.

Discussion of Results

The results of this experiment indicate that the model, GIL, is a useful description of the process of learning ill-defined categories and of classifying novel stimuli into them. It is evident from the results that both central tendency and degree of category variability were learned by the observers during the training process and that knowledge of these two distribution parameters was employed appropriately in the

use of a likelihood ratio decision rule. Manipulation of category variability affected the membership function in a way which was predicted only by this model. The models of category learning which describe learning as prototype abstraction only,or as criterion learning only,could not predict the significant effect a change in category variability would have on the membership function.

These results provide an answer to the question of what is learned during perceptual learning,asked by Gibson and Gibson (1955) over twenty years ago. They support the finding of Attneave (1957) that subjects learn three characteristics of a category: (1) its central tendency; (2) the dimensions along which its members differ; and (3) the degree of dispersion among the category members. Although this assertion was widely known in the literature and had been substantiated by current investigators (Posner and Keele, 1968, 1970; Homa, et al., 1973, 1975, 1976), the implications of this for decision-making, i.e., for the classification of novel stimuli, was not understood. These authors could not predict the effect within category variability would have on classification decisions, since they were operating under a model of decision-making for which knowledge of category variability was

irrelevant, i.e., a model based on distance to the prototype. This model, which grew out of the work of Shepard (1962a, 1962b, 1964) and Kruskal (1964), defined similarity in terms of distance using either a Euclidean or a city-block metric--neither of which took account of the degree of within-category variability.

The results of Reed (1972) were also consistent with the theory of a likelihood ratio decision rule. He found that on the average, only 76 percent of his prototypes were correctly identified when they were close together, but that 90 percent, were correctly identified when they were far apart. A decision rule based only on distance to the prototype cannot predict these results, since the prototype is always zero distance to itself. However, the proportion of observers who misclassified the prototype may have been a function of the likelihood that the prototype was itself a distortion of the other. A likelihood ratio decision rule predicts this result--many more confusions of the distribution means are expected when d' is low than when it is high. Since the distance models do not allow for any misclassifications of the prototypes, Reed could not explain his results.

The likelihood ratio decision rule predicted by GIL, on the other hand, was a strict. deterministic decision criterion. This implies a step in the membership function, as occurred in Experiment 3, which in turn, implies a strict finite, category boundary based on likelihood ratio. The fact that a step function did not occur in this experiment argues for a modification of this theory. It appears that the membership function of ill-defined categories is unbounded (except by the occurrence of the next category), implying a probabilistic rather than a deterministic decision rule. The model which appears to be most descriptive of these data indicates that the probability of assigning a novel member to an ill-defined category is a monotonic function of likelihood ratio. Support for this hypothesis of a probabilistic decision model is only tentative, however, since it is derived from grouped, rather than individual, data. It may be that individual observers follow a strict deterministic, likelihood ratio decision criterion, but that each person sets his criterion at a different location on the decision axis. This could lead to a continuous curve when all were averaged together. However, inspection of the individual data revealed the same continuous pattern of responding for the individuals as exhibited by the group as a whole.

But unfortunately, the data for each observer is not sufficient to warrant drawing any strong conclusions from it, as there were only 100 observations for each spread over eight or twelve levels of likelihood ratio. The important question of whether the decision rule is deterministic or a probabilistic function of likelihood ratio must await future experimentation. That it is based on likelihood ratio is strongly indicated, however.

The use of a likelihood ratio decision rule rather than one based on distance to the prototype may indicate that judgments of physical similarity may differ fundamentally from judgments of class membership. Stimuli judged for physical similarity are generally expressed as points in a Cartesian space with psychological distance defined by a Minkowski metric (Shepard, 1962a, 1962b, 1964; Kruskal, 1964; Reed, 1972, 1973; Garner, 1974). Psychological distance between stimuli can evidently not be expressed in such a metric when degree of category membership is measured, unless category variances are equal, as it does not take into account category variability. The likelihood ratio rule implies that a metric based on category standard deviation may be more appropriate here.

These four experiments, taken together, have demonstrated that perceptual learning is self-directing.

and self-terminating and can occur without external reinforcement, as suggested by Gibson (1969). They have demonstrated that learning ill-defined categories includes both the degree of variability as well as the central tendency of the category, and that it most probably occurs thorough estimation and reestimation of these parameters during the observation period. In addition, the experiments have demonstrated that novel items are classified according to a likelihood ratio decision rule, and that the observer uses his current estimates of both the degree of category variability as well as its central tendency in forming his criterion. Models which describe learning as criterion learning only (Kac, 1962; Dorfman and Biderman, 1971; Dorfman, Saslow and Simpson, 1975; Thomas, 1973, 1975), or as prototype abstraction only (Posner and Keele, 1968, 1970; Reed, 1972, 1973; Homa and Chambliss, 1975; Homa, et al., 1973; Homa and Vosburgh, 1976; Goldman and Homa, 1977); and which describe the decision rule as based only on the distance to the prototype (Posner and Keele, 1968, 1970; Reed, 1972, 1973; Homa and Chambliss, 1975; Homa et al., 1973; Homa and Vosburgh, 1976; Goldman and Homa, 1977) are unable to account for the results of these experiments. Only

GIL, a model in which learning is based on parameter estimation and which employs a likelihood ratio decision rule, proved to be more consistent with the data.

APPENDIX

A method is outlined here for obtaining maximum likelihood estimates of the five parameters of a mixture of two normal distributions from a sample. An observation, z_n, representing a value on some variable z on trial n is drawn randomly from an unknown mixture of two normal distributions. The log of the likelihood function of the observation is then expressed in terms of the (five) unknown distribution parameters: the two means, the two variances and the single proportion parameter:

Let us assume a unidimensional variable, Z, such that

$$-\infty \leq z \leq \infty$$

Let us assume the following probability density for Z

$$(1) \qquad p(z_i) = P\,(z_i|P,\mu_x,\sigma_x,\mu_y,\sigma_y)$$

$$= P\, f(z_i|\mu_x,\sigma_x) + (1-P)\, f(z_i|\mu_y,\sigma_y)$$

where $f(z_i|\mu,\sigma) = (2\pi)^{-1/2}\sigma^{-1} \exp\,[-(z_i-\mu)^2/2\sigma^2]$

and where

$$i = 1,\ldots,n$$

After a random sample of n observations we have the following likelihood function

$$(2) \qquad Ln(z_1,\ldots,z_i,\ldots,z_n|P,\mu_x,\sigma_x,\mu_y,\sigma_y) = \prod_{i=1}^{n} p(z_i|P,\mu_x,\sigma_x,\mu_y,\sigma_y)$$

taking the logarithm

$$\log Ln = \sum_{i=1}^{n} \log \; p(z_i | P, \mu_x, \sigma_x, \mu_y, \sigma_y) \tag{3}$$

Maximum likelihood estimates of the parameters are then found by computing the five partial derivatives of the likelihood function, setting each to zero, and solving the five resulting equations through Newton's iterative procedure for solving systems of non-linear equations (see, e.g., Edwards, 1973, p. 181-186). Derived estimates of the distribution parameters are then used to estimate the likelihood ratio of each observation, while the estimate of the mixture parameter is used to estimate prior odds (if not known).

Posterior odds of the observation coming from each population are then computed as a function of prior odds and likelihood ratio, using Bayes' Theorem, and used along with payoffs to estimate the expected value of assigning the observation to the available categories. The category which maximizes the observer's expected value is chosen. This process is repeated after each observation, current parameter estimates providing the starting point for the next series of iterations. The observations are used to estimate and to re-estimate the distribution parameters of both categories, while at the same time current estimates are used to classify the observations themselves.

REFERENCES

Attneave, F. Transfer of experience with a class schema to identification learning of patterns and shapes. Journal of Experimental Psychology, 1957, 54, 81-88.

Attneave, F. and Arnoult, M. D. The quantitative study of shape and pattern perception. Psychological Bulletin, 1956, 53, 452-471.

Barlett, F. C. Remembering: A study in experimental and social psychology. London and New York: Cambridge University Press, 1932.

Bruner, J. S. On perceptual readiness. Psychological Review, 1957, 64, 123-152.

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

Carterette, E. C., Friedman, M. P., and Wyman, M. J. Feedback and psychophysical variables in signal detection. Journal of the Acoustical Society of America, 1966, 39, 1051-1055.

Dorfman, D. D., and Biderman, M. A learning model for a continuum of sensory states. Journal of Mathematical Psychology, 1971, 8, 264-284.

Dorfman, D. D., Saslow, C. F., and Simpson, J. C. Learning models for a continuum of sensory states reexamined. Journal of Mathematical Psychology, 1975, 12, 178-211.

Edwards, Jr., C. H. Advanced Calculus in Several Variables. New York: Academic Press, 1973.

Franks, J. J. and Bransford, J. D. Abstraction of visual patterns. Journal of Experimental Psychology, 1971, 90, 65-74.

Garner, W. R. The processing of information and structure. New York: Wiley, 1974.

Gibson, E. J. Improvement in perceptual judgments as a function of controlled practice or training. Psychological Review, 1953, 50, 401-431.

_____. Principles of perceptual learning and development. New York: Appleton, 1969.

Gibson, J. J. and Gibson, E. J. Perceptual learning: differentiation or enrichment? Psychological Review, 1955, 62, 32-41.

Goldman, D. and Homa, D. Integrative and metric properties of abstracted information as a function of category discriminability, instance variability, and experience. Journal of Experimental Psychology: Human Learning and Memory, 1977, 3, 375-385.

Green, D. M. and Swets, J. A. Signal detection theory and psychophysics. New York: Wiley, 1966.

Hammerton, M. An Investigation into changes in decision criteria and other details of a decsion making task. Psychonomic Science, 1970, 21, 203-204.

Hersh, H. M. and Caramazza, A. A fuzzy set approach to modifiers and vagueness in natural language. Journal of Experimental Psychology: General, 1976, 105, 254-276.

Homa, D. and Chambliss, D. The relative contributions of common and distinctive information on the abstraction from ill-defined categories. Journal of Experimental Psychology: Human Learning and Memory, 1975, 104, 351-359.

Homa, D., Cross, J., Cronell, D., Goldman, D., and Schwartz, S. Prototype abstraction and classification of new instances as a function of the humber of instances defining the prototype. Journal of Experimental Psychology, 1973, 101, 116-122.

Homa, D. and Vosburgh, R. Category breadth and the abstraction of prototypical information. Journal of Experimental Psychology: Human Learning and Memory, 1976, 2, 322-330.

Kac, M. A note on learning signal detection. IRE Transactions on Information Theory, 1962, IT-8, 126-128.

Kruskal, J. B. Multidimensional scaling by optimizing goodness of fit to a non-metric hypothesis. Psychometrika, 1964, 29, 1-27.

Kubovy, M. and Healy, A. The decision rule in probabilistic categorization: What it is and how it is learned. Journal of Experimental Psychology: General, 1977, 106, 427-446.

Kubovy, M., Rapoport, A. and Tversky, A. Deterministic vs. probabilistic strategies in detection. Perception and Psychophysics, 1971, 9, 427-429.

Larkin, W. Response mechanisms in detection experiments. Journal of Experimental Psychology, 1971, 91, 140-153.

Lee, W. and Janke, M. Categorizing externally distributed stimulus samples for three continua. Journal of Experimental Psychology, 1964, 68, 376-382.

Levine, M., Leitenberg, H., and Richter, M. The blank trials law: the equivalence of positive reinforcement and non-reinforcement. Psychological Review, 1964, 71, 94-103.

Lewin, Irwin P. Information integration in numerical judgments in decision processes. Journal of Experimental Psychology: General, 1975, 104, 39-53.

Luce, R. D. Detection and recognition. Vol. 1 in R. D. Luce, R. R. Bush, and E. Galanter (Eds.). Handbook of Mathematical Psychology. New York: Wiley, 1963.

Mood, A. Introduction to the Theory of Statistics. New York: McGraw-Hill, 1950.

Neisser, U. Cognitive Psychology. New York: Appleton-Century-Crofts, 1967.

Norman, M. F. Markov Processes and Learning Models. New York: Academic Press, 1972.

Oldfield, R. C. Memory mechanisms and the theory of schemata. British Journal of Psychology, 1954, 45, 14-23.

Peterson, M. J., Meagher, R. B., Chait, H., and Gillie, S. The abstraction and generalization of dot patterns. Cognitive Psychology, 1973, 4, 378-398.

Posner, M. I. and Keele, S. W. On the genesis of abstract ideas. Journal of Experimental Psychology, 1968, 77, 353-363.

_____. Retention of abstract ideas. Journal of Experimental Psychology, 1970, 83, 304-308.

Reed, S. K. Pattern recognition and categorization. Cognitive Psychology, 1972, 3, 383-407.

_____. Psychological processes in pattern recognition. New York: Academic Press, 1973.

Rosch, E. On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), Cognitive development and the acquisition of language. New York: Academic Press, 1973.

_____. Human categorization. In N. Warren (Ed.), Advances in cross-cultural psychology (Vol. 1). London: Academic Press, 1977.

Rosch, E. and Mervis, C. B. Family resemblances: Studies on the internal structure of categories. Cognitive Psychology, 1975, 7, 573-605.

Schoeffler, M. S. Theory for psychological learning. Journal of the Acoustical Society of America, 1965, 37, 1124-1133.

Shepard, R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function. I. Psychometrika, 1962a, 27, 125-140; II. Psychometrika, 1962b, 27, 219-246.

_____. Attention and the metric structure of the stimulus space. Journal of Mathematical Psychology, 1964, 1, 54-87.

Swets, J. A., Tanner, W. P., Jr., and Birdsall, T. G. Decision processes in perception. Psychological Review, 1961, 68, 301-340.

Tanner, T. A., Jr., Rauk, J. A., and Atkinson, R. C. Signal recognition as influenced by information feedback. Journal of Mathematical Psychology. 1970, 7, 259-274.

Tanner, W. P., Jr., and Swets, J. A. A decision making theory of visual detection. Psychological Review, 1954, 61, 401-409.

Thomas, E.A.C. On a class of additive learning models: error correcting and probability matching. Journal of Mathematical Psychology, 1973, 10, 241-264.

_____. Criterion adjustment and probability matching. Perception and Psychophysics, 1975, 18, 158-162.

Tversky, A. and Kahneman, D. Judgment under uncertainty: Heuristics and biases. Science, 1974, 185, 1124-1131.

Ward, L. M. Use of Markov -encoded sequential information in numerical signal detection. Perception and Psychophysics, 1973, 14, 337-342.