INFORMATION TO USERS

This dissertation 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.

**University Microfilms**
300 North Zeeb Road
Ann Arbor, Michigan 48106
A Xerox Education Company

72-22,331

SILER, William MacDowell, 1920-
SIGNAL DETECTION VERSUS THRESHOLD THEORY:
TESTS ON GOLDFISH.

The City University of New York, Ph.D., 1972
Biology

University Microfilms, A XEROX Company, Ann Arbor, Michigan

# SIGNAL DETECTION VERSUS THRESHOLD THEORY: TESTS ON GOLDFISH

by

WILLIAM SILER

A dissertation submitted to the Graduate
Faculty in Biology in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy, the City University
of New York.

1972

This manuscript has been read and accepted for the Executive Committee in Biology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

April 12, 1972
Date

William N. Tavolga
Chairman of Examining Committee
Professor William N. Tavolga

April 15, 1972
Date

Louis G. Moriber
Executive Officer
Professor Louis G. Moriber

Max A. Woodbury
Prof. M. Woodbury

Duke University Medical Center
Institution

George Gourevitch
Prof. G. Gourevitch

Hunter College
Institution

James A. Organ
Prof. J.A. Organ

City College
Institution

Erich G. Heinemann
Prof. E. Heinemann

Brooklyn College
Institution

John J. Lee
Prof. J.J. Lee

City College
Institution

The City University of New York

## ABSTRACT

SIGNAL DETECTION VERSUS THRESHOLD THEORY: TESTS ON GOLDFISH.

By William Siler. Advisor: Professor William N. Tavolga.

Tests of signal detection and threshold theory were carried out on goldfish, by comparing observed Receiver Operating Characteristic Curves with those to be expected from the theories being tested. An avoidance-conditioning paradigm was employed, with sound as the conditioned stimulus and shock as the unconditioned stimulus, using a specially-constructed automatic control apparatus. Equal-variance signal-detection theory was found to be consistent with the data, with two modifications. First, random responses unrelated to signal or noise, as observed by Heinemann in pigeons, were sometimes present, substantially altering the predicted ROC curve; in addition, the probability of random response was itself (unlike that in pigeons) a random variable. Secondly, the criterion employed by the fish was also found to be a random variable. A method for correction for criterion variability, the "small-group" method, was devised, which also alters the shape of the ROC curves. Simple threshold theories were not consistent with the data. However, under the explicit and implicit restrictions of threshold methods, it is shown that they can produce replicable results, even though based on invalid threshold methodology.

## ACKNOWLEDGEMENTS

The consistent help of Professor Tavolga is most gratefully acknowledged, not only for suggesting the problem here attacked, but for his strict but always kindly criticism and suggestions throughout the course of my graduate work. Dr. Eric Heinemann, by suggesting what turned out to be the "Heinemann Modification", played a critical part in this work. Thanks are due to the U.S.P.H.S., whose Research Grant RR-00291 supported the computer and biomathematical aspects of the work. Thanks are due to my committee, Drs. E. E. Suckling, G. Gourevitch , John Lee, Eric Heinemann, and Max Woodbury, for reading and criticising the manuscript. Miss Barbara Auerbach is thanked for her help in training and maintaining the test subjects, and Dr. Hugo Freudenthal for his help in the laboratory space problem. Above all, the support of my wife, Dr. Phyllis Cahn, made this work possible: she supplied laboratory space and equipment, hot meals, hosts of references, read and criticised the manuscript, and was completely indispensible.

TABLE OF CONTENTS PAGE

TABLE OF CONTENTS (Cont'd.) PAGE

**LIST OF TABLES** **PAGE**

LIST OF ILLUSTRATIONS PAGE

LIST OF ILLUSTRATIONS (Cont'd.) PAGE

## A. INTRODUCTION

### I. Threshold theory.

The concept of the sensory threshold was central to the development of psychophysics and psychophysical methods, and remained so, unchallenged, for a century. In their brief review of threshold theory, Green and Swets (1966) trace the idea back to the ancient Greeks. In modern thought, Leibnitz (1714) laid the basis for threshold theory when he put forward the notions of "consciousness" and of the "unconscious", but failed to define the concept of threshold itself. The first clear exposition of the threshold concept is due to Herbart (1824), who applied the concept to ideas, rather than sensations. Herbart defined the "threshold of consciousness" as the "boundary which an idea appears to cross as it passes from the totally inhibited state into some (any) degree of actual ideation." He also introduced the notion that a given idea had to compete with others to cross the consciousness threshold, reminiscent of the "signal-to-noise ratio" idea put forward much later. By 1824, then, the threshold concept had been clearly defined, but not as yet applied to sensory systems and sensations.

In the course of his experiments on touch, Weber (1834) arrived at a measurement of sensibility, or acuteness of sensation, in the "just noticeable difference" (jnd) concept, and used this measurement to relate sensation to stimulus in what was later formalized as "Weber's Law".

In the jnd concept, the idea of a threshold is firmly (if implicitly) fixed: certain stimuli (in this case the difference between paired presentations) are divided into two classes, noticeable and not noticeable; and a certain stimulus value (the jnd) falls on the borderline between the two classes. This is, although Weber did not himself generalize this far, a threshold. The full generalization was done by Fechner (1860); in his monumental Elemente der Psychophysik he created a whole new field of science, Psychophysics, based squarely on the idea of the sensory threshold, and laid the goundwork on which generations of scientists, interested in the investigation of sensory systems through behavioral techniques, and later in the study of learning, have based their researches.

Fechner's reasoning is summarized by Corso (1963) in his review of the threshold concept. Being mainly concerned with establishing the functional relation between mind and body, and realizing that mental magnitudes such as sensation were not directly accessible, he established a metric relationship between stimulus and response by searching for limens as measures of sensitivity. Distinguishing between absolute and differential sensitivity, Fechner laid down three specific methods for threshold determination: the methods of limits, of average error, and of constant stimuli. By the use of well-conceived methods, he was able to improve the reproducibility of tests such as those of Weber; and although new methods have been added in the hundred years since his techniques were published, Fechner's methods remain at the center of psychophysical research.

While Fechner established firmly the threshold concept as a primary tool in psychophysical research, the concept by no means led inescapably to a definition of threshold which could be experimentally implemented. Intrigued by the elegance of the Weber-Fechner law, Fechner defined the threshold in terms of that law:

$$\text{Sensation} = \text{Constant}\ (\log(\text{Stimulus}) - \log(\text{Threshold})).$$

Stevens (1961a), using the power law proposed by the 19th Century Belgian physicist Plateau to replace the Weber-Fechner logarithmic function, gave a definition parallel to that of Fechner but different from it:

$$\text{Sensation} = \text{Constant}\ (\text{Stimulus} - \text{Threshold})^n$$

in which $n$ is an experimentally-determined parameter for a given perceptual continuum. In both Fechner's and Stevens' formulations sensation vanishes when the stimulus equals the threshold; but experimentally the threshold would have to be established by extrapolation of data, and the precise threshold value would depend upon the (different) scaling laws postulated. Also, in neither formula is "sensation" defined.

A further problem arose in threshold determination which led to a third definition. Experimentally, it was very quickly found out that, with stimuli in the neighborhood of the threshold, there were statistical fluctuations in the stimulus-response relationship; a given stimulus would sometimes arouse a response, and sometimes not. Fechner devised elaborate statistical procedures to handle this effect. Nevertheless, both Fechner's and Stevens' formulations have three disadvantages: "sensation" is not defined; extrapolation is much more error-susceptible than interpolation when dealing with experimental data; and the threshold depends upon acceptance of a particular scaling law. To avoid these drawbacks, another definition came into use, and was formalized by Guilford (1954); the threshold was defined as "that low stimulus quantity that arouses a response 50 per cent of the time". (Such a definition holds for a simple detection task. However, in a two-alternative forced-choice task, a correct response is expected 50 per cent of the time by pure chance; in this case, a 75 per cent correct response rate is taken as the threshold point.) Since Guilford's definition leads directly to experiment, is unambiguous, and does not require acceptance of a particular scaling law, besides being consonant with the statistical variations observed, it has been widely accepted, in spite of its *ad hoc* nature.

The problem of responses in the absence of stimuli remained. To account for these, a number of additional threshold models have been developed: the high-threshold, low-threshold, and neural-quantum models. These are reviewed by Green and Swets (1966), and will be briefly discussed here.

The "high-threshold" model of Blackwell (1953) is of particular interest for several reasons. It is a one-parameter model, making unambiguous testing of model validity against data fairly easy; it is most directly related to Guilford's commonly accepted threshold definition; the well-known "correction for chance success" frequently applied to threshold data follows from this model; and a somewhat different formulation of the psychological assumptions, designed to be applicable to lower animals, yields the same model, as shown in the "Theoretical Results" section of this paper.

Of the other threshold models, the "two-state low-threshold" model of Luce (1963), having only two parameters, can also be tested against data for validity. The "single-threshold multi-state" theory of Swets, Tanner and Birdsall (1955, 1961) and the "two-threshold three-state" theory of Atkinson (1963) have a number of free parameters, making model validity tests difficult. The "neural-quantum" models, proposed by von Bekesy (1930), revised and extended by several others as reviewed by Corso (1956) and Stevens (1961b), depend upon the postulate of many thresholds, with a "neural-quantum" jump between them. As in the "multiple-threshold" and "multiple-state" models, the number of free parameters gives such flexibility in model predictions that testing for validity becomes very difficult.

The state of the art in animal psychophysics is presented in Stebbins (1970). In his review of this book, Guttman (1972) gave an idea of the importance in animal work of the threshold concept today: "The book is intended to be methodological ... but happily it is more besides. Compositely the papers afford a view not elsewhere available of the state of the art, together with some of the best products of the art, and offer an intimation of how far the art might be extended. _The detailed information supplied concerning the measurements of thresholds and similar parameters_ is what one would ordinarily have to obtain from the investigators in person...." (emphasis added).

Another example of the importance of the threshold concept in recent years is given by the Nobel Prize winner Georg von Bekesy (1960), in a book which summarizes 34 years of the author's researches. In his index, von Bekesy has almost 60 entries for thresholds or limens, and the book includes (pp. 207-271) an entire chapter on auditory thresholds. von Bekesy also gives here an indication of the reproducibility of threshold measurements. Presenting a design for a new audiometer, he tests _the instrument itself_ by seeing whether thresholds of hearing on a human subject taken by the instrument are reproducible (pp. 83-85). A more powerful indication of threshold reproducibility is shown by the fact that the well-accepted S.P.L. unit for sound pressure is based on the hearing threshold of humans at 1000 Hz.

Since the present work is concerned with tests of the threshold theory on fish, the reproducibility of hearing threshold measurements on fish is of interest. Several investigators have been concerned with the ability of the goldfish, Carassius Auratus, to perceive auditory signals of various frequencies. The curve of threshold intensities against the frequencies at which these thresholds were measured constitutes an audiogram. Such audiograms have been measured by Jacobs and Tavolga (1967); Fay (1969); Weiss (1966) and Enger (1967). Similar paradigms (avoidance conditioning in a shuttlebox with shock as the U.S.) were used by Weiss and by Jacobs and Tavolga; Fay employed classical conditioning of respiration with shock as the U.S.; and Enger utilized positive reward of food. In spite of the widely different paradigms employed by Fay, Jacobs and Tavolga, and Enger, their audiograms were in substantial agreement; but the audiogram of Weiss, though taken with a paradigm very similar to that of Jacobs and Tavolga, was different. A careful set of experiments by Wodinsky (unpublished MS), whose criticism is supported by the work of Yarczower and Switalski (1969), indicated that Weiss' procedure was suspect. In actual testing, Weiss omitted shock, to avoid "experimental neurosis"; but Wodinsky's data showed that if shock were omitted extinction would very quickly set in, causing Weiss' threshold to be artifactually high. In fact, Weiss' threshols did tend to be higher than the other investigators found. In this case, a difference is reported thresholds led to tests of the test method itself, the difference in thresholds being attributed after careful check to an inappropriate test procedure.

An extreme example of threshold reproducibility occurred in two separate tests of low-frequency sonic sensitivity of the grunt Haemulon sciurus, a reef fish, carried out first by Tavolga and Wodinsky (1963) and subsequently by Cahn, Siler and Wodinsky (1969). Similar paradigms, but quite different test appatus, were employed. At very low frequencies, Tavolga and Wodinsky observed two distinct thresholds for this animal; with little experimental evidence but brilliant insight, they predicted that one of the two observed thresholds was due to sound pressure sensitivity by the ear, and the other to particle motion sensitivity by the lateral line. With more advanced acoustical test facilities, Cahn et al not only replicated the dual thresholds, but were able to demonstrate that the second one was indeed due to particle motion sensitivity.

For many years, then, the threshold concept has dominated behavioral tests of sensory system sensitivity, and served as the cornerstone of the science of Psychophysics. The intuitive appeal of the idea was buttressed by the reproducibility of the thresholds and limens obtained.

## II. Signal Detection Theory.

While psychophysicists were grappling with the theoretical and experimental problems of detection of stimuli near the border of detectibility with the aim of investigating sensory system sensitivity and sensory psychology, theoretical advances made in other fields were utilized by a group of scientists primarily concerned with detection problems arising from military electronic equipment developed during World War II, especially radar and sonar. The military detection problem was of more than academic importance: to put it crudely, one could get killed if one failed to detect an enemy, or a great deal of wasted effort could be spent if one detected an enemy who wasn't there. The problem of noise combined with a possible signal, heretofore largely ignored by biologists, came forcefully to the fore.

From the problem of reliable communication over telephone lines and their maximally efficient use arose the information-theoretic approach of Shannon and Weaver (1949). Weiner (1949) combined statistical correlational techniques with the time-varying voltage phenomena (familiar to electrical engineers) to enormous effect; in fact, signal detection theory is an extension of Weiner's approach of combining statistical techniques with signal analysis. Thurstone (1927) had developed the idea that a stimulus produced a Gaussian probability distribution of effects. By the late forties, then, a solid mathematical basis for handling the description of mixed signal and noise had been laid down. The critical element to be added was "...decision made by individuals under conditions of risk and uncertainty" (Green and Swets, 1966, p. 7). This theory was developed by Wald (1950) and by Morgenstern and von Neumann (1953); with the addition of statistical decision theory to signal analysis, signal detection theory came into being. The earliest signal detection papers are unavailable, being classified; among the earliest non-classified papers are those of Tanner and Swets (1953), and Swets, Tanner and Birdsall (1955), both Technical Reports of the Electronic Defense Group at the University of Michigan. In the regular literature, there is the early paper of Tanner and Swets (1954). Since then, the papers on signal detection as applied to radar and sonar constitute a major literature; since this paper is concerned primarily with tests on lower animals, no attempt is made to survey that literature.

Unlike threshold theory, <u>uncertainty on the part of the observer is central to signal detection theory</u>. Threshold theory assumes, in its simplest form, that if the observer senses the signal (stimulus), he will respond; signal detection theory makes no such assumption. Threshold theory is not concerned with noise and its properties; signal detection theory is based on analysis of the noise "stimulus" which competes with the signal for recognition by the observer.

Consider the 50% detectibility point normally considered to be a "threshold"; the threshold theorist might say that half the time the subject hears the signal, and half the time he does not; the signal detection theorist would say that half the time the subject decides that the incoming stream of sensory information meets his criterion for responding "Signal", and half the time fails to meet this criterion. The difference between these two concepts is by no means purely semantic; qualitatively different predictions for the performance of subjects in detection tasks result from accepting one or the other idea.

Given the idea of observer uncertainty, there are then four possibilities: The observer may respond "No Signal" when there is none, a *correct rejection*; he may respond "Signal" when there is none, a *false alarm*; he may respond "No Signal" when signal is present, a *miss*; or he may respond "Signal" when signal is present, *a hit*. These possibilities may be concisely summarized without loss of information by two probabilities, the "Hit probability", written as the conditional probability of responding "Signal" (S) under the condition that signal is present (s), or P(S/s); and the "False Alarm probability", the probability of responding "Signal" (S) under the condition of noise alone (n), or P(S/n). A plot of P(S/s) against the corresponding P(S/n), called for unfortunate historical reasons a Receiver Operating Characteristic or (more concisely) an ROC curve, is basic to the interpretation of experimental results of detection tasks for signal detection theory. The shape of experimentally-measured ROC curves is of the greatest importance in deciding whether threshold theory or signal detection theory holds for a particular experiment and subject, since the shape of the curves predicted by the two theories (with minor exceptions later discussed) are quite different; in this paper, the touchstone for testing the validity of a given theory has been the shape of the experimentally-measured ROC curves.

As an exposition of signal detection and threshold theories, replete with references, Green and Swets (1966) is an excellent text. For the convenience of the reader not familiar with signal detection theory, a very brief presentation of the basic ideas will be given here.

The observer is confronted with a continuous stream of incoming sensory information, which may be signal plus noise or noise alone. From moment to moment, the observer infers from this information the likelihood that it is signal (say $l_s$) or simply noise (say $l_n$). The likelihood ratio ($L = l_s/l_n$) is then a measure of the degree of assurance the observer has that signal, rather than simply noise, is being received.

Having determined from the incoming sensory information a likelihood ratio, the observer must now decide whether to respond signal (S) or noise (N). It is assumed in signal detection theory that the observer adopts a certain criterion for responding Signal; if the likelihood ratio at any moment exceeds the criterion, the response is S; if the likelihood ration does not exceed the criterion, the response is N.

From von Neuman's theory of games, and in particular the theory of statistical decisions (Wald, 1950), the choice of an optimum criterion (i.e. likelihood ratio of signal to noise) is based upon the expected rewards and punishments for correct and incorrect decisions; that is, the <u>payoff</u> <u>matrix</u>. Typical payoff matrices are shown in Figure 1. If a heavy penalty is attached to a false-alarm, and a light reward to a hit, the subject is likely to choose a very stiff criterion, i.e. a very high value of the likelihood ratio for responding S; conversely, a very loose criterion may be employed, so that the response is S for very low values of the likelihood ratio. In the extreme limit of the loosest possible criterion, the subject always responds S, giving hit probability P(S/s) and false-alarm probability P(S/n) both equal to 1; with the other extreme of the tightest possible criterion, both P(S/s) and P(S/n) are zero.

### CONVENTIONAL THRESHOLD PAYOFF MATRIX

RESPONSE

| Condition | | S | N |
|---|---|---|---|
| | s | 0 | 1 |
| | n | 0 | 0 |

(Entries are shock probabilities)
Shock on MISS only.

### PAYOFF MATRIX 2

RESPONSE

| Condition | | S | N |
|---|---|---|---|
| | s | 0 | 1 |
| | n | 1 | 0 |

(Entries are shock probabilities)
Shock on MISS or FALSE ALARM.

### PAYOFF MATRIX 3

RESPONSE

| Condition | | | |
|---|---|---|---|
| | s | v | 1 |
| | n | v | 0 |

v = Difficulty of hurdle crossing as compared to shock.

Shock on MISS only; Hurdle cross on HIT or FALSE ALARM.

Figure 1

Typical ROC curves for the simplest and widely-accepted signal detection model, the "equal-variance "Gaussian model, and for the two simplest threshold models, the "high-threshold" theory (implicit in most threshold testing) are shown in Figure 2, and the somewhat more complex "two-state threshold" theory of Luce, (1963) is shown in Figure 3. The "high-threshold" model, proposed by Blackwell (1953), is of particular interest in threshold theory for lower animals since, although originally based on the assumption that inter-trial crossings (false alarms) were based on "guesses" on the part of the subject that signal was present which had only a random chance of being correct, it can also be derived on the assumption that false alarms are random events unrelated to the signal in any way. In fact, Cahn and Siler (unpublished) so derived the high threshold theory in the course of the experiments reported in Cahn, Siler and Wodinsky (1969), and tested the hypothesis of randomness of false alarms by checking the distribution of intertrial crossing intervals; this distribution was found to be approximately of the Poisson type, as expected from the assumption that false alarms occurred with equal probability in time. Luce's two-state threshold model is slightly more complex, in that the ROC curve consists not of one straight line segment, as in high-threshold theory, but of two; its appearance is still much closer to the high-threshold curve than to signal-detection curves, being easily distinguished from the latter.

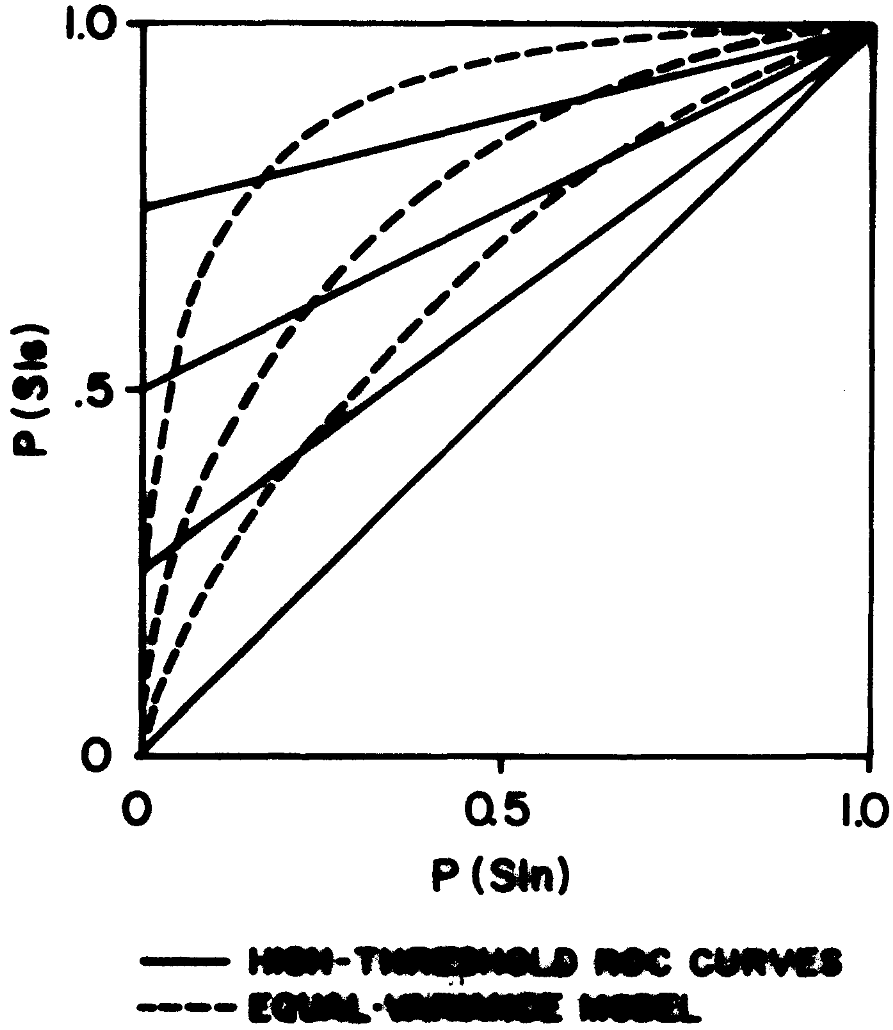
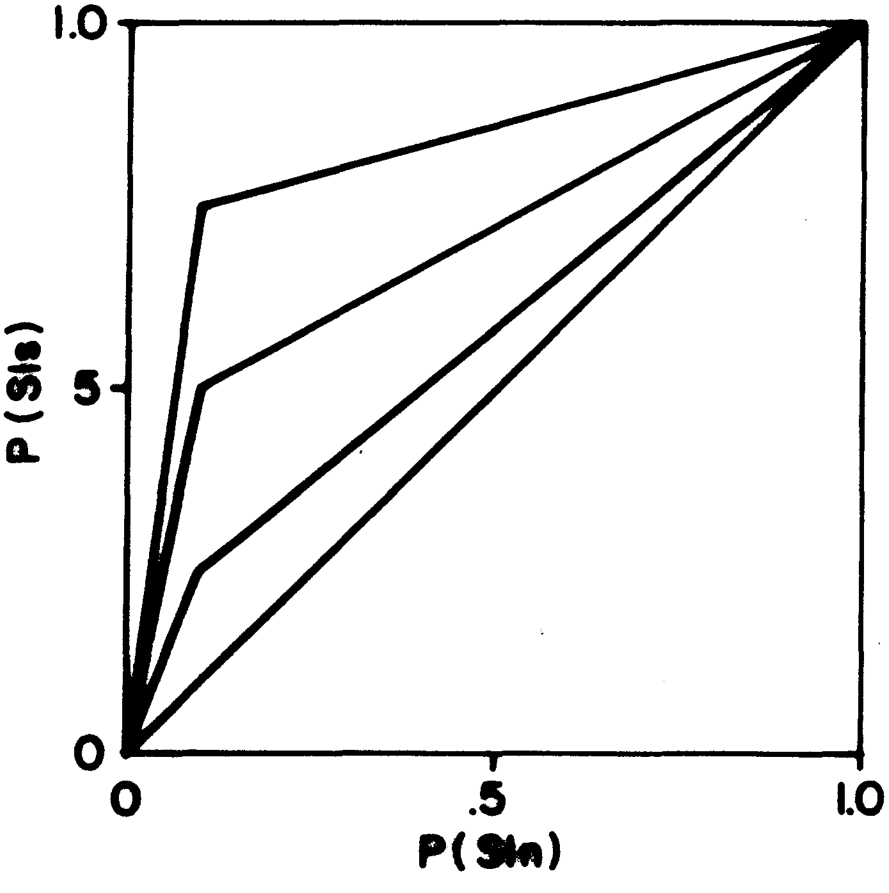

1.0
.5
0
P (S|s)
0
0.5
1.0
P (S|n)
HIGH-THRESHOLD ROC CURVES
EQUAL-VARIANCE MODEL

Figure 2



TWO-STATE THRESHOLD
ROC CURVES

Figure 3

A special characteristic of equal-variance signal detection ROC curves is that they plot as straight lines on double probability paper with a slope of 1. A modification of the equal-variance model which has received some acceptance, the "unequal-variance" model, also yields straight lines on double probability paper, but with slopes somewhat less than unity.

Signal detection theory than yields much more complex problems than simple threshold theory. Experimentally, hit probabilities alone do not suffice, but false alarm probabilities must be recorded as well; rather complex curves may be plotted, requiring a wealth of data; and a detectibility index must be found which is (hopefully) independent of the subject's criterion for responding or not to the ambiguously detected signal.

Fortunately, the equal-variance signal detection model yields a very simple criterion-free detectibility index, called "d-prime". A d' value of zero amounts to zero detectibility; higher d' values indicate increased signal detectibility. To each d' value there corresponds a unique ROC curve, and (if one accepts the equal-variance model unreservedly) a d' value may be calculated from a single measurement of hit probability P(S/s) and false alarm probability P(S/n).

Experimental evidence for the validity of signal-detection theory for humans was very strong as far back as five years ago (Green and Swets, 1966). At that time, discussing the evidence for the two classes of theories, they state: "There may be a sensory threshold....On the other hand, the existence of a sensory threshold has not been demonstrated. Furthermore - and this is the moral of the story - the threshold that is tenable is not compatible with several commonly used forms of the classical psychophysical methods. A threshold likely to be exceeded in the absence of a signal is not measured by the yes-no forms of the methods of limits or constants without a large number of blank trials, nor by means of the correction for chance success." This, coupled with their rejection on experimental grounds of the high-threshold theories (in which the threshold is not likely to be exceeded in the absence of a signal) constitutes a devastating criticism of the type of threshold tests currently performed on lower animals. The mounting evidence for signal detection theory, and its apparent irreconcilability with simple threshold theory, led Swets (1961) to publish a strong attack on threshold theory in a paper entitled simply "Is there a sensory threshold?"

In discussing possible utility of low-threshold theories, Green and Swets (1966) noted that "Two of them give promise of illuminating learning processes in the detection setting, but in their present forms the low-threshold theories do not provide an adequate basis for experimentation on substantive problems in sensory psychology." This paper is concerned with sensory psychology rather than learning; but it is worth noting that current literature in learning theory relies heavily on signal detection theory, and that the promise of a contribution here of threshold theory has by-and-large not materialized; see for example the review by Banks (1970).

In summary, signal detection theory is based on the well-developed sciences of signal analysis and statistical decision theory, and is mathematically attractive. Further, the experimental evidence for its validity in humans is so great that it is accepted widely, and has achieved this acceptance in a very short period of time (approximately 15 years since the early paper of Tanner, Swets and Green, 1956; 10 years since Swets' (1961) challenge to threshold theory; and five years since the publication of the first comprehensive text of Green and Swets (1966).

## III. Tests of Signal Detection Theory on Lower Animals

In view of the extensive application of signal detection theory to humans with great success, the sharpness of Swets (1961) attack on threshold theory, and the importance of sound psychophysical methods for sensory system tests on animals with whom verbal communication is impossible, the literature on signal detection theory tests in lower animals seems unbelievably sparse.

The first animal tests of signal detection theory were carried out on rats in Skinner boxes by Hack (1963). While he was able to calculate detection indices and variance ratios using the unequal variance S-D model, there was a great deal of scatter in his data points, and no error analysis of the derived parameters was carried out; in consequence, his results must be viewed as equivocal at best. Similar tests by Nevin (1964) produced similar results; he achieved only a limited range of false-alarm probabilities, and while his ROC curve is suggestive of signal detection theory, the data are fitted nearly as well by high-threshold theory. Rilling and McDiarmid (1965), using pigeons, achieved two points on each of a number of ROC curves at different signal-to-noise ratios. Their results seem to support signal detection theory, but the limited number of points on each ROC curve again left some doubt. Based on these results, Green and Swets said in 1966; "Results to date in animal psychophysics are barely suggestive of possibilities."

By 1967, Blough (1967) had been able to generate a complete set of ROC curves for pigeons in a visual discrimination task. Although his method of data analysis is somewhat open to question, a rating method being employed which guarantees that the ROC curve will be monotonically increasing with false-alarm probability, the shape of the curves was clearly very much closer to signal detection theory than to threshold theory. Nevin (1965) published a beautifully subtle test of signal detection theory, re-analyzing the pigeon wave-length discrimination data of Boneau, Holland and Baker (1965); he used a first-order Markov chain approach to the effect of previous reinforcement on response probability based on signal detection theory, producing a theoretical curve in good agreement with the data. However, Nevin himself qualified his demonstration, saying "it is by no means clear that decision theory is applicable in detail to operant discrimination experiments of this sort". Similar reservations were voiced by Subowski and Spevack (1968), using signal detection analysis of discriminant conditioning in pigeons; their ROC curves were clearly not of the threshold theory type (except possibly for Luce's two-state threshold model), but were also not clearly of the standard signal detection theory type. Their paper ends with the rather interesting remarks: "The ambiguities in interpretation may make (signal detection theory) methods appear unsatisfactory. However, the ambiguities can be reduced by the generation of data, and the alternative may be considered: to simply ignore the problems which (signal detection theory) methods can help to solve."

Before 1970, then, no paper from substantial criticism had appeared which demonstrated unequivocally that one or the other of the signal detection models, rather than any low-rank threshold model, applied to any animal other than humans. The dissertation of Hobson (1970) appears free from any major criticism: families of ROC curves were generated for pigeons, in a perfectly straightforward fashion, which showed clearly that the standard

Gaussian signal detection model, with no more than mildly unequal variances between signal and noise distributions, applied to these animals. By 1970 this demonstration capped the mounting body of evidence that pigeons were basically signal detectors, and that threshold theory did not apply to them. However, the attempts to demonstrate that rats were signal detectors remained totally unconvincing, and no efforts to demonstrate that signal detection theory applies to other species than man, rat and pigeon have been published. Unpublished work by Tavolga on fish had indicated the strong possibility that these animals were signal detectors, but ambiguities in the data prevented their publication.

From the theoretical viewpoint, differences between the performance of pigeons and conventional signal detection theory had been noticed by Heinemann, Avin, Sullivan and Chase (1969), and a modification to the signal-detection-theory component of stimulus generalization curves was proposed with considerable impact on the shape of theoretical ROC curves. They introduced two auxiliary probabilities; the probability that the subject is not attending to the stimulus dimension, and the probability that the subject will give a response even though he is not attending to the stimulus. This latter probability is closely related to the probability of a random non-stimulus-related response hypothesized by Cahn and Siler (unpublished), in which they had rederived the high-threshold model. When combined with signal detection theory, a new formulation emerges, derived but unpublished by Heinemann, which we will call the Heinemann modification to signal detection theory, and which combines signal detection, "attention" and random activity. This modification will be discussed in detail later in this paper.

## IV. Objective of This Study.

The particular objective of this study was to determine whether goldfish (Carassius auratus) function according to threshold or signal detection theory. The goldfish was selected as test subject for several reasons. It is considerably lower in the phyletic scale than the pigeon, phyletic differences appearing morphologically in the amount of cerebral cortex present, and also in terms of behavioral responses (Bitterman, 1965; Behrend, Powers and Bitterman, 1970). Bitterman, testing learning patterns of fish, turtles, pigeons, rats and monkeys, found that there were phyletically-ranked differences; Behrend et al found fish showed "probability matching" on probabilistic reinforcement of two stimuli, but rats did not, optimizing their payoffs instead. Much threshold testing has been carried out on goldfish, particularly in acoustic sensitivity, so that the goldfish might be called the "white rat" of fish. Cahn, Siler and Fujiya (1972) have measured the goldfish threshold in terms acceptable both to threshold and to signal detection theorists, i.e. signal-to-noise ratio, as about 25 dB, giving a point of departure for the tests here reported. Finally, the goldfish is not unrepresentative of a very large group of lower animals on which sensory tests using threshold theory have been applied; tests of the validity of threshold theory and methods were felt to be of significant importance both in the interpretation of data collected in the past and in the theory and methods to be properly employed in the future, especially so in the light of the success of signal detection theory on the only two species on which it had been adequately tested, and in view of Swets (1961) sharp challenge to the validity of the threshold concept itself.

## B. MATERIALS AND METHODS

### I. Subjects.

Common goldfish of 6 to 7 cm fork length were purchased from two local pet stores, the Freeport Aquarium and Macy's Department Store. The fish were held in two twenty-gallon glass aquaria, each divided into six partitions by pierced plastic tank dividers; during training, two fish (easily distinguishable) were held in each compartment, and during final testing, one fish was held in each compartment. Recirculated and filtered fresh water was used, one-half gallon in each aquarium being replaced by tap water daily. The aquaria were cleaned and vacuumed daily to remove algae and feces. Commercial "Tetramin" fish food was presented twice daily.

## II. Training procedures.

The training paradigm of Tavolga and Wodinsky (1963) was followed. Six fish were trained concurrently. For each fish, a shuttlebox with hurdle obstacle was placed in a ten-gallon aquarium, at the bottom of which was placed a University underwater swimming-pool speaker. Stainless steel shock electrodes covered each side of the shuttleboxes. White noise, band-limited from 300 to 800 Hz, was supplied constantly to all tanks; other stimulus functions (signal, light when used, and shock) were switched sequentially from tank to tank after each trial. For stimulus control, a small analog-hybrid laboratory computer (Siler and Lubowsky, 1969) was used.

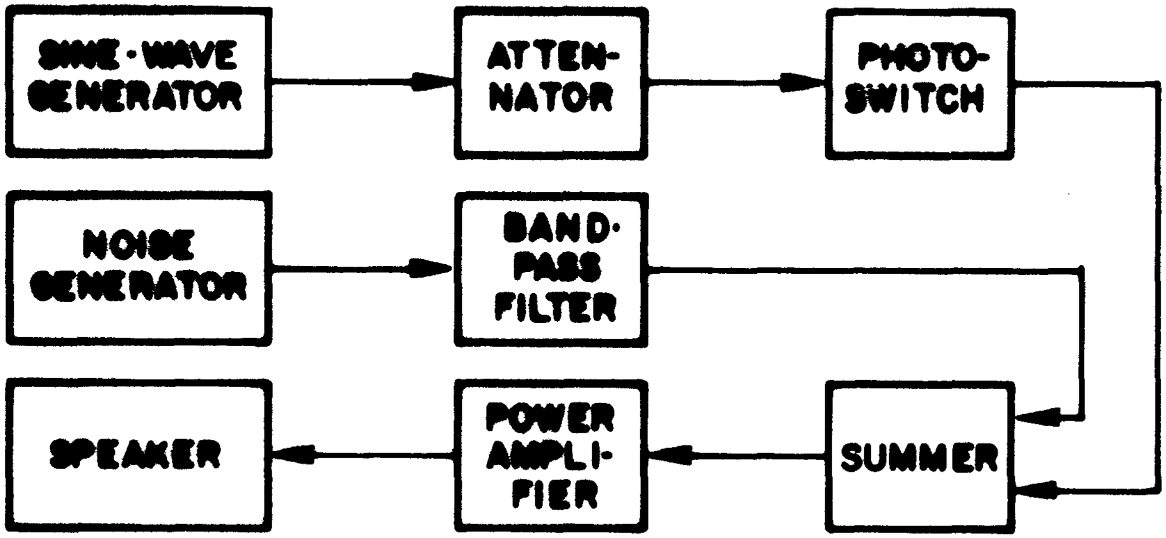
First in the trial sequence was the addition to the noise of a 500 Hz sine wave signal at a 40 dB signal-to-noise ratio; this figure was chosen as comfortably in excess of the 25 dB S/N threshold for goldfish at 400 Hz reported in Cahn, Siler and Fujiya (1972). In early stages of training, an auxiliary conditioned stimulus, an overhead Tensor light, was used; the light was eliminated as soon as avoidances were observed. Following initial presentation of the 500 Hz conditioned stimulus, a ten-second CS-US interval was allowed, during which the shock penalty could be avoided by swimming over the hurdle to the opposite side of the training tank. If the fish avoided, the conditioned stimulus was discontinued, and the trial terminated; if not, shock was administered as the unconditioned stimulus as 100 msec AC pulses with a two-second pulse repetition rate. If the fish failed to escape the shock by swimming over the hurdle and thus terminating all stimuli, after a few shock pulses the fish was guided over the hurdle by a plastic paddle.

During training, no particular effort was made to control false alarms (inter-trial crossings), except in those cases when the fish simply swam continuously from one end of the shuttlebox to the other; in those cases, if no reduction in crossing rate was noticed in a few days, the water level over the hurdle was lowered to give a "mild discouragement" to that particular behavior. False alarms were counted for a 30-second period during the midpoint of the training session. For the first five training sessions, six trials constituted a session; for the sixth through fifteenth sessions, ten trials per session were given; for sessions 16, 17 and 18, fifteen trials per session; and twenty trials per session thereafter. Sessions were conducted once daily, with some slight irregularity. After a fish was admitted to final testing, he was trained on those days in which he was not used for final test to avoid any extinction.

Sound generating equipment included a Heathkit sine wave generator for the 500 Hz signal; a General Radio noise generator for the masking noise; a Krohn-Hite filter, with 24 dB/octave rolloff, to band-limit the noise so that the noise pass-band was geometrically centered on the signal frequency; an Allen LR-1 phototransistor audio switch to control the signal with a 100 msec rise time, thus avoiding switching transients; and two Allen power amplifiers to drive the speakers, one being used to furnish signal and noise to the tank containing the fish under trial, the other furnishing noise alone to the other five tanks. Manual switching from one tank to another was used. The sound equipment is block-diagrammed in Figure 4; the stimulus sequence is block-diagrammed in Figure 5.

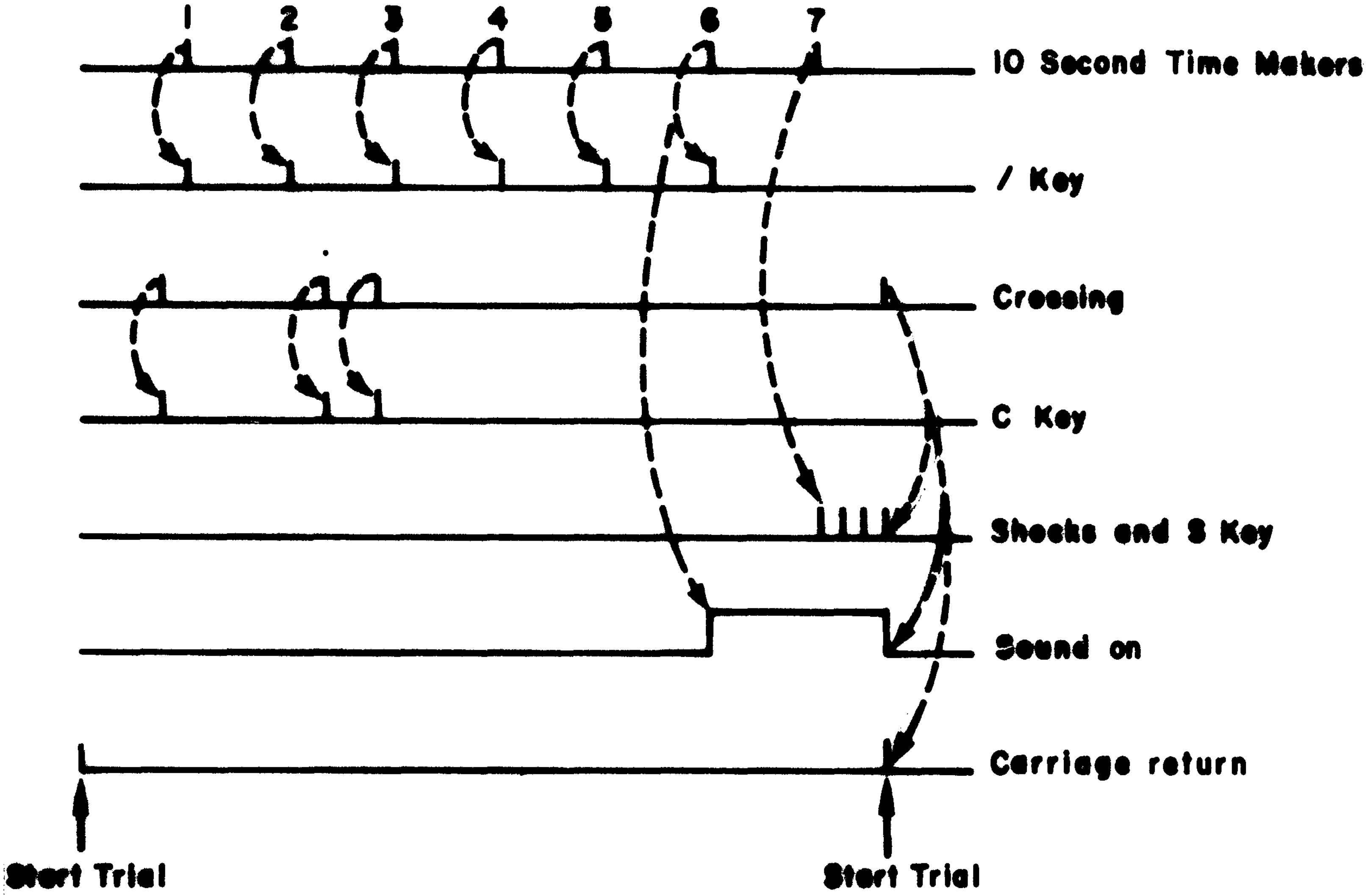
A fish was deemed ready for final testing in the acoustic enclosure later described when two criteria had been met: the avoidance rate had reached 90% or higher; and the probability of avoidance was consistently higher than the probability of a false alarm. In this sense, the training procedure was slightly different from that of Tavolga and Wodinsky (1963); in fact, they also required that the false alarm rate be quite low, of the order of one or two per minute. Since the objective here was not simply to obtain a conventional threshold but to trace out complete ROC curves, it was felt that the two criteria specified above were evidence that the subject had indeed learned the task. Fish that could not learn were eliminated.

In selecting fish for final testing, the experience of Nevin (1964) was kept in mind; in these tests of signal detection theory on rats, great difficulty was encountered in obtaining a sufficient range of false-alarm rates to



**BLOCK DIAGRAM:**
**SOUND PRODUCTION EQUIPMENT**

Figure 4



STIMULUS CONTROL DIAGRAM

Figure 5

constitute an adequate test of signal detection theory. For this reason, it was felt desirable to enter final testing with subjects that exhibited a variety of false-alarm rates, to maximize the possibility of obtaining complete ROC curves on at least some subjects. Fishes that were selected for final testing, numbers 2,3,6,8,11 and 12, were picked because they exhibited a high percentage of avoidances; a good ratio of avoidances to false-alarms; and a variety of false-alarm rates. Of these, fishes 8 and 11 were selected for testing their response to a payoff matrix defined in terms of shock probabilities, fish 8 having a low false-alarm rate, and fish 11 a high false-alarm rate; fishes 2,3,6 and 12 were selected for criterion-varying attempts by changing the difficulty of crossing the hurdle between the shuttlebox partition, their false-alarm rates ranging from low to high.

## III. Final testing apparatus.

a. Acoustical enclosure. The test enclosure is diagrammed in Figure 6. The basic housing was constructed of three expanded polystyrene picnic boxes. The lower, in normal upright position, held the shuttlebox, and was lined with one-half inch thick Johns-Mansville perforated acoustic tile. The second, inverted on top of the first with an adapter ring, contained a viewing window; a 45° inclined mirror; a 4-watt fluorescent lamp; and, mounted on its "top", a six-inch Jensen "high-fidelity" speaker (not acoustically suspended). The entire second layer was lined

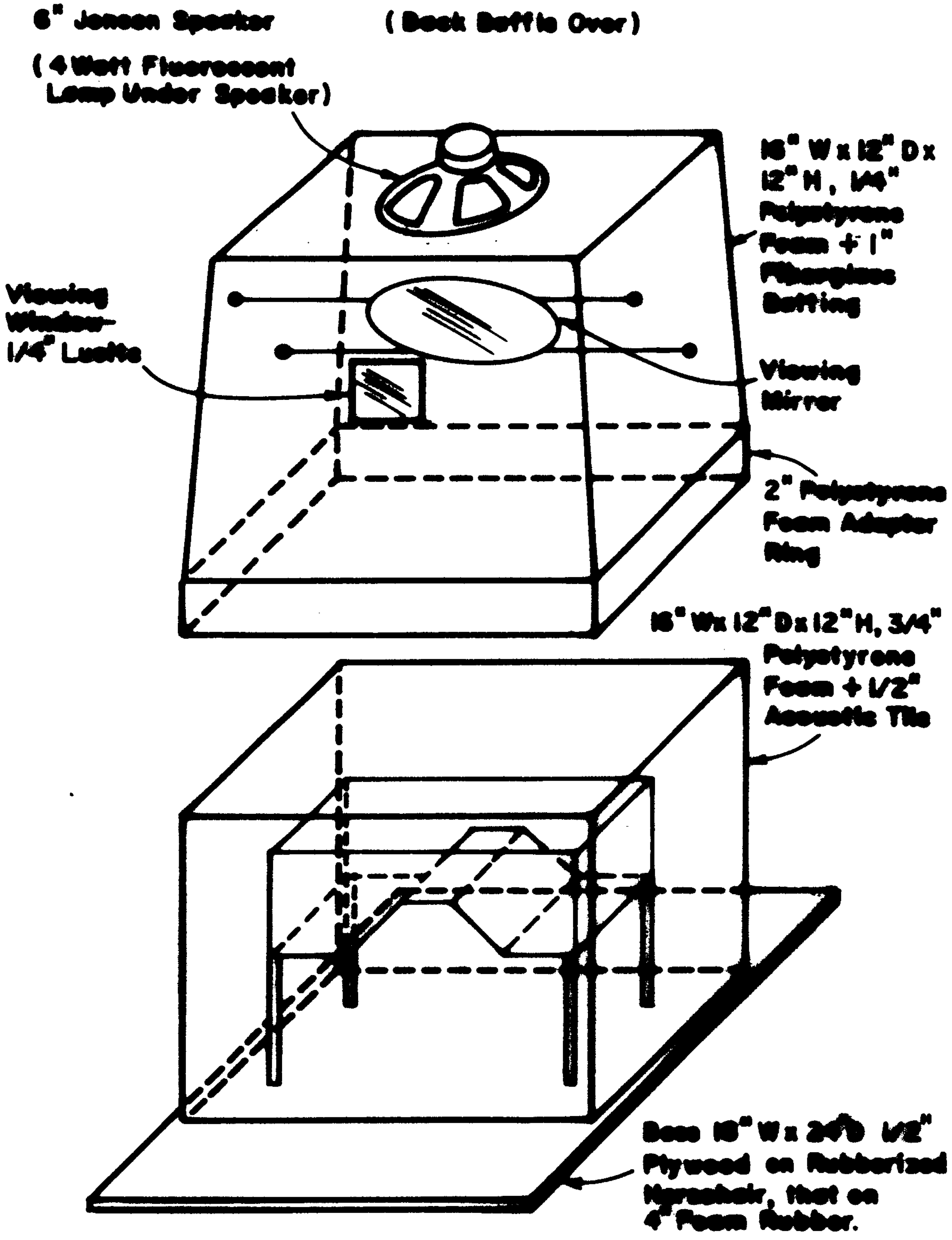


Figure 6 TESTING ENCLOSURE

with one-inch thick Fiberglas acoustic batting. The third layer, another polystyrene foam picnic basket inverted on second, was used as an "infinite baffle" for the speaker, to help match the impedance presented to the speaker by the bottom enclosure. It was, like the second, lined with one-inch Fiberglas acoustic batting.

The entire assembly was mounted on a square of 1/2" plywood, which served for mounting the Lafayette Radio photocell and light source crossing detector. The light source, mounted on a light plywood framework, projected its beam through a hole in the bottom picnic basket just above the shuttlebox hurdle; the photocell detector and relay received its light through a similarly-located hole in the bottom picnic basket directly opposite the light source.

Power spectra of the electrical signal furnished the speaker, and the actual acoustical spectra measured within the enclosure, are shown in Figure 7. It is obvious that the response of the enclosure is nonlinear, and that it is therefore unsuited toward such experiments as critical band measurement. However, the response is at least smooth, and since the noise was invariant throughout the entire experiment, the enclosure was considered adequate.

b. Electrical control and output. The volume of data which had to be taken precluded anything except a fully-automated final testing set-up. An additional laboratory computer (Siler and Lubowsky, 1969) was employed to im-

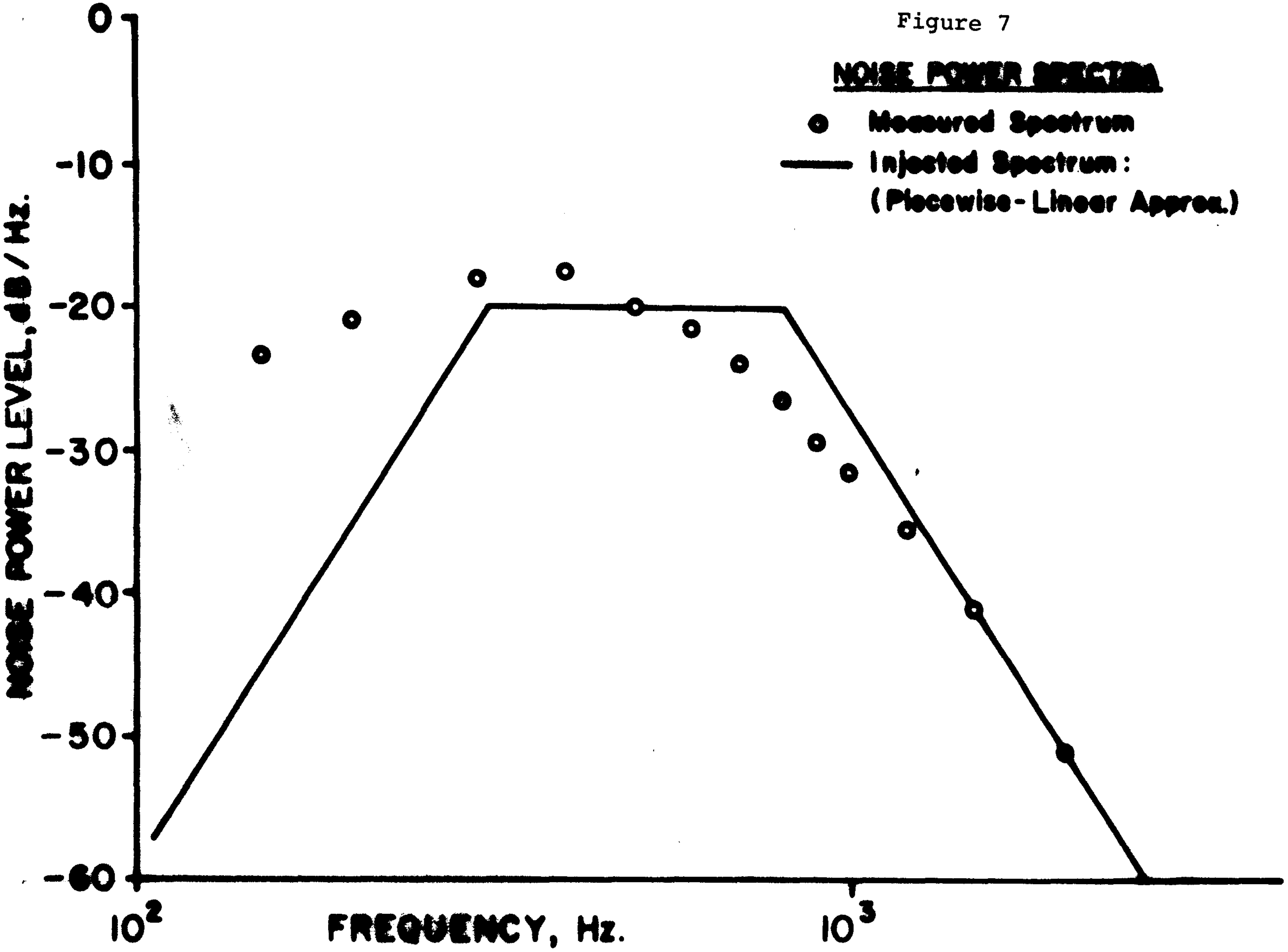


Figure 7

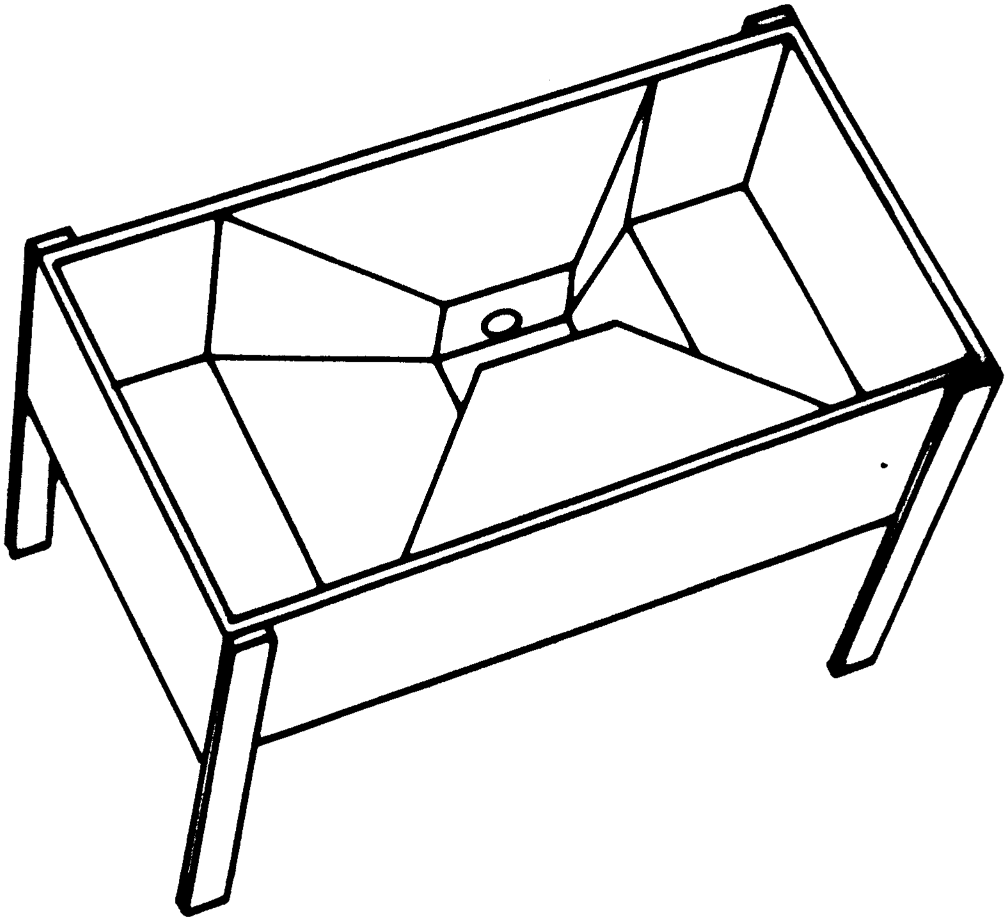
NOISE POWER SPECTRA

plement the control functions necessary in addition to those already implemented for training.

Following a suggestion by Tavolga, a second-hand electric Smith-Corona typewriter was purchased as the basic output device. Four push-solenoids drove the s,c,/ and carriage-return keys. A fixed schedule was set up (See Figure 5), as follows: at the initiation of a trial, the slash key was pushed every 10 seconds as a time marker. Crossings (false-alarms) during the succeeding 60 seconds were typed by the letter "c". After 60 seconds, the signal was presented; during signal presentation, the next crossing terminated signal, shock (if any), and started a new cycle. If shock were delivered due to failure to avoid, an "s" was typed. In this way, the experiment ran itself once started, and hard copy output on the typewriter was obtained. Data such as date, fish number, signal and noise intensity, were manually entered. Normal duration of a final session was 50 trials, although occasionally sessions were extended as long as 100 trials, or (rarely) terminated earlier due to intractable fish behavior. Data from the typewriter output were manually tabulated.

c. Tabulation of results. Since any single crossing in the 10-second CS-US period was considered a hit, any single (or multiple) crossing in a period between 10-second time markers with no signal presentation was considered a false-alarm. This means, of course, that two or more crossings during the programmed ten-second no-signal periods were considered a single false-alarm. For reasons discussed in the Results section, the first 20 seconds following a trial were ignored; since there was no evidence of anticipatory crossings, the following four 10-second periods were considered as legitimate false-alarm opportunities. For each 60-second trial period, then, there were four opportunities to false-alarm, and one opportunity to hit or miss. All data tabulation was carried out on this basis.

d. Criterion control methods. The only measurable payoff matrices that can be constructed in avoidance conditioning such as carried out here is to vary the probability of shock in the various squares of the payoff matrix. Such an attempt was made in the cases of fishes number 8 and 11. For the other fishes reported, numbers 2, 3, 6, and 12, the element in the payoff matrix varied was the difficulty of crossing the hurdle. Primary method used was a "gate", shown in Figure 8. Occasionally, other methods were employed, such as a platform to reduce the effective height of the hurdle, reducing light intensity sharply inside the test enclosure, or reducing water height over the hurdle. The obvious problem with such methods is that, not being dimensionally comparable with the shock penalty for a "miss", the payoff matrix does not permit computation of an optimum criterion for response according to signal detection theory. On the other hand, the fish may have its own ideas as to what constitutes an optimum



SHUTTLE BOX, WITH GATE IN PLACE

Figure 8

criterion or even the concepts upon which an "optimum" criterion should be based; it is quite possible, that a fish, never having met von Neumann, may be no better off than a hunch player at Las Vegas. In any event, the method employed to control the criterion, although non-mathematical, may be subject to mathematical analysis. See Discussion, section III.

IV. Sequence of Experimental Procedures. Following successful training, the fish were placed in the enclosure for final training and testing. Two fish (8 and 11) were selected for adaptation to a well-defined shock payoff matrix, with shock probability one for miss or false-alarm, and shock probability zero for hit or correct rejection. Four fish (2,3,6 and 12) were selected for varying the payoff matrix by varying the hurdle-crossing difficulty. For fishes 2,3,6 and 12, after acclimation to the enclosure, first step was to obtain a conventional threshold. Following this, and using a signal-to-noise level approximately equal to that obtained for a threshold, runs were started for obtaining ROC curves. Normally fifty trials were conducted under constant conditions without interruption, at which time the typewriter paper had to be changed. The hit and false-alarm probabilities were then checked, and any tendency for the false-alarm probability to vary spontaneously in one or the other direction was noted. Conditions for the next session were then determined in order to elicit different points on the ROC curve. For example, the hurdle gate might be removed or replaced, or the water level might be increased or decreased over the hurdle. The experimental design was then sequential rather than fixed, in order to trace out as complete as ROC curve as possible with nearly maximal efficiency.

By fixed experimental design, we mean the experimental design concept of R. A. Fisher, (1925) in which the experiments to be done are planned in advance of their execution, with a view toward treating the experimental results by analysis of variance, or a similar statistical technique.

However, as mathematical models have become more complex, and greater and greater volumes of data have had to be acquired for analysis, the concept of sequential experimental design has lately received much attention; see (for example) Woodbury (1970). In general, the concept of sequential experimental design involves the advance planning not of the experiments themselves, but of the way in which the next experiment will be determined from the results of preceding experiments, with the specific aim of optimizing the amount of experimentation required to achieve a given end. It is this concept which was employed here.

The sequential-design concept has been well-stated by Chernoff (1969): "Sequential analysis was originally introduced as an alternative to fixed sample size experimentation to take advantage of the possibility that the first few observations provide enough information for it to be unnecessary to gather more costly data. In cases when there are alternative experiments which may be carried out to obtain additional data, the relative goodness of these experiments may depend on facts which are unknown but on which the data has bearing. Thus as information accumulates, it may be used to select subsequent experiments which may be more efficient."

Modern sequential design usually entails complex statistical calculations between experiments to enable the quickest possible reduction of uncertainty, and maximally efficient experimentation to be carried out. Compared to such procedures as outlined in Woodbury (1970), the sequential procedures here utilized were very naive; but the same concept applies to all sequential designs, including that here utilized.

The basic procedure is described in Green and Swets (1966, p. 87); "The simplest and most direct test of the decision-theory analysis of the yes-no task is to trace out an empirical ROC curve for a signal (and noise background) of a given strength. To obtain this curve, the observer must change his decision criterion from one set of trials to another, thus producing several different points in the unit square having the proportion of hits, P(S/s), as the ordinate and the proportion of false-alarms, P(S/n), as the abscissa. The form of the empirical ROC curve is then compared with the form that is predicted by decision theory, and with the forms that are predicted by alternative theories of the decision process...The observer can be induced to change his decision criterion in any of several ways." The problem of sequential effects must now be considered. Green and Swets (1966, pp.259-265) discuss the problem of sequential decisions, but "under the assumption that d' remains constant over successive observations." If the observer's criterion varies during observations grouped for analysis, a reduction in the computed d' may be expected. This problem is also discussed by Green and Swets (1966, p. 411) who say; "We can assume...that sequential effects depress P(C) (or equivalently, d'). In the simplest case, if the observer shifts back and forth between two decision criteria...his ROC point will lie on a straight line connecting the two points that would correspond to the two criteria...Procedures for determing the size of the effect on P(C) of sequential tendencies - procedures which would permit a correction for them - have not been devised. Sequential effects are commonly ignored in sensory studies, for lack of an alternative, and with the belief that they are small." The lack of procedures for correction for sequential effects led to the development of the "small-group" model described under Theoretical Results in this paper, to permit correction for sequential effects given a varying criterion.

In Tables I through VI, in which the fish data are presented, the test conditions employed are also given, in the sequence in which they were employed.

## V. Data Analysis.

a. Grouping of raw data points. Since the subjects had four opportunities to false-alarm for every hit opportunity, false-alarm probabilities were known with considerably greater accuracy than hit probabilities for any given number of trials. Accordingly, for each fish, a certain number of consecutive trials were grouped together, and the hit and false-alarm probabilities determined simply by dividing the number of scores by the number of scoring opportunities, this being the maximum-likelihood estimate for a Bernoulli process. The number of trials in the initial groupings was always ten, such as trials 1-10, 21-30, etc. (Runs of ten such as 4-13 were never tabulated to avoid possible introduction of subjectivity.) After the initial grouping, further groups were formed by specific ranges of false-alarm probabilities. From these final groups of initial groups, final hit and false-alarm probabilities were estimated in the same way as for the initial groups. In addition, variance estimates of these probabilities (again by maximum-likelihood methods) were made by the formula, in which p is the maximum-likelihood probability estimate, $\sigma_{\hat{p}}^2 = p(1-p)/N$, N being the number of scoring opportunities.

This method of data grouping raises some previously unexplored theoretical questions. In particular, one may ask whether or not the shape of the ROC curve is altered by such a grouping method, and if so, how? The theoretical analysis of this question presented later in this paper indicated that the conventional equal-variance ROC curves are indeed altered if the false-alarm probability is a random variable, slightly so for initial groups of ten trials (40 false-alarm chances) and markedly for trial-by-trial groupings (4 false-alarm chances.) To test this theory, trial-by trial groupings were made for one fish, number 2.

b. Testing of models. Models considered for testing were the high-threshold model, the most satisfactory for justification of the staircase or up-down threshold-testing method of Tavolga and Wodinsky (1963); the Luce two-state threshold model; the equal - and unequal-variance Gaussian signal-detection models; the Heinemann modification of the Gaussian signal detection model, mentioned in the introduction; and finally, the "small-group" modification to the Gaussian equal-variance model mentioned in the paragraph above, and derived later in this paper. Testing of models was very simple: is it possible to vary the model parameters in such a way as to fit the data, or is it not? Fortunately, for the models tested, the answers to this question, as discussed in later sections of this paper, were unequivocal.

Two important qualifications must be placed upon these "unequivocal" answers. First, no other models than those listed above were considered. There is an unsolved problem, the model uniqueness problem, which has been concisely stated by Bellman (personal communication): "One cannot validate models. One can, however, invalidate them." Acceptance of a model here implies that all other models have been invalidated of those models tested; it does not imply that the accepted model is proved to be true; it does imply that the accepted model is consistent with all known data. Other untested (and perhaps as yet unformulated) models may apply as well, or better. Rejection of a model implies much more confidence; it a model is clearly inconsistent with the data, it is rejected as unsatisfactory.

The second problem is the parameter uniqueness problem, which is much more tractable, although still a subject of controversy for non-linear models. The question here is - - given a model which fits the data with certain specific values of its parameter values, how wide is the range of parameter values with which an acceptable fit could still be obtained? For our purposes, this problem is less plaguing than in some other problems, since the primary concern here is whether a model will fit the data at all, rather than with the specific parameter values obtained. In addition, the number of parameters used is rather small (one or two), which greatly eases the parameter uniqueness problem, and it is not here considered serious. Standard errors of parameters, when given, were calculated by linear model theory.

Fitting of models to data was carried out by digital computer, using a multi-dimensional Newton-Raphson algorithm to reduce the sum of squares of differences between observed and model-predicted hit probabilities to a minimum. For this purpose, the MODFIT program available at Downstate Medical Center, Brooklyn, N. Y. was used on the IBM 360/44 digital computer of Downstate's Medical Computer Science Program.

## C. RESULTS

For convenience, results obtained in the course of this work are presented in two major sections: first, theoretical results; second, experimental results. The theoretical section is given first, since the analysis of the experimental results and the method of their presentation depends upon certain of the theoretical results. Some experimental results are given in the "theoretical results" section, as justification for the theory there developed.

### I. Theoretical results.

This section will present the derivation of some of the models tested whose derivation has been previously unpublished, to lay the groundwork for the model-fitting and data-grouping methods employed in the "experimental results" section.

a. High-threshold model. This model has been previously derived by Blackwell (1953), under the assumption that the observer paid continuous attention to the detection task, and that false alarms were "random guesses" that the signal was present when, in fact, it was not. Cahn and Siler (unpublished) independently derived a high-threshold theory mathematically identical to Blackwell's, but under assumptions psychologically different.

Suppose a subject has a fixed probability P(D) of detecting a signal when present, and that every time he detects the signal, he responds S. Suppose also that the subject, for unknown reasons completely unrelated to the presence or absence of signal, has a probability P(r) of exhibiting the behavior interpreted by the observer as response S, this random response having a Bernoulli distribution. The probability of a false-alarm is then simply

$$P(S/n) = P(r)$$

and the hit probability is given by

$$P(S/s) = P(D) + P(r) - P(D)P(r)$$

or

$$P(S/s) = P(D) + (1-P(D))P(r)$$

These relations give ROC curves which intersect the P(S/s) axis at P(D), with P(r) = P(S/n) = 0, and climb linearly upward toward the point P(S/s) = P(S/n) = P(r) = 1, as the random activity P(r) increases. The theory is then mathematically identical to Blackwell's (1953). Typical "high-threshold" ROC curves are shown in Figure 2.

The particular reason for this derivation was the subjective impression by the author and by biologists experienced in threshold testing of fish that there appeared to be a behavioral difference between hits and false-alarms: hits appeared to take place "purposefully", with a short latency between the apparent recognition of a signal and the actual hurdle crossing, but false-alarms appeared to be much more leisurely events. In Blackwell's theory, both hits and false-alarms result from the subject responding to what he believes to be a signal, and there is no reason to anticipate behavioral differences between hits and false-alarms; but in the above, hits and false-alarms stem from entirely different mechanisms, and behavioral differences between them may be expected. It is worth noting that in signal detection theory, both hits and false-alarms spring from the subject's belief that the incoming stream of sensory information meets his criterion for signal, and therefore no behavioral difference between them should be observable; only the Heinemann modification, later discussed, relates responses due to apparent signal recognition under signal detection theory with random responses which could be behaviorally different.

b. The "Small-Group" method: alteration in theoretical ROC curves when false-alarm rate is a random variable. That the false-alarm rate may vary rapidly and spontaneously over a wide range with a constant payoff matrix is shown in Figure 9, in which in only 100 trials, grouped by consecutive 10-trial averaged sequences, a trained fish (number 3) traced out nearly a complete ROC curve quite spontaneously under invariant conditions. Variability assigned to random activity has been reported and analyzed for pigeons by Heinemann, Avin, Sullivan and Chase (1969). Similar random variation in false-alarm rate to that shown in Figure 9 was observed many times in the course of the experimental work later reported. I here propose a mathematical formulation for averages of a number of small trial sequences, grouped by observed false-alarm probabilities for each small sequence, and based on the Gaussian signal-detection model.

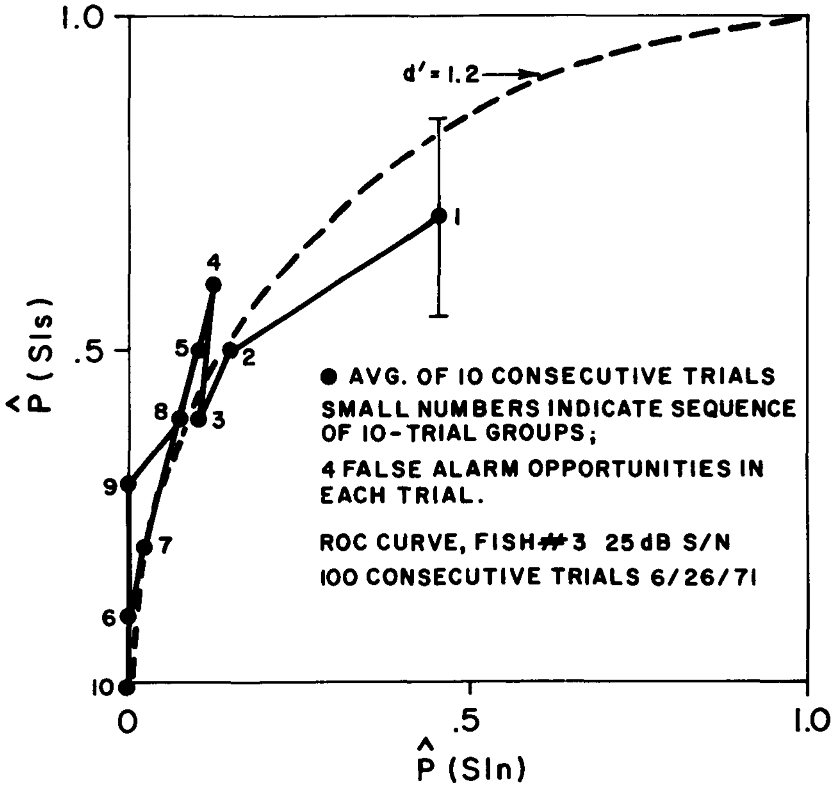

1.0
.5
0
.5
1.0
$\hat{P}$ (S|s)
$\hat{P}$ (S|n)
d′ = 1.2
1
2
3
4
5
6
7
8
9
10
● AVG. OF 10 CONSECUTIVE TRIALS
SMALL NUMBERS INDICATE SEQUENCE OF 10-TRIAL GROUPS;
4 FALSE ALARM OPPORTUNITIES IN EACH TRIAL.
ROC CURVE, FISH #3 25 dB S/N
100 CONSECUTIVE TRIALS 6/26/71

Figure 9

The complete derivation of the "Small-Group" model is given in the Appendix. Assumptions on which the derivation is based are:

- a large number of trials of a subject has been conducted in a simple detection task; hit and false-alarm probabilities are observed to vary widely.
- Trials are grouped in small sequences of a certain length, with M opportunities to false-alarm in each sequence.
- All sequences in which N false-alarms are observed are grouped together, and the hit and false-alarm probabilities for all sequences in each group are averaged.
- Actual hit and false-alarm probabilities are related by some model such that the hit probability P(S/s) is some function of the false-alarm probability P(S/n), i.e. if

  $P(S/s) = y$

  and $P(S/n) = x$

  then $y = y(x)$
- The *a priori* probability distribution for actual false-alarm probabilities is uniform.

Given these assumptions, we derive the expected value for the hit probabilities of each group in terms of the observed false-alarm probability for each group

$$z = P(S/n) = N/M$$

to be:

$$E[P(S/s)/P(S/n)] = \frac{\int y(x)\ P(z/x)\ dx}{\int P(z/x)\ dx}$$

in which y(x) is the model relating hit and false-alarm probabilities; z is the observed false-alarm probability for a particular group, with $z = P(S/n)$; and $P(z/x)$ is given by

$$P(z/x) = \binom{M}{N} x^N (1-x)^{M-N}$$

Integrations are carried out over the range of the "true" false-alarm probability x, i.e. from 0 to 1.

As shown in the Appendix, although the assumption of equally-probable false-alarm ratios is hard to justify, the "peakiness" of the P(z/x) function tends to make the calculations insensitive to this assumption, the insensitivity increasing with increasing sequence length.

Digital computer calculations show, for short sequences, considerable difference between the standard ROC curves and those with a limited number of trials selected for grouping, as shown by Figures 10 and 11. In particular, the expected hit probability for small sequences with zero false-alarm rate is not zero, but some larger value depending on the sequence length and the detectibility index d'; similarly, the expected hit probability for a false-alarm rate of unity is not one, but a slightly smaller value.

One should ask if these calculations are purely theoretical, or have some basis in fact, especially in view of the arbitrary prior distribution chosen for the Bayesian estimation procedure. Figure 12 shows an extreme case

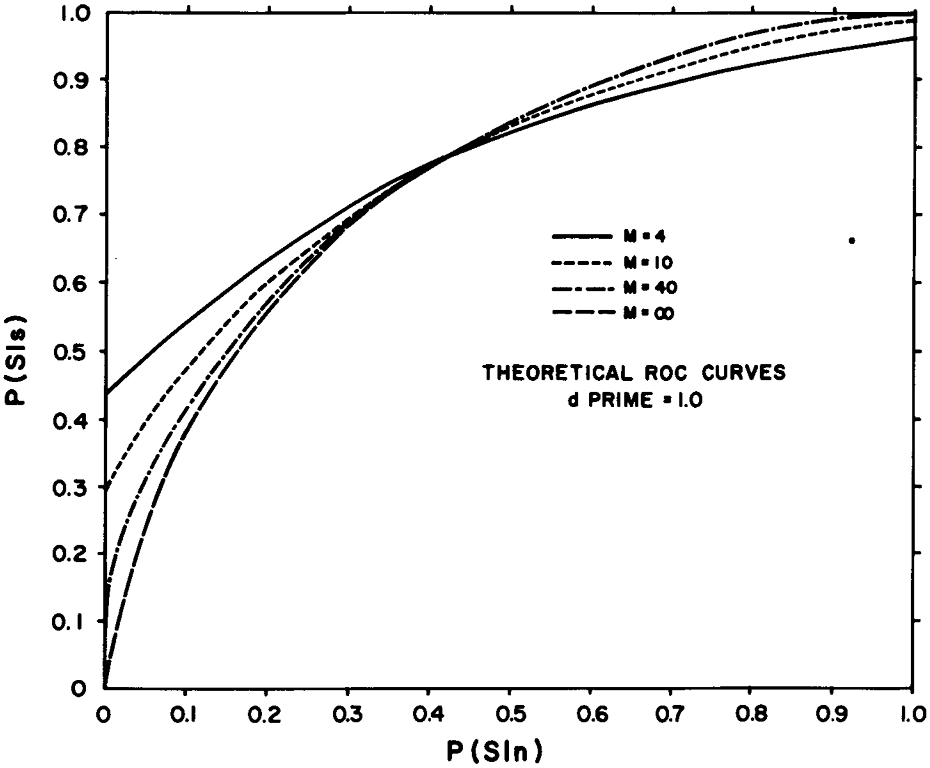

1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
P(S|s)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
P(S|n)
M = 4
M = 10
M = 40
M = ∞
THEORETICAL ROC CURVES
d PRIME = 1.0

Figure 10

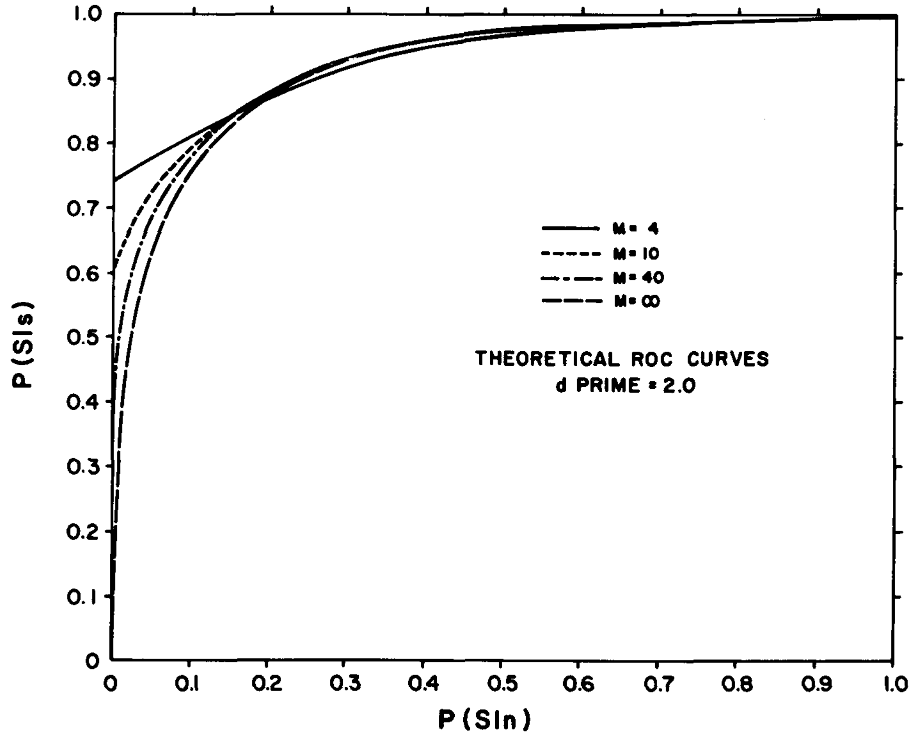

1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
P(S|s)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
P(S|n)
M = 4
M = 10
M = 40
M = ∞
THEORETICAL ROC CURVES
d PRIME = 2.0

Figure 11

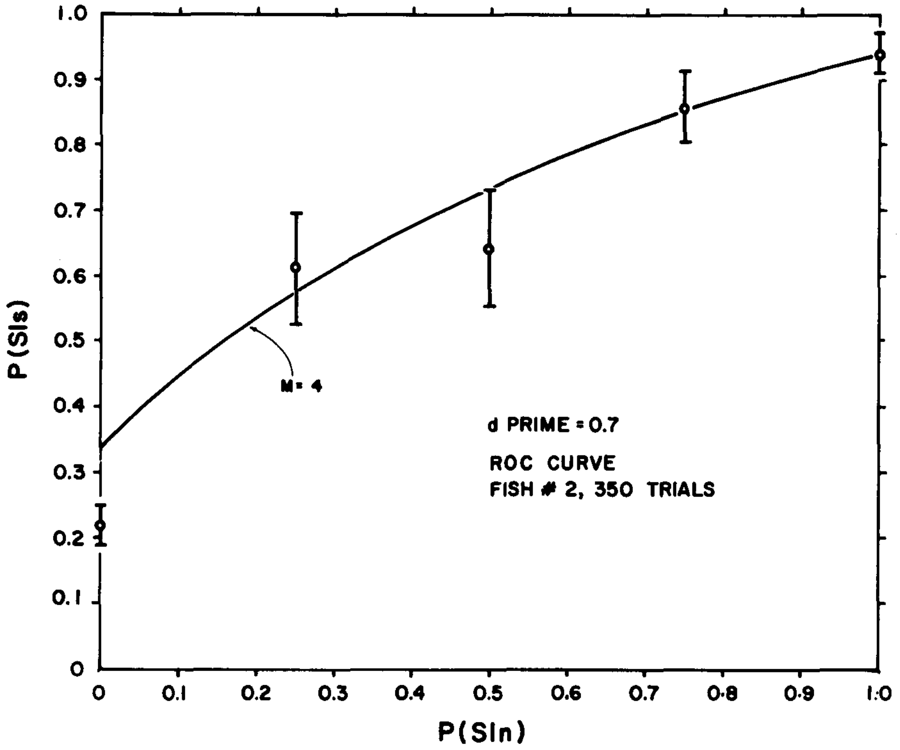

1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
P(SIs)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
P(SIn)
M=4
d PRIME = 0.7
ROC CURVE
FISH # 2, 350 TRIALS

Figure 12

of small grouping, in which the subject had only four opportunities to false-alarm per trial, and the results were grouped on a trial-by-trial basis, giving an M of only four. While agreement with the theoretical model is not perfect, indicating perhaps an oversimplified choice of the prior distribution, the qualitative agreement is encouraging; it indicates that the argument here presented has merit, and that the "grouping by false-alarm probability" method may serve to derive ROC curves when the subject's criteria varies spontaneously.

c. The "Heinemann Modification" to signal detection theory ROC curves. The fundamental concepts in this modification were formulated in Heinemann, Avin, Sullivan and Chase (1969). The derivation given below was the result of Heinemann's suggestion that the effect of random activity superimposed on signal detection theory be explored, and in results duplicates those previously derived but unpublished by Heinemann. The specific derivation given is my own,and may differ in detail from the original Heinemann derivation.

The Heinemann model corrects the conventional signal detection models for two effects; random responses similar to those described in the high-threshold model derivation given first in the "Results" section; and reduced response rates due to a lack of attention to the stimulus. First, the correction for random responses will be presented.

Say $S_1$ is a "Signal" response due to a decision by the subject that a signal is present, and $S_2$ is a random "Signal" response unrelated to the presence or absence of signal in the incoming sensory information stream. Define $P(S_2) = r$, the probability of a random response. The conditional probabilities $P(S_1/s)$ and $P(S_1/n)$ are defined by signal detection theory, as specified by the unequal-variance Gaussian model.

Then the joint probabilities are given by:

$$P(S_1 \vee S_2/s) = P(S_1/s)(1-r) + r$$

and

$$P(S_1 \vee S_2/n) = P(S_1/n)(1-r) + r$$

in which V denotes the logical "OR" operator.

For the Gaussian model,

$$P(S_1/s) = \Phi((k-m_s)/\sigma_s)$$

and

$$P(S_1/n) = \Phi((k-m_n)/\sigma\ n)$$

in which $\Phi$ is the cumulative Gaussian function, $k$ is the criterion, $m_s$ and $m_n$ are the Gaussian means for signal and noise respectively, and $\sigma_s$ and $\sigma_n$ are the standard deviations for signal and noise respectively (Green and Swets, 1966, p. 59).

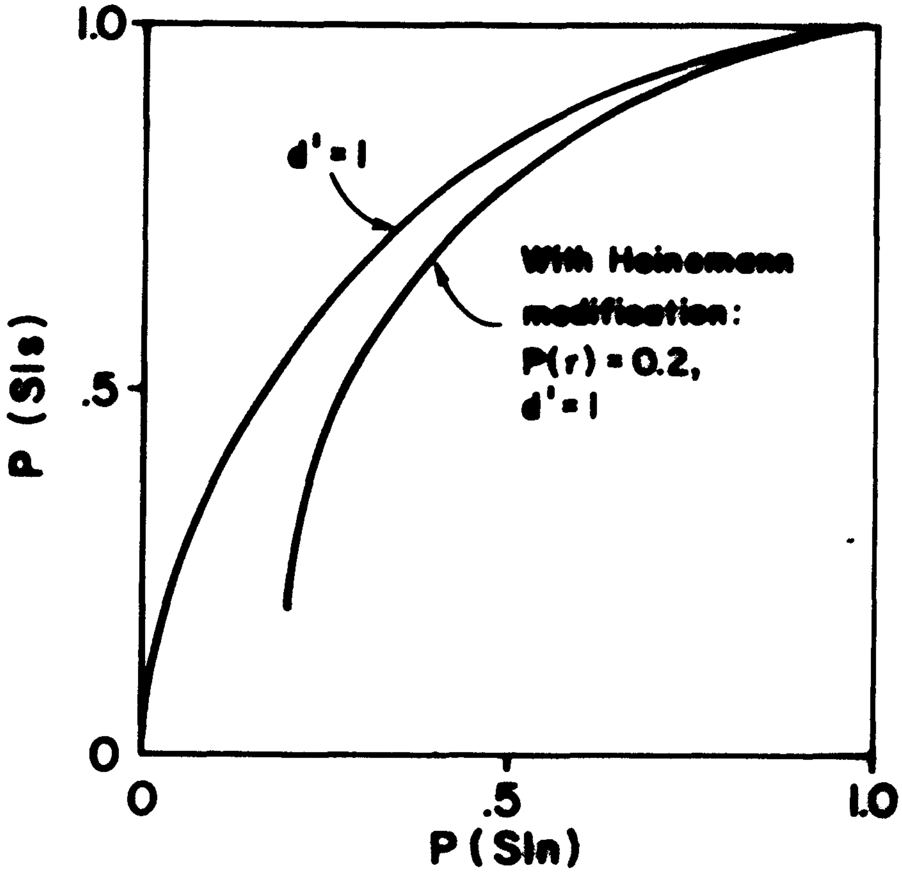
For combined random plus signal-detection-motivated activity, we achieve the final results:

$$P(S/s) = \Phi((k-m_s)/\sigma_s)(1-r) + r$$

and

$$P(S/n) = \Phi((k-m_n)/\sigma_n)(1-r) + r$$

A plot of a typical result is given in Figure 13, in which equal variance is assumed, with $d' = 1$, and $r = 0.2$ . It is immediately seen that a substantial modification of the conventional ROC curve has occurred.



**HEINEMANN MODIFICATION
TO
EQUAL-VARIANCE ROC CURVE**

Figure 13

Suppose now that the subject has probability P(A) of attending to the stimulus dimension, i.e. to the potential stimulus, in the absence of random activity. Denote by P(S/s) and P(S/n) the hit and false-alarm probabilities to be expected when attention is constantly paid to the stimulus, and P'(S/s) and P'(S/n) are the corresponding probabilities when P(A) is the probability of attention. Then the hit and false-alarm probabilities are altered by inattention to become:

$$P'(S/s) = P(A)P(S/s)$$

and $$P'(S/n) = P(A)P(S/n)$$

The resulting ROC curve starts at the point

$$P'(S/s) = P'(S/n) = 0$$

but instead of terminating at the (1,1) point as usual, curves upward to the right and terminates at the point

$$P'(S/s) = P'(S/n) = P(A)$$

Again, a substantial modification of the conventional ROC curve has occurred, but of quite a different nature.

Now consider the combined effects of inattention and of random activity. In this case, unlike Blackwell's model discussed first in the results section, somewhat different predictions result depending upon whether the random activity is assumed to result from random guesses as to the presence of signal, as Blackwell assumed, or whether it is assumed to have no relationship to the presence or absence of signal, as assumed in the Cahn and Siler derivation of the high-threshold model. In the first of these two possibilities, the random activity constituting random guesses about the signal, inattention would *reduce* the random activity probability to a maximum value equal to the attention probability P(A); in the second case, the random activity (being unrelated to the signal) would *not* be reduced by inattention, and the random response probability r could range to a maximum of 1, no matter what the attention probability might be. Unfortunately for model-testing purposes, outside of the restriction above on the maximum value of r, the models yield predictions which cannot be distinguished. If $P''(S/s)$ and $P''(S/n)$ are the response probabilities modified for both inattention and random responses, and P(S/s) and P(S/n) are the unmodified probabilities, then the "random guess" hypothesis yields:

$$P_1''(S/s) = (1-rP(A))P(A)\ P(S/s) - rP(A)$$

and

$$P_1''(S/n) = (1-rP(A))P(A)\ P(S/n) - rP(A)$$

and the assumption that random activity is unrelated to signal yields:

$$P_2''(S/s) = (1-r)P(A)P(S/s) - r$$

and

$$P_2''(S/n) = (1-r)P(A)P(S/n) - r$$

the corresponding terms rP(A) and r in these two pairs of expressions being indistinguishable experimentally except for the limitation on their maximum values.

d. *Summary of Theoretical Results*. Typical plots of the four models to be tested are shown in Figure 14. The quantitative and qualitative differences between the model predictions for the ROC curve are apparent. It now remains to see which model or models are adequate to explain the experimental results.

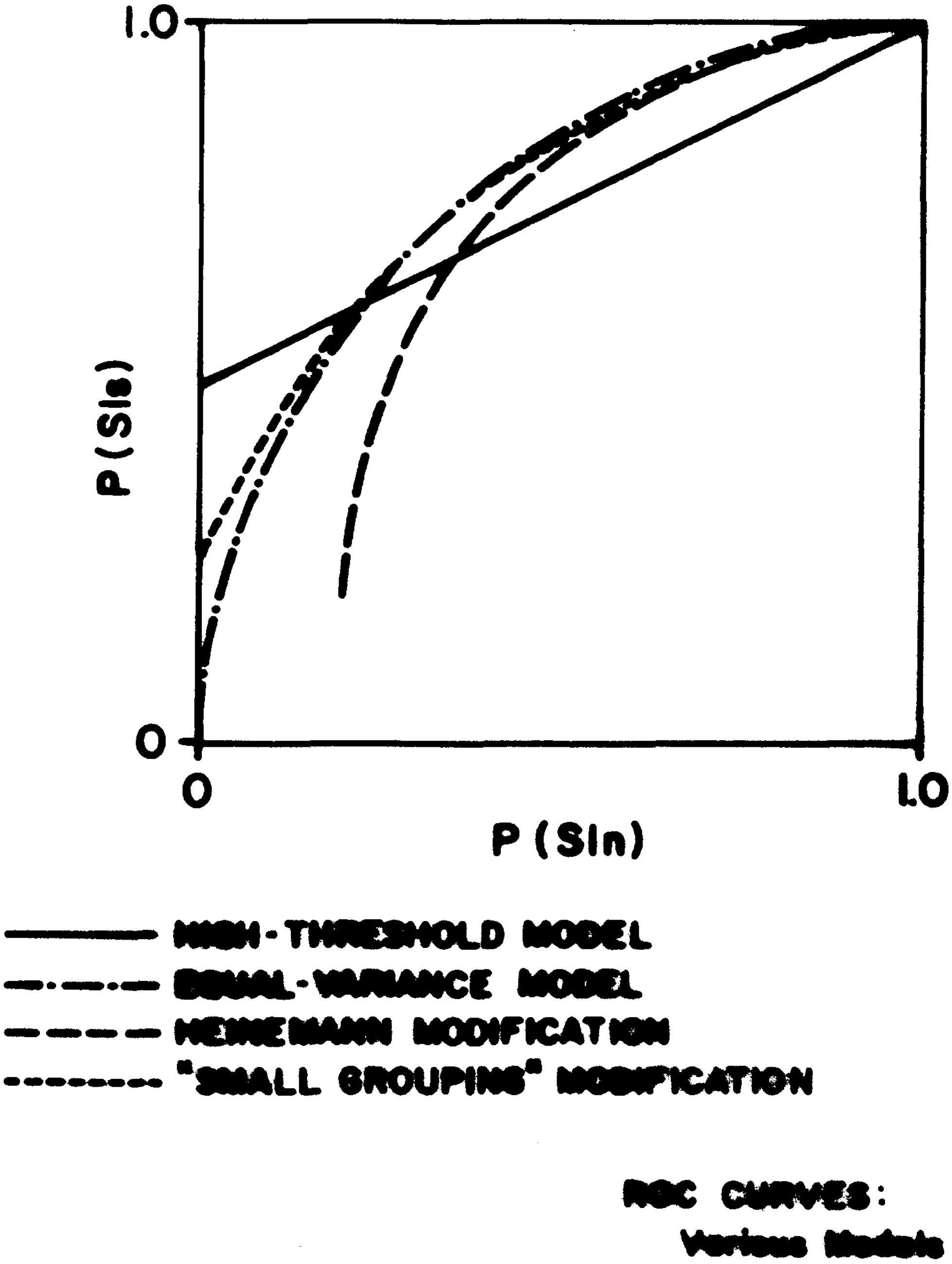
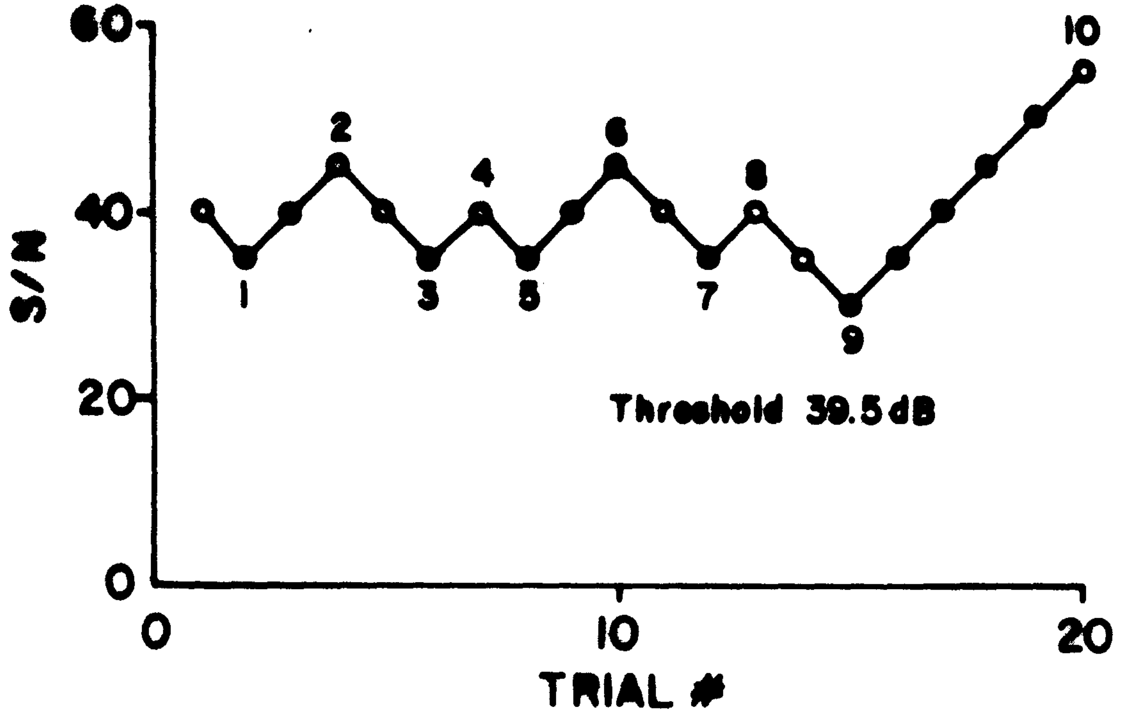

1.0
0
P(S|s)
0
1.0
P(S|n)
HIGH-THRESHOLD MODEL
EQUAL-VARIANCE MODEL
HEINEMANN MODIFICATION
"SMALL GROUPING" MODIFICATION
ROC CURVES:
Various Models

Figure 14

a. Fish 12. To reduce a normally-high false-alarm rate, this fish was threshold-tested with the gate (Figure 8) in place. Average false-alarm probability during the threshold test was 0.15; except for the first six trials, not counted as part of the threshold run, the false-alarm rate was approximately constant. The threshold plot itself is shown in Figure 15. The threshold value of 39.5 dB is high for a goldfish critical ratio, but there is no reason to doubt its validity. Average latency was 5.5 seconds.

Following the threshold test, an ROC curve was obtained at a S/N ratio of 40 dB, very close to the threshold value. Trials were continued at 40 dB S/N, using various combinations of no gate, one side of the gate only, and the full gate, to obtain a wide range of false-alarm ratios by inducing the fish to change his criterion. The results, summarized by 10-trial sequences, are shown in Table I; the final ROC curve is shown in Figure 16. It is obvious that the high-threshold model fails completely to fit the data, and that the conventional equal-variance signal detection model, fitted by eye with d' of 1.0, gives a good fit. Two sets of points are given in Figure 16; averages for one run of (usually) 50 consecutive trials; and averages of 10-trial sequences grouped by false-alarm ratios, as appropriate for the "small-group" model earlier described. For the latter points, while the equal-variance signal detection model gives a significantly better fit than the high-threshold model, the "small-group" model fit is significantly better still; see "Results" Section IIg.



FISH # 12
THRESHOLD
S/N In dB
6/13/71 WS

Figure 15

TABLE I: Fish # 12 Data. 10 Trial Sequences Averaged.

S/N = 40 dB

| P(S/s) | P(S/n) | Conditions |
|---|---|---|
| (6/13/71-12-2) | | $\bar{c}$ gate |
| .5 | .025 | |
| .5 | .100 | |
| .6 | 0 | |
| .3 | .025 | |
| .4 | .025 | |
| .48 | .055 | |
| (6/13/71-12-3) | | $\bar{c}$ 1/2 gate |
| .9 | .275 | |
| .7 | .250 | |
| .3 | .050 | |
| .633 | .192 | |
| (6/13/71-12-4) | | $\bar{s}$ gate |
| .7 | .200 | |
| .1 | .100 | |
| .5 | .200 | |
| .7 | .275 | |
| .8 | .450 | |
| .560 | .245 | |
| (6/14/71-12-1) | | $\bar{s}$ gate |
| 1.0 | 1.0 | |
| .9 | .950 | |
| 1.0 | 1.0 | |
| 1.0 | 1.0 | |
| .975 | .988 | |
| (6/14/71-12-2) | | $\bar{c}$ gate |
| .1 | 0 | |
| .3 | .075 | |
| .5 | .150 | |
| .300 | .075 | |
| (6/18/71-12-1) | | $\bar{c}$ gate |
| .4 | .025 | |
| .8 | .200 | |
| .7 | .175 | |
| .5 | .050 | |
| .7 | .125 | |
| .620 | .115 | |

| P(S/s) | P(S/n) | Cond's |
|---|---|---|
| (6/23/71-12-1) | | $\bar{s}$ gate |
| .8 | .600 | |
| 1.0 | .750 | |
| 1.0 | 1.0 | |
| .933 | .783 | |
| (6/23/71-12-2) | | $\bar{s}$ gate,<br>$\bar{s}$ light |
| .3 | .050 | |
| 0 | 0 | |
| 0 | 0 | |
| 0 | 0 | |
| .1 | .025 | |
| 0 | .025 | |
| .067 | .017 | |
| (7/15/71-12-1) | | $\bar{c}$ platform |
| 1.0 | 1.0 | |
| 1.0 | .900 | |
| .8 | .600 | |
| .933 | .833 | |
| (6/18/71-12-2) | | S/N down 10 dB |
| .8 | .325 | |
| .7 | .350 | |
| .7 | .450 | |
| .4 | .225 | |
| .9 | .375 | |
| .70 | .345 | d'=0.9 |

**TABLE I.** Cont'd: Data Grouped by P(S/n)

S/N = 40 dB

$P(S/n)=0$

| P(S/s) | P(S/n) |
|---|---|
| .6 | 0 |
| .1 | 0 |
| 0 | 0 |
| 0 | 0 |
| 0 | 0 |
| .14 | 0 |

$0<P(S/n)<.05$

| P(S/s) | P(S/n) |
|---|---|
| .5 | .025 |
| .3 | .025 |
| .4 | .025 |
| .1 | .025 |
| 0 | .025 |
| .4 | .025 |
| .283 | .025 |

$.05<P(S/n)<.1$

| P(S/s) | P(S/n) |
|---|---|
| .3 | .050 |
| .3 | .075 |
| .5 | .050 |
| .3 | .050 |
| .35 | .056 |

$.1\leq P(S/n)<.2$

| | |
|---|---|
| .5 | .100 |
| .1 | .100 |
| .5 | .150 |
| .7 | .175 |
| .50 | .130 |

$.2\leq P(S/n)<.3$

| | |
|---|---|
| .9 | .275 |
| .7 | .250 |
| .7 | .200 |
| .5 | .200 |
| .7 | .275 |
| .8 | .200 |
| .4 | .225 |
| .671 | .232 |

$.3\leq P(S/n)<.5$

| | |
|---|---|
| .8 | .450 |
| .8 | .325 |
| .7 | .350 |
| .7 | .450 |
| .9 | .375 |
| .78 | .390 |

$.5\leq P(S/n)<.8$

| | |
|---|---|
| .8 | .600 |
| 1.0 | .750 |
| .8 | .600 |
| .867 | .650 |

$.8\leq P(S/n)<1.0$

| | |
|---|---|
| 1.0 | 1.0 |
| .9 | .950 |
| 1.0 | 1.0 |
| 1.0 | 1.0 |
| 1.0 | 1.0 |
| 1.0 | 1.0 |
| 1.0 | .900 |
| .986 | .979 |

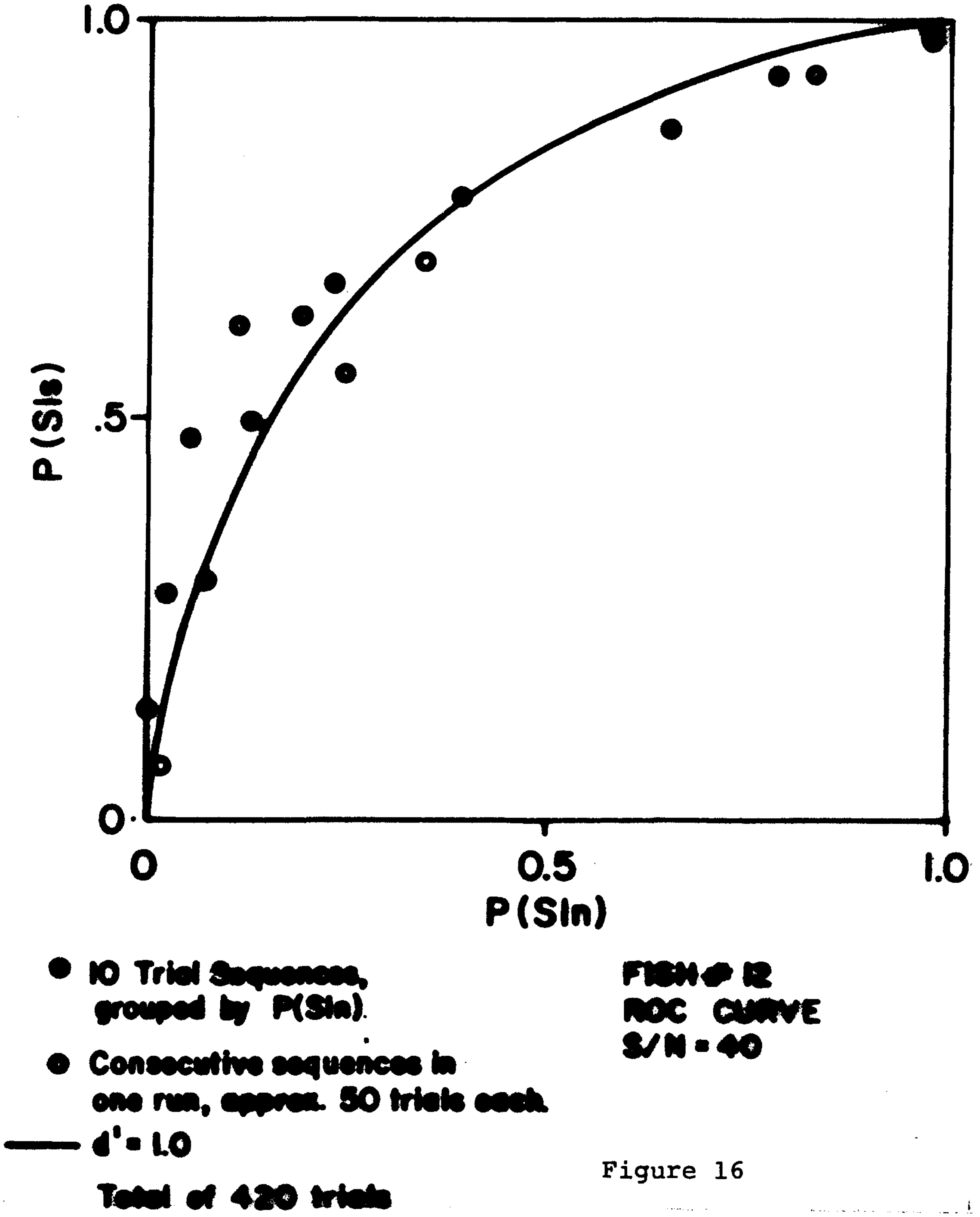

1.0
.5
0
0
0.5
1.0
P(S|s)
P(S|n)
● 10 Trial Sequences, grouped by P(S|n).
⊙ Consecutive sequences in one run, approx. 50 trials each.
—— d' = 1.0
Total of 420 trials
FISH # 12
ROC CURVE
S/N = 40

Figure 16

To determine if the high 40 dB threshold value was due to very high internal noise, an additional set of 50 trials was run with the signal and noise both reduced by 10 dB. Results are shown in Table I; the average d' remained at 0.9, indicating that internal noise was not the limiting factor.

b. Fish 2. An extremely variable false-alarm rate exhibited by this fish prohibited carrying out a conventional threshold test. S/N ratio was arbitrarily set at 35 dB for ROC curve determination. Although the initial false-alarm probability was usually very close to unity, within occasional groups of ten trials obvious abrupt criterion changes occurred; in such cases, that particular group of ten trials was eliminated from the ten-trial-average groupings, but not from the trial-by-trial groupings later described. Tests were carried out with and without gate; even with only these two conditions, a wide range of false-alarm probabilities was observed. The ten-trial sequences are tabulated in Table II, and the corresponding ROC curve is given in Figure 17. The sequences grouped by false-alarm ratio are well fitted by the equal-variance signal detection model with d' of 0.8; a better fit, but not significantly so, was obtained by the small-group model, as shown in Section IIg. The 50-trial sequence points show more scatter, points in the middle of the curve being depressed below the best-fit curve for the points grouped by P(S/n); this is to be expected in the presence of a varying

**TABLE II: Fish # 2 Data. 10-Trial Sequences Averaged.**

S/n = 35 dB

| P(S/s) | P(S/n) | Conditions |
|---|---|---|
| (6/28/71-2-1) | | $\bar{s}$ gate |
| 1.0 | 1.0 | |
| 1.0 | .875 | |
| 1.0 | 1.0 | |
| 1.0 | .958 | |
| (6/28/71-2-2) | | $\bar{c}$ gate |
| .5 | .100 | |
| .8 | .450 | |
| .2 | 0 | |
| .8 | .475 | |
| .4 | .150 | |
| .8 | .250 | |
| .4 | .225 | |
| 0 | 0 | |
| 0 | 0 | |
| .433 | .183 | |
| (6/29/71-2-1) | | $\bar{s}$ gate |
| 1.0 | 1.0 | |
| 1.0 | 1.0 | |
| 1.0 | .975 | |
| 1.0 | .992 | |
| (6/29/71-2-2) | | $\bar{c}$ gate |
| .4 | .150 | |
| .6 | .275 | |
| .5 | .175 | |
| .6 | .325 | |
| .3 | .050 | |
| .480 | .195 | |
| (6/29/71-2-3) | | $\bar{c}$ gate |
| .3 | 0 | |
| .3 | 0 | |
| .1 | 0 | |
| .233 | 0 | |

| P(S/s) | P(S/n) | Cond's |
|---|---|---|
| (7/10/71-2+1) | | $\bar{s}$ gate |
| .9 | .725 | |
| .8 | .575 | |
| .4 | .250 | |
| 0 | .025 | |
| .525 | .394 | |
| (7/11/71-2-1) | | $\bar{s}$ gate |
| .9 | .775 | |
| .9 | .700 | |
| 0 | .025 | |
| .60 | .500 | |

## TABLE II Cont'd: Data Grouped by P(S/n)

S/N = 35 dB

$0 \leq P(S/n) < .05$

| P(S/s) | P(S/n) |
|---|---|
| .2 | 0 |
| 0 | 0 |
| 0 | 0 |
| .3 | 0 |
| .1 | 0 |
| 0 | .025 |
| 0 | .025 |
| .113 | .006 |

$.05 \leq P(S/n) < .1$

| P(S/s) | P(S/n) |
|---|---|
| .3 | .050 |
| .3 | .050 |

$.1 \leq P(S/n) < .2$

| P(S/s) | P(S/n) |
|---|---|
| .5 | .100 |
| .4 | .150 |
| .4 | .150 |
| .5 | .175 |
| .45 | .144 |

$.2 \leq P(S/n) < .3$

| | |
|---|---|
| .4 | .225 |
| .8 | .250 |
| .6 | .275 |
| .4 | .250 |
| .55 | .250 |

$.3 \leq P(S/n) < .4$

| | |
|---|---|
| .6 | .325 |
| .6 | .325 |

$.4 \leq P(S/n) < .5$

| | |
|---|---|
| .8 | .450 |
| .8 | .475 |
| .8 | .462 |

$.5 \leq P(S/n) < .6$

| | |
|---|---|
| .8 | .575 |
| .8 | .575 |

$.6 \leq P(S/n) < .7$

None

$.7 \leq P(S/n) < .8$

| | |
|---|---|
| .9 | .725 |
| .9 | .775 |
| .9 | .700 |
| .90 | .733 |

$.8 \leq P(S/n) < .9$

| | |
|---|---|
| 1.0 | .875 |
| 1.0 | .875 |

$.9 \leq P(S/n) < 1.0$

| | |
|---|---|
| 1.0 | 1.0 |
| 1.0 | 1.0 |
| 1.0 | 1.0 |
| 1.0 | 1.0 |
| 1.0 | .975 |
| 1.00 | .995 |

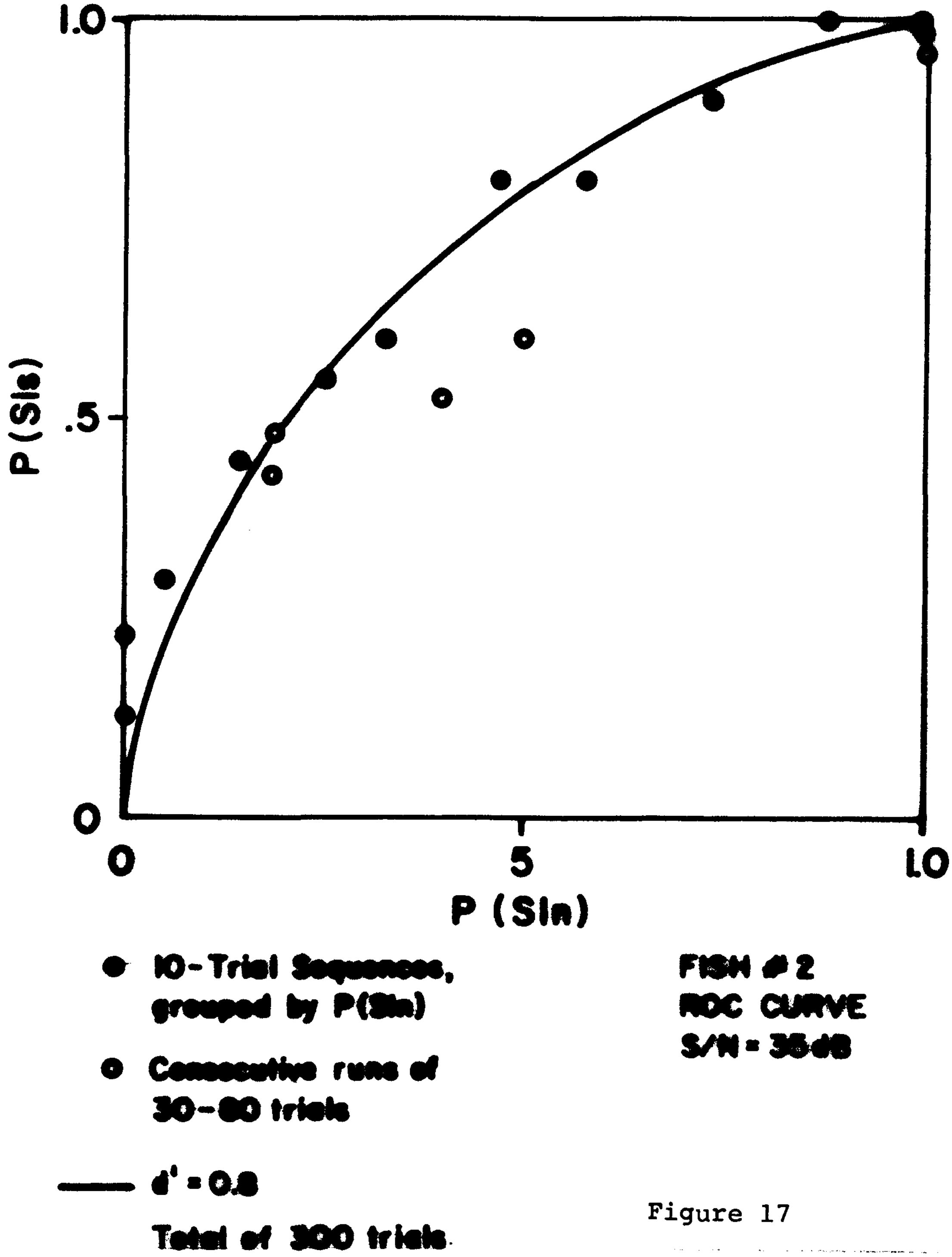

1.0
.5
0
P (S|s)
0
5
1.0
P (S|n)
● 10-Trial Sequences, grouped by P(S|n)
○ Consecutive runs of 30-80 trials
d' = 0.8
Total of 300 trials
FISH # 2
ROC CURVE
S/N = 35dB

Figure 17

criterion (Green and Swets, 1966, p. 411). While the 50-trial points could be fitted by the high-threshold model almost as well as by the signal detection model, neither fit is very good, the fit of the small-group model to the points grouped by P(S/n) being much better.

As a further test of the small-group model, the data for fish 2 were retabulated, grouping on a trial-by-trial basis on observed P(S/n). In this case, those 10-trial sequences previously eliminated because of abrupt criterion changes were now included. The results are shown in Figure 12, previously referred to in the Theoretical Results section. These data were fitted fairly well by the small-group model, except for the point at zero P(S/n): considering the arbitrary prior distribution for P(S/n), the fit can be considered satisfactory. The d' value yielded by the trial-by-trial grouping, 0.7, is almost the same as that yielded by the ten-trial sequences, 0.8; in view of the marked difference in the shapes of the curves fitted, this agreement is encouraging. Under no circumstances can the high-threshold or the Heinemann models apply to the points grouped by P(S/n).

c. Fish 3. This fish appeared remarkably steady during training and final testing; his usual resting position between trials was facing the hurdle, in the center of the shuttlebox partition, with virtually no "random" motion around the side of the shuttlebox which he occupied. A normal threshold test was carried out, without gate, giving a threshold of 18.5 dB with false-alarm probability of 0.24. Because the false-alarm rate at threshold was somewhat higher than usually considered allowable, ROC testing began at S/N level of 30 dB; the results, plotted in Figure 18, indicated a rather high d' value of about 2, so testing at S/N 30 was discontinued, and extensive testing at S/N equal to 25 was carried out.

The ten-trial sequence data are given in Table III, and plotted, together with the 50-trial sequences, in Figure 19. Neither the high-threshold nor the signal-detection models would fit these data very well. The shape of the curve indicated that perhaps, in spite of the apparently stable behavior of the fish, that the type of random activity taken into account by the Heinemann model might be occurring. The data points using the 50-trial sequences should then permit determination whether day-by-day variation in random activity, combined with a relatively constant criterion, might be taking place. Figure 19 shows that nearly all the 50-trial sequence data points are included between the equal-variance model with d' = 1.6, and the Heinemann modification with d' = 1.6 and random response probability r = 0.15. In fact, the 50-trial points in Figure 19 closely resemble the diagram in Figure 13 illustrating the effect of the Heinemann modification, drawn some months before the data in Figure 19 were taken. Some random activity might have been forecast from the

<u>TABLE III</u>: Fish # 3 data. 10-Trial Sequences Averaged.

S/N = 25 dB

| P(S/s) | P(S/n) | Conditions |
|---|---|---|
| (6/10/71-3-3) | | $\bar{s}$ gate |
| .3 | 0 | |
| .5 | .250 | |
| .8 | .350 | |
| .7 | .250 | |
| .5 | .100 | |
| .50 | .190 | |
| (6/10/71-3-4) | | $\bar{s}$ gate |
| .6 | .050 | |
| .8 | .225 | |
| .6 | .225 | |
| .6 | .225 | |
| .7 | .200 | |
| .66 | .185 | |
| (6/12/71-3-1) | | $\bar{s}$ gate |
| .7 | .250 | |
| .6 | .275 | |
| .9 | .025 | |
| .7 | .100 | |
| .8 | .200 | |
| .72 | .170 | |
| (6/12/71-3-2) | | $\bar{s}$ gate |
| .8 | .075 | |
| .9 | .025 | |
| .9 | .275 | |
| 1.0 | .450 | |
| .8 | .675 | |
| .88 | .300 | |
| (6/15/71-3-1) | | $\bar{s}$ gate |
| .9 | .075 | |
| .8 | .075 | |
| .5 | .050 | |
| .8 | .125 | |
| .5 | .250 | |
| .70 | .115 | |

| P(S/s) | P(S/n) | Cond's |
|---|---|---|
| (6/22/71-3±1) | | $\bar{c}$ gate |
| .2 | .075 | |
| .3 | .100 | |
| .6 | .250 | |
| .37 | .142 | |
| (6/24/71-3-1)<br>(First 50 trials) | | $\bar{s}$ gate |
| .6 | .125 | |
| .5 | .200 | |
| .6 | .300 | |
| .3 | .200 | |
| .5 | 0 | |
| .50 | .165 | |
| (Second 50 trials) | | |
| .5 | .050 | |
| .1 | .025 | |
| .5 | 0 | |
| .3 | .050 | |
| .3 | .050 | |
| .34 | .035 | |
| (6/26/71-3-1) | | $\bar{c}$ gate |
| .7 | .450 | |
| .5 | .150 | |
| .4 | .100 | |
| .6 | .125 | |
| .5 | .100 | |
| .54 | .185 | |
| (6/26/71-3-1) cont'd) | | |
| .1 | 0 | |
| .2 | .025 | |
| .4 | .075 | |
| .3 | 0 | |
| 0 | 0 | |
| .20 | .020 | |

**TABLE III Cont'd:** Fish # 3 Data. 10-Trial Sequences Averaged.

S/N = 25 dB

| P(S/s) | P(S/n) | Conditions |
|---|---|---|
| (6/26/71-3-2) | | $\bar{s}$ gate |
| .4 | 0 | |
| .3 | .250 | |
| .3 | .025 | |
| .4 | .200 | |
| .3 | .175 | |
| .34 | .130 | |
| (6/26/71-3-2 cont'd) | | |
| 0 | 0 | |
| .5 | .050 | |
| .7 | .050 | |
| .6 | .175 | |
| .9 | .150 | |
| .54 | .085 | |
| (6/26/71-3-3) | | $\bar{c}$ gate |
| .5 | 0 | |
| .6 | .075 | |
| .6 | .075 | |
| .8 | .125 | |
| .7 | .250 | |
| .64 | .105 | |
| (7/13/71-3-1) | | $\bar{s}$ gate |
| .7 | .150 | |
| .6 | .025 | |
| .7 | 0 | |
| .7 | .125 | |
| .5 | .150 | |
| .64 | .090 | |
| (7/13/71-3-2) | | $\bar{c}$ platform |
| .5 | .300 | |
| .2 | .050 | |
| .2 | .075 | |
| .4 | .225 | |
| .2 | .150 | |
| .30 | .150 | |

| P(S/s) | P(S/n) | Cond's |
|---|---|---|
| (7/13/71-2-3) | | $\bar{c}$ platform |
| .2 | .050 | |
| .2 | .075 | |
| .2 | .025 | |
| .4 | .075 | |
| .25 | .056 | |
| (7/18/71-3-1) | | $\bar{c}$ platform |
| .1 | .125 | |
| .3 | .050 | |
| .4 | .200 | |
| .7 | .200 | |
| .3 | .025 | |
| .36 | .120 | |

**TABLE III** Cont'd: Fish # 3 Data Grouped by P(S/n).

S/N = 25 dB

$0 \leq P(S/n) < .05$

| P(S/s) | P(S/n) |
|---|---|
| .3 | 0 |
| .9 | .025 |
| .9 | .025 |
| .5 | 0 |
| .1 | .025 |
| .5 | 0 |
| .1 | 0 |
| .2 | .025 |
| .3 | 0 |
| 0 | 0 |
| .5 | 0 |
| .6 | .025 |
| .7 | 0 |
| .2 | .025 |
| .3 | .025 |
| .407 | .012 |

$.05 \leq P(S/n) < .1$

| P(S/s) | P(S/n) |
|---|---|
| .6 | .050 |
| .8 | .075 |
| .9 | .075 |
| .8 | .075 |
| .5 | .050 |
| .2 | .075 |
| .5 | .050 |
| .3 | .050 |
| .3 | .050 |
| .4 | .075 |
| .6 | .075 |
| .6 | .075 |
| .2 | .050 |
| .2 | .075 |
| .2 | .050 |
| .2 | .075 |
| .4 | .075 |
| .3 | .050 |
| .2 | .050 |
| .432 | .063 |

$.1 \leq P(S/n) < .2$

| P(S/s) | P(S/n) |
|---|---|
| .5 | .100 |
| .7 | .100 |
| .8 | .125 |
| .3 | .100 |
| .6 | .125 |
| .5 | .150 |
| .4 | .100 |
| .6 | .125 |
| .5 | .100 |
| .8 | .125 |
| .7 | .150 |
| .1 | .125 |
| .7 | .125 |
| .5 | .150 |
| .2 | .100 |
| .527 | .120 |

$.2 \leq P(S/n) < .3$

| | |
|---|---|
| .5 | .250 |
| .7 | .250 |
| .8 | .225 |
| .6 | .225 |
| .6 | .225 |
| .7 | .200 |
| .7 | .250 |
| .6 | .275 |
| .8 | .200 |
| .9 | .275 |
| .5 | .250 |
| .6 | .250 |
| .5 | .200 |
| .3 | .200 |
| .7 | .250 |
| .4 | .225 |
| .4 | .200 |
| .7 | .200 |
| .611 | .231 |

$.3 \leq P(S/n) < .4$

| | |
|---|---|
| .8 | .350 |
| .6 | .300 |
| .5 | .300 |
| .633 | .317 |

$.4 \leq P(S/n) < .5$

| | |
|---|---|
| 1.0 | .450 |
| .7 | .450 |
| .85 | .450 |

$.5 \leq P(S/n) < .6$

None

$.6 \leq P(S/n) < .7$

| | |
|---|---|
| .8 | .675 |
| .8 | .675 |

$.7 \leq P(S/n) \leq 1.0$

None

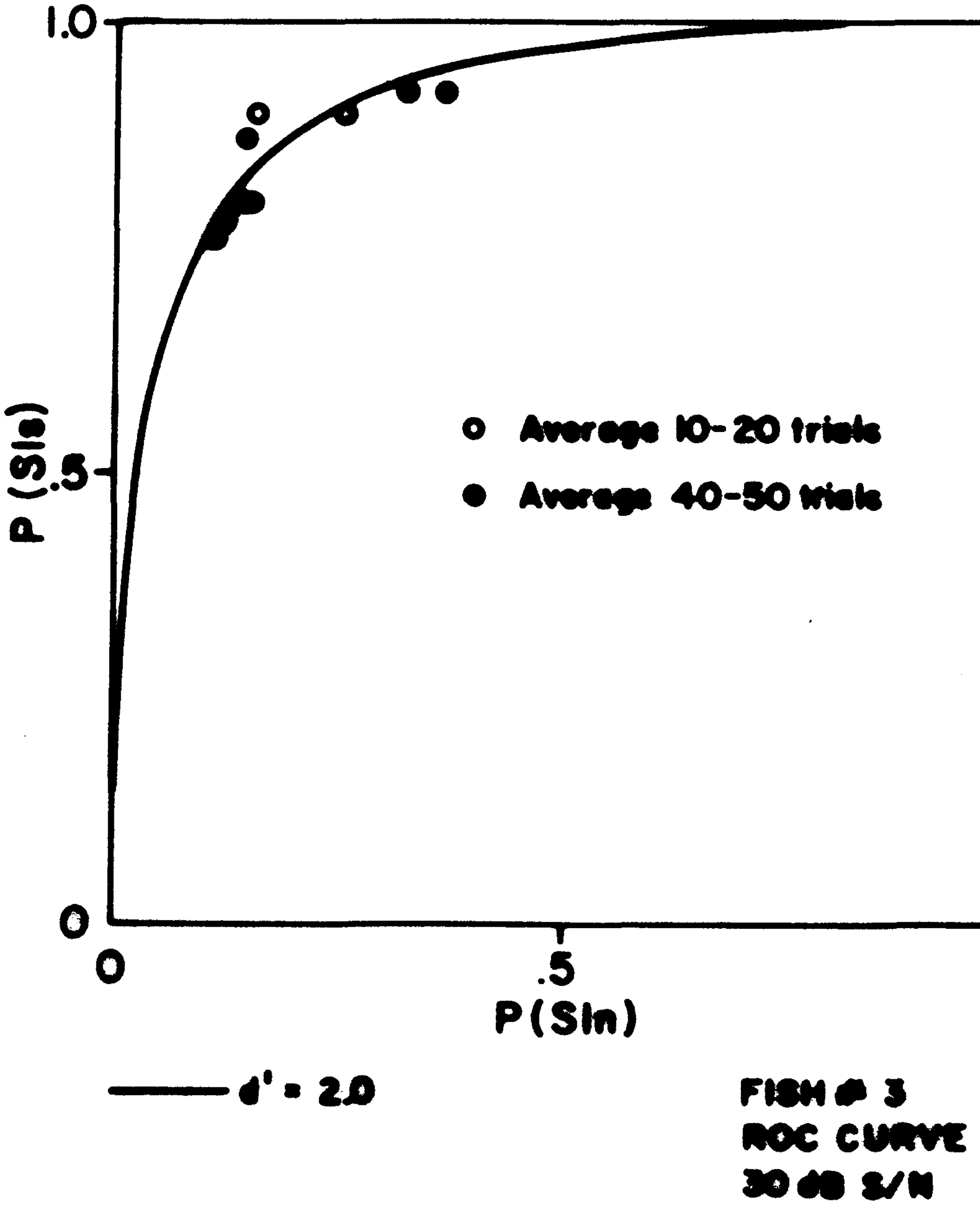

1.0
.5
0
P (S|s)
0
.5
1.0
P(S|n)
○ Average 10-20 trials
● Average 40-50 trials
d' = 2.0
FISH # 3
ROC CURVE
30 dB S/N

Figure 18

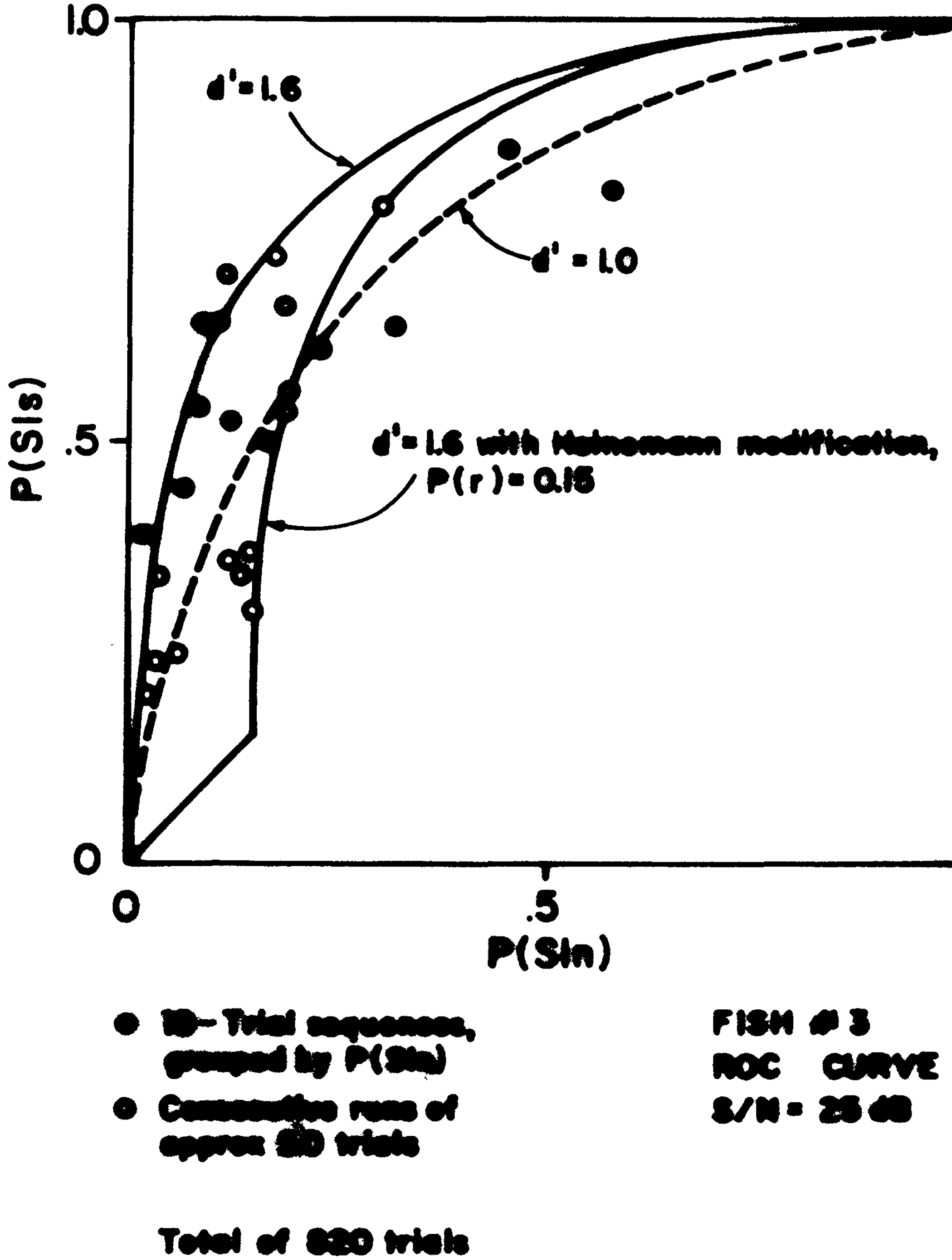
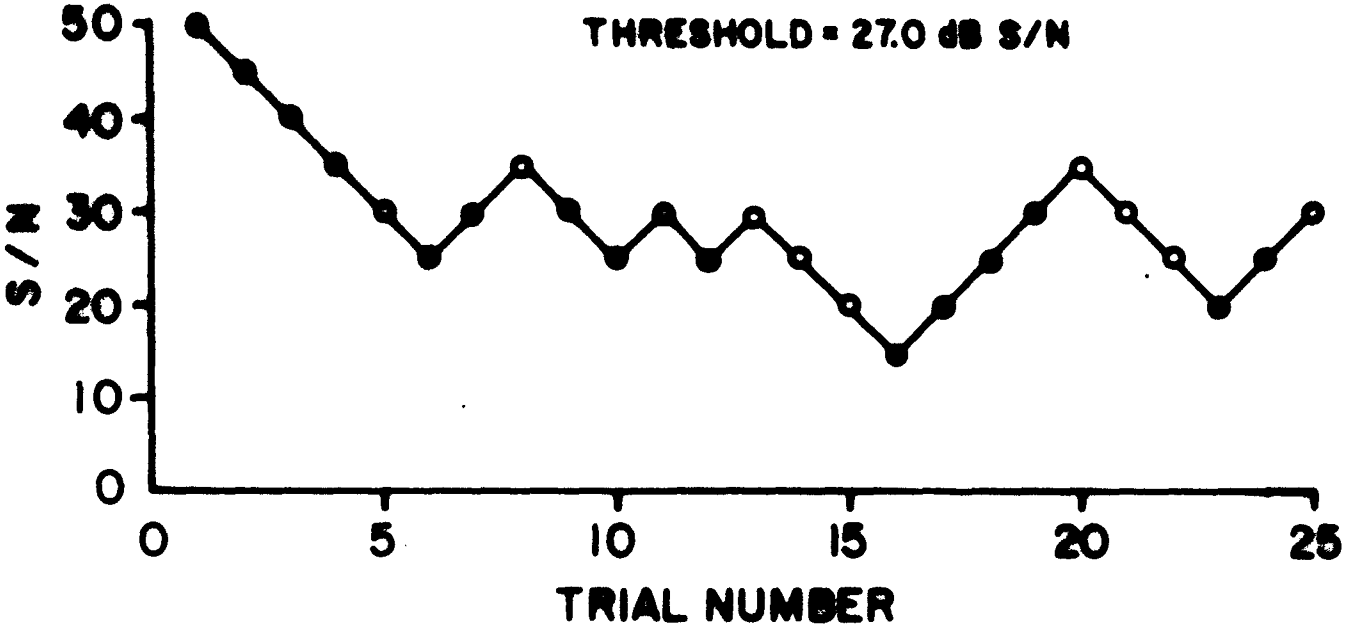
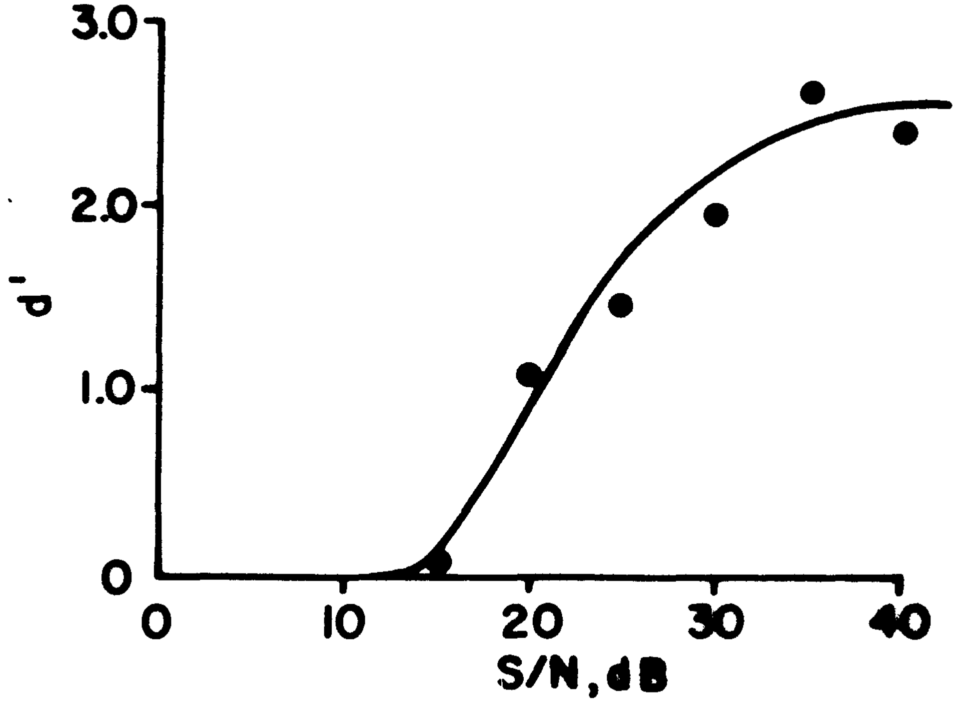

1.0
.5
0
P(S|s)
0
.5
1.0
P(S|n)
d' = 1.6
d' = 1.0
d' = 1.6 with Heinemann modification,
P(r) = 0.15
● 10-Trial sequences, grouped by P(S|n)
○ Consecutive runs of approx 20 trials
Total of 820 trials
FISH # 3
ROC CURVE
S/N = 25 dB

Figure 19



**FISH # 6**
**THRESHOLD # 3**

Figure 20



FISH # 3
PSYCHOMETRIC FUNCTION

Figure 21

slightly erratic threshold data, shown in Figure 20. The major difference between Fish 3's performance and the Heinemann model is that the random response probability is itself a random variable, varying roughly uniformly between the limits of 0 and 0.15.

Data for a psychometric function of d' versus S/N ratio were taken at S/N ranging from 15 to 40 dB. The results, shown in Figure 21, indicate an initial sharp rise in d' with S/N, leveling off at a d' of about 2.5.

d. Fish 6. Threshold tests on this fish showed immediately that erratic responses to the signal were taking place. Three thresholds, shown in Figures 22, 23 and 24, were taken, with threshold values ranging from 24 to 32 dB, depending upon which date and which 10 reversals were selected. ROC points were collected at 25, 30 and 35 dB S/N, all showing erratic behavior. The data, with various methods of grouping, are shown in Figures 25 and 26, and are tabulated in Table IV.

Only the Heinemann model can account for such data, since of the models under consideration, only this model can account for data points which lie anywhere between the conventional signal-detection ROC curve and the positive diagonal. All other models fail completely. Due to the extreme variability, one cannot assign d' values to the various S/N ratios with any confidence, except that they seem to be at least 1.5 or higher, and probably monotonically increasing with S/N. Apparently, the probability

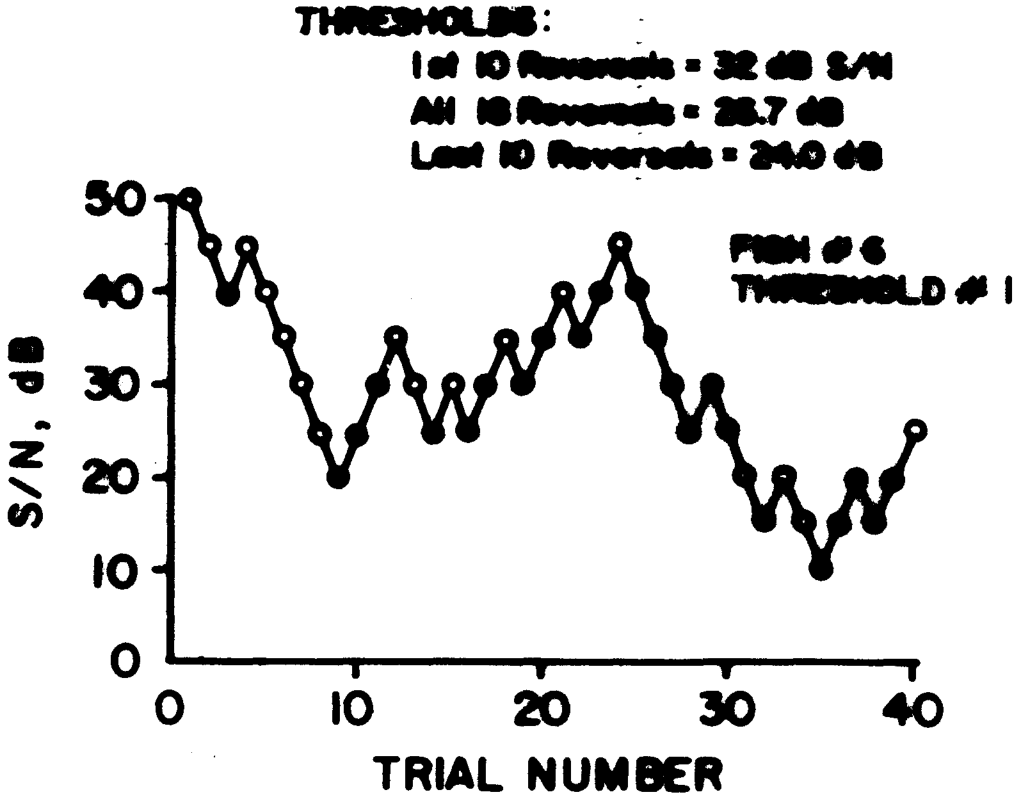


Figure 23

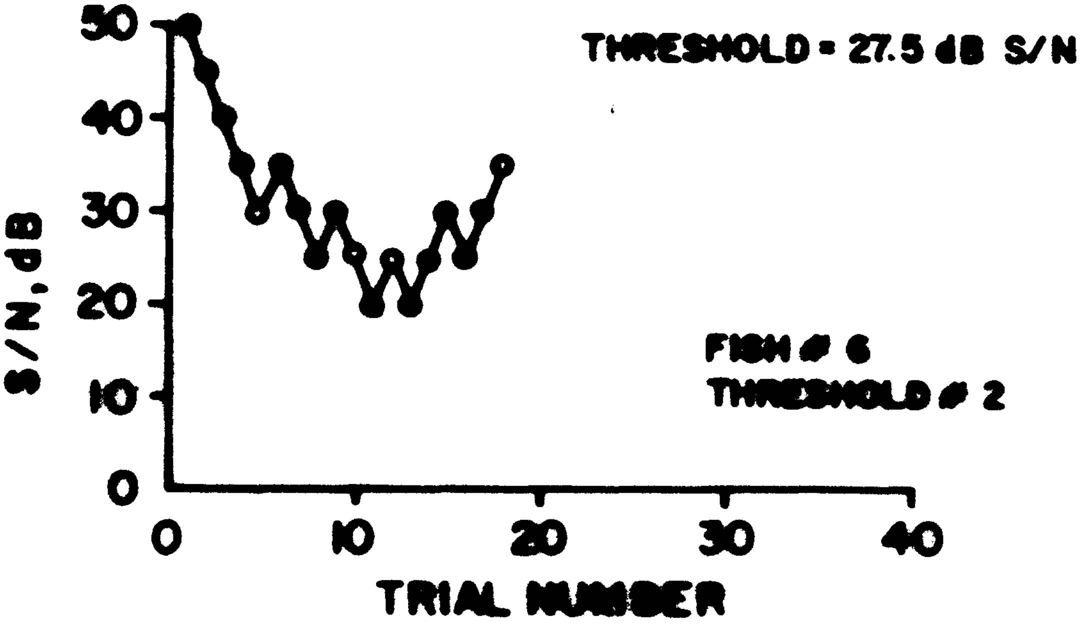


Figure 22

Figure 24

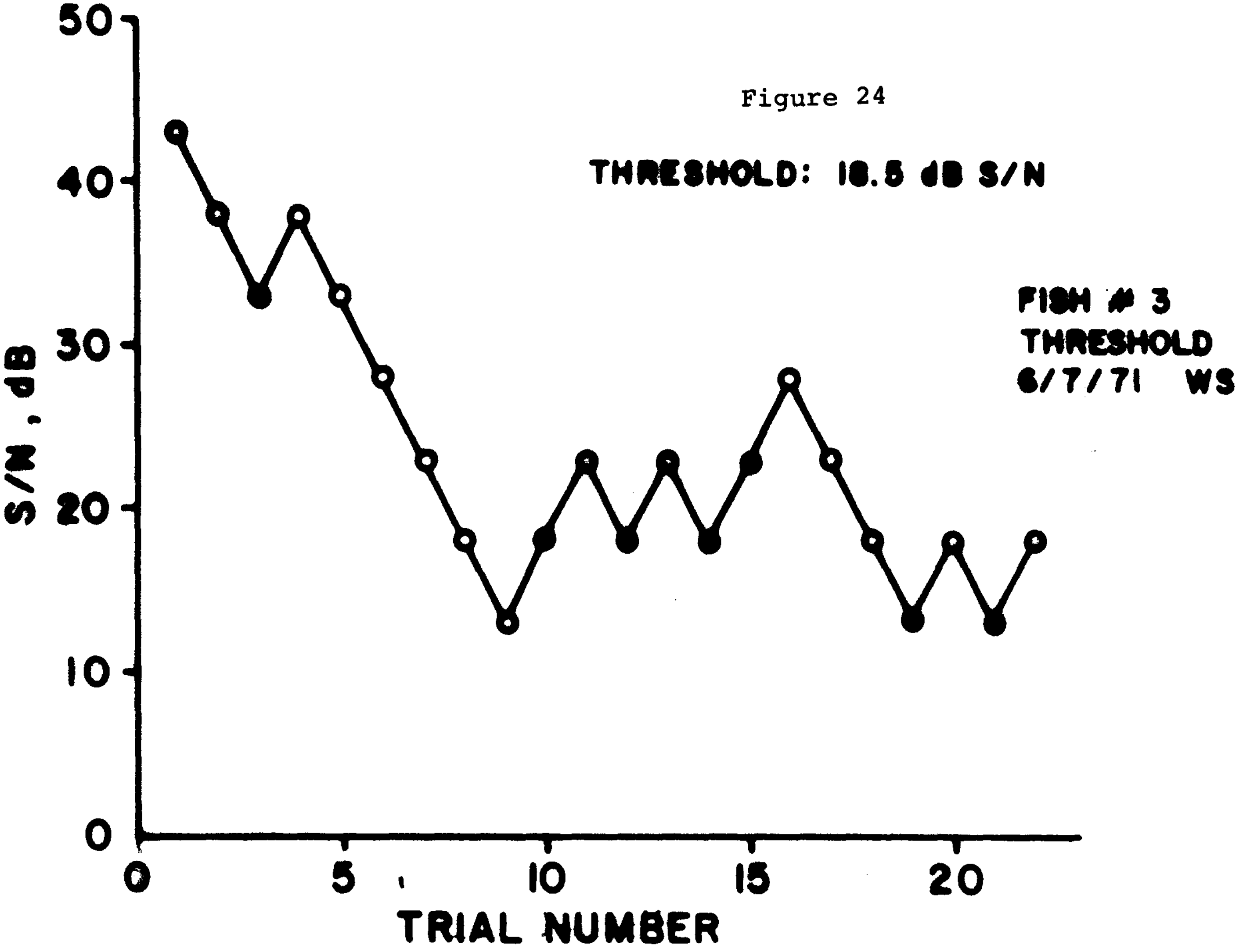

THRESHOLD: 18.5 dB S/N
FISH # 3
THRESHOLD
6/7/71 WS
S/N, dB
50
40
30
20
10
0
0
5
10
15
20
TRIAL NUMBER

**TABLE IV:** Fish # 6 Data. 10 Trial Sequences Averaged.

S/N = 25 dB

| P(S/s) | P(S/n) | Conditions |
|---|---|---|
| (6/26/71-6-1) | | $\bar{c}$ gate |
| .5 | .075 | |
| .4 | .100 | |
| .5 | .025 | |
| .3 | .100 | |
| .3 | .025 | |
| .40 | .065 | |
| (6/26/71-6-2) | | $\bar{c}$ gate |
| .3 | .025 | |
| .6 | 0 | |
| .3 | 0 | |
| .2 | 0 | |
| .2 | 0 | |
| .32 | .005 | |
| (6/26/71-6-3) | | $\bar{s}$ gate |
| .2 | .250 | |
| .5 | .225 | |
| .8 | .175 | |
| .8 | .300 | |
| .8 | .050 | |
| .62 | .200 | |
| (6/27/71-6-1) | | $\bar{s}$ gate |
| .8 | .475 | |
| .8 | .575 | |
| .6 | .450 | |
| .6 | .575 | |
| .6 | .675 | |
| .66 | .550 | |
| (6/27/71-6-2) | | $\bar{c}$ gate |
| .8 | .450 | |
| .4 | .450 | |
| .3 | .400 | |
| .2 | .400 | |
| .4 | .450 | |
| .42 | .430 | |

| P(S/s) | P(S/n) | Cond's |
|---|---|---|
| (6/28/71-6-1) | | $\bar{s}$ gate |
| .8 | .650 | |
| .7 | .450 | |
| .5 | .625 | |
| .8 | .600 | |
| .6 | .650 | |
| .68 | .595 | |
| (6/28/71-6-2) | | $\bar{c}$ gate |
| 1.0 | .825 | |
| .9 | .800 | |
| .8 | .775 | |
| .7 | .550 | |
| .9 | .475 | |
| .8 | .475 | |
| .7 | .400 | |
| .83 | .614 | |
| (6/28/71-6-3) | | $\bar{c}$ gate<br>$\bar{s}$ light |
| .2 | .125 | |
| .6 | .425 | |
| .4 | .275 | |
| (6/29/71-6-1) | | $\bar{c}$ gate |
| .7 | .375 | |
| 1.0 | .550 | |
| .9 | .675 | |
| .9 | .800 | |
| .9 | .650 | |
| .88 | .610 | |
| (6/29/71-6-2) | | $\bar{c}$ gate |
| .6 | .575 | |
| 1.0 | .550 | |
| .7 | .500 | |
| 1.0 | .700 | |
| .8 | .525 | |
| .82 | .570 | |

<u>TABLE IV Cont'd</u>: 10 Trial Sequences Averaged.

S/N = 25 dB

| P(S/s) | P(S/n) | Conditions |
|---|---|---|
| (6/29/71-6-3) | | $\bar{c}$ gate, water level over hurdle cut in half. |
| .8 | .175 | |
| .3 | .175 | |
| .3 | .300 | |
| .4 | .075 | |
| .6 | .075 | |
| .48 | .160 | |

TABLE IV Cont'd: Data Grouped by P(S/n).

S/N = 25 dB

$0 \leq P(S/n) < .05$

| | |
|---|---|
| .5 | .025 |
| .3 | .025 |
| .3 | .025 |
| .6 | 0 |
| .3 | 0 |
| .2 | 0 |
| .2 | 0 |
| .343 | .011 |

$.05 \leq P(S/n) < .1$

| | |
|---|---|
| .5 | .075 |
| .8 | .050 |
| .4 | .075 |
| .6 | .075 |
| .575 | .069 |

$.1 \leq P(S/n) < .2$

| | |
|---|---|
| .4 | .100 |
| .3 | .100 |
| .8 | .175 |
| .2 | .125 |
| .8 | .175 |
| .3 | .175 |
| .467 | .142 |

$.2 \leq P(S/n) < .3$

| | |
|---|---|
| .2 | .250 |
| .5 | .225 |
| .35 | .238 |

$.3 \leq P(S/n) < .4$

| | |
|---|---|
| .8 | .300 |
| .7 | .375 |
| .3 | .300 |
| .60 | .325 |

$.4 \leq P(S/n) < .5$

| | |
|---|---|
| .8 | .475 |
| .6 | .450 |
| .8 | .450 |
| .4 | .450 |
| .3 | .400 |
| .2 | .400 |
| .4 | .450 |
| .7 | .450 |
| .9 | .475 |
| .8 | .475 |
| .7 | .400 |
| .6 | .425 |
| .600 | .442 |

$.5 \leq P(S/n) < .6$

| | |
|---|---|
| .8 | .575 |
| .6 | .575 |
| .7 | .550 |
| 1.0 | .550 |
| .6 | .575 |
| 1.0 | .550 |
| .7 | .500 |
| .8 | .525 |
| .775 | .550 |

$.6 \leq P(S/n) < .7$

| | |
|---|---|
| .6 | .675 |
| .8 | .650 |
| .6 | .625 |
| .8 | .600 |
| .6 | .650 |
| .9 | .675 |
| .9 | .650 |
| .743 | .646 |

$.7 \leq P(S/n) < .8$

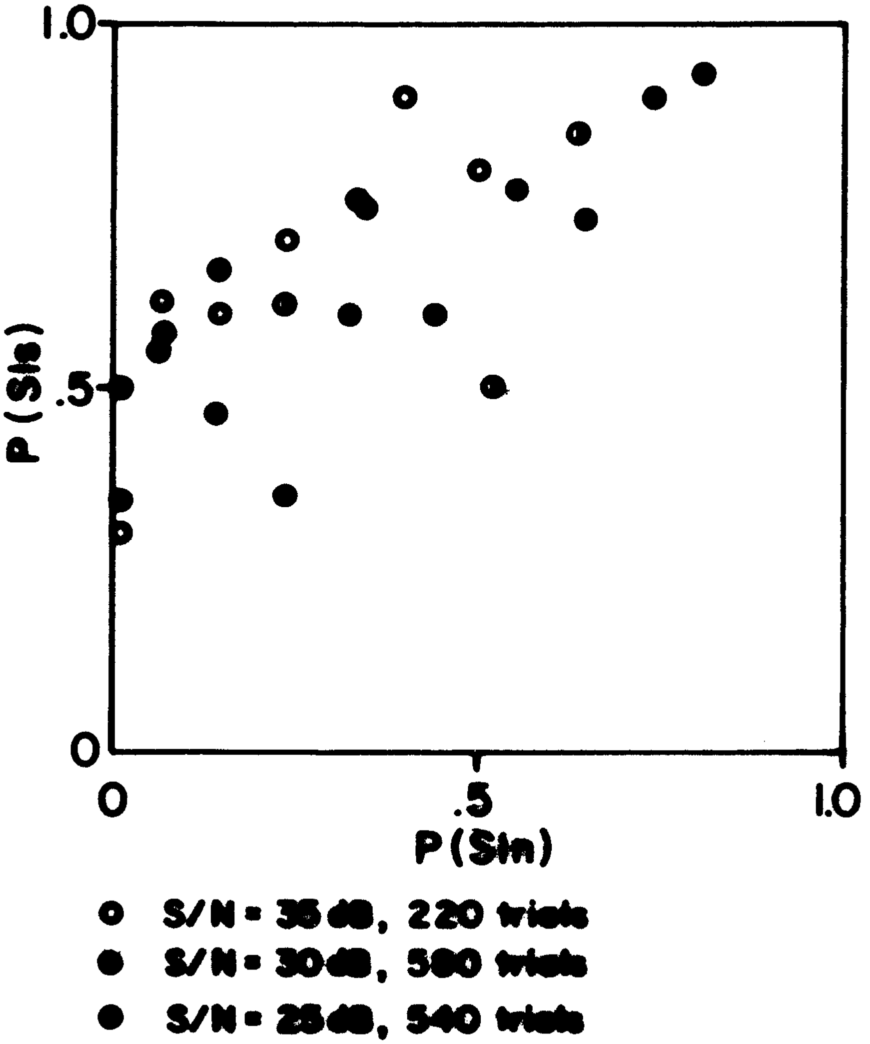
| | |
|---|---|
| .8 | .775 |
| 1.0 | .700 |
| .90 | .738 |

$.8 \leq P(S/n) < .9$

| | |
|---|---|
| 1.0 | .825 |
| .9 | .800 |
| .9 | .800 |
| .933 | .738 |

$.9 \leq P(S/n) \leq 1.0$

None



FISH, # 6
ROC, grouped by
P(S|n)
Various S/N ratios

Figure 25

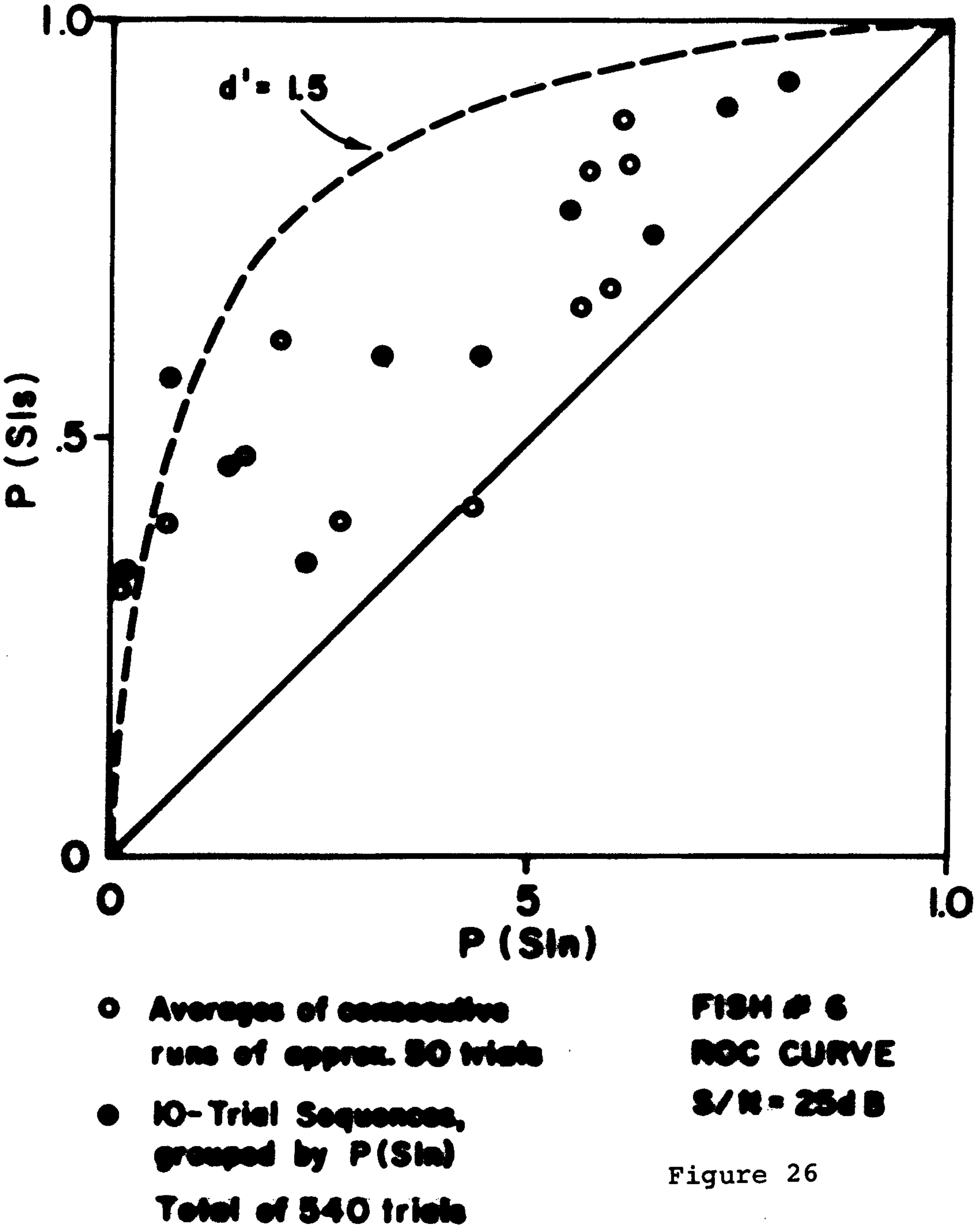

1.0
.5
0
P (S|s)
d' = 1.5
0
.5
1.0
P (S|n)
○ Averages of consecutive runs of approx. 50 trials
● 10-Trial Sequences, grouped by P(S|n)
Total of 540 trials
FISH # 6
ROC CURVE
S/N = 25dB

Figure 26

of a random response varied almost everywhere between the theoretical limits of zero and unity.

e. Fish 8. This fish, together with fish 11 (next reported) was selected for testing whether or not fish could have their criteria varied in a well-defined way, from which an optimum criterion could be calculated, by using a payoff matrix in which the various entries were probabilities of shock (see Figure 1, Payoff Matrix 2). This study was not concerned with learning, but rather the way in which trained fish respond to signal and noise, and learning dynamics were ignored throughout the rest of the study; but in this case, since the learning of a novel payoff matrix was required, it was felt that some attention must be paid to the learning process.

After transfer from the training tanks to the final test enclosure, the fish was confronted with a somewhat novel environment, to which acclimatization might well be necessary. In order to check the way in which fish adjusted to the new environment, 100 trials were run with the same payoff matrix as used in training (i.e. no shock on false-alarms), and his responses noted. These are listed in Table V; first groupings are (as usual) by ten-trial sequences, and then, to smooth out the statistical irregularities to be expected in a small-sample Bernoulli process, 20-trial moving averages of the results were also tabulated. The results of Table V are clear; for the first 50 trials the fish failed to respond at better than the chance level

TABLE V: Fish # 8 Data. 10-Trial Sequences Averaged.

| P(S.s) | P(S/n) | d' | Conditions |
|---|---|---|---|
| (6/11/71-8-1) | | | Learning, No shock on false-alarms, S/N = 43 dB. No gate. |
| .9 | .96 | -.4 | |
| 1.0 | 1.0 | ? | |
| 1.0 | 1.0 | ? | |
| .8 | .84 | -.2 | |
| .3 | .38 | -.2 | |
| .7 | .42 | +.7 | |
| .6 | .40 | +.7 | |
| .6 | .26 | +.5 | |
| .4 | .30 | +.3 | |
| .7 | .34 | +.9 | |
| Same run, 20-trial moving averages; | | | |
| .95 | .98 | -.4 | |
| 1.0 | 1.0 | ? | |
| .90 | .92 | -.1 | |
| .55 | .61 | -.2 | |
| .50 | .40 | +.25 | |
| .65 | .41 | +.6 | |
| .60 | .33 | +.7 | |
| .50 | .28 | +.6 | |
| .55 | .32 | +.6 | |
| Last 50 trials: | | | |
| .60 | .344 | +.65 | $\beta$ = 1.04 ; optimum $\beta$ = zero. |
| (6/13/71-8-2) | | | S/N = 40 dB. No shock on false-alarm for first 10 trials; shock on false-alarm thereafter. |
| 1.0 | 1.0 | ? | |
| 1.0 | 1.0 | ? | |
| 1.0 | 1.0 | ? | |
| 1.0 | 1.0 | ? | |
| 1.0 | 1.0 | ? | |
| 1.0 | .675 | ? | |
| .8 | .500 | +.85 | |
| .9 | .275 | 1.9 | |
| 1.0 | .425 | ? | |
| 1.0 | .675 | ? | |
| 1.0 | .300 | ? | |
| Last 50 trials: | | | |
| .94 | .435 | 1.75 | $\beta$ = 0.31 ; optimum $\beta$ = 6.0 |

(in fact somewhat worse); then abruptly, after the fiftieth trial, the d' index shifted to about 0.65, and remained there throughout the succeeding 50 trials except for statistical fluctuations. The learning of the task in a new environment was then definitely an "Aha!" phenomenon. The criterion during the first 50 trials could scarcely be calculated, since the effective d' over these trials was very slightly negative, but judging from the probabilities of hits and false-alarms, the criterion was extremely loose, roughly zero ("always respond S"). During the second fifty trials, the criterion could readily be calculated, since it remained roughly constant over the 50 trials, as $\beta = 1.04$, on the tight side from the optimal zero (from Green and Swets, 1966, p. 23), ignoring the difficulty of hurdle crossing.

The succeeding 50 trials showed a somewhat similar picture. After ten trials with no shock on false-alarm to enable the subject to readjust to the test vessel, during which he exhibited the optimum criterion of zero, he was suddenly switched to a payoff matrix in which the probability of shock on a miss or false-alarm was unity, and the shock probability for correct rejection or hit was zero (payoff matrix 2 in Figure 1). The results are shown in Table V. For the first 30 trials after switching the matrix, the subject continued to maintain a zero criterion; then, abruptly, the false-alarm probability dropped, and for the next sixty trials both the hit and false-alarm probabilities remained roughly constant, with statistical fluctuations. The d' value over the last 50 trials jumped to about 1.75 over that of 0.65 experienced with the "no shock on false-alarm" payoff matrix. The criterion, calculated over the last fifty trials, dropped somewhat from the 1.04 calculated from the last fifty trials with Payoff Matrix 1, even though the optimum criterion had sharply increased from zero to a very stiff value of 6.00.

Further tests with this animal would have been exciting; unfortunately, due to his highly suboptimal choice of criterion, he had experienced hundreds of shocks, and showed definite signs of disorientation in the holding tank (having difficulty in holding a vertical position) and although no signs of tetany were evident, testing was discontinued.

f. Fish 11. Like Fish 8, this fish was scheduled for testing for adaptability to a payoff matrix in which the various entries were the same shock probabilities as for Fish 8. Results of this test series are summarized in Table VI; a set of fifty trials was run to establish a d' value for the test level of 40 dB S/N; this value was surprisingly low, being only d' = 1.2. The criterion adopted by this fish for the payoff matrix with no shock on false-alarms was 0.50. Next, a set of fifty trials was run with the payoff matrix set for shock on false-alarms. During these fifty trials, the false-alarm rate (instead

**TABLE VI**: Fish # 11 Data. 10-Trial Sequences Averaged.

S/N = 40 dB; Shock Payoff Matrix Trials.

| P(S/s) | P(S/n) | Conditions |
|---|---|---|
| (6/14/71-11-1) | | No shock on false-alarms. With gate. |
| .9 | .400 | |
| 1.0 | .700 | |
| .9 | .400 | |
| .8 | .500 | |
| .8 | .450 | |
| .88 | .490 | d' = 1.2; = 0.50 |
| (6/16/71-11-4) | | Shock on false alarms. No gate. |
| .8 | .275 | |
| .5 | .425 | |
| .8 | .925 | |
| 1.0 | .800 | |
| 1.0 | .900 | |
| .82 | .665 | d' = 0.5; = 0.71 |
| (6/21/71-11-1) | | Shock on false-alarms. No gate. |
| .9 | .575 | |
| .4 | .625 | |
| 1.0 | .850 | |
| 1.0 | .950 | |
| .7 | .875 | |
| .80 | .775 | d' = 0.1; = 0.93 |
| (Same run) | | No shock on false-alarms. No gate. |
| .7 | .400 | |
| 1.0 | .650 | |
| .85 | .525 | d' = 0.95; = 0.59 |

**TABLE VI Cont'd:** 10-Trial Sequences Averaged.

S/N = 30 dB. No shock on false-alarms; check for d'.

| P(S/s) | P(S/n) | Conditions |
|---|---|---|
| (6/26/71-11-1) | | $\bar{s}$ gate |
| 1.0 | .975 | |
| .9 | .825 | |
| .9 | .725 | |
| 1.0 | .825 | |
| .95 | .838 | |
| (6/26/71-11-2) | | $\bar{c}$ gate |
| .6 | .375 | |
| .5 | .650 | |
| .9 | .675 | |
| .7 | .575 | |
| .8 | .725 | |
| .70 | .605 | |
| (6/26/71-11/2) | | Cont'd. |
| .8 | .700 | |
| .9 | .800 | |
| .9 | .825 | |
| .867 | .775 | |

| P(S/s) | P(S/n) | Cond's |
|---|---|---|
| (6/27/71-11-1) | | $\bar{c}$ gate |
| .8 | .475 | |
| .5 | .625 | |
| .9 | .575 | |
| .6 | .425 | |
| .4 | .225 | |
| .64 | .465 | |
| (6/27/71-11-1) | | Cont'd. |
| .5 | .125 | |
| .3 | .175 | |
| .6 | .350 | |
| .3 | .225 | |
| .425 | .219 | |

**TABLE VI Cont'd:** Data Grouped by P(S/n).

S/N = 30 dB.

| $0 \leq P(S/n) < .1$ | | $.1 \leq P(S/n) < .2$ | | $.2 \leq P(S/n) < .3$ | |
|---|---|---|---|---|---|
| P(S/s) | P(S/n) | P(S/s) | P(S/n) | P(S/s) | P(S/n) |
| None | | .5 | .125 | .4 | .225 |
| | | .3 | .175 | .3 | .225 |
| | | .4 | .150 | .35 | .225 |

| $.3 \leq P(S/n) < .4$ | | $.4 \leq P(S/n) < .5$ | | $.5 \leq P(S/n) < .6$ | |
|---|---|---|---|---|---|
| .6 | .375 | .8 | .475 | .7 | .575 |
| .6 | .350 | .6 | .425 | .9 | .575 |
| .60 | .362 | .7 | .450 | .8 | .575 |

| $.6 \leq P(S/n) < .7$ | | $.7 \leq P(S/n) < .8$ | | $.8 \leq P(S/n) < .9$ | |
|---|---|---|---|---|---|
| .5 | .650 | .9 | .725 | .9 | .825 |
| .5 | .625 | .9 | .750 | 1.0 | .825 |
| .8 | .675 | .8 | .725 | .9 | .800 |
| .60 | .650 | .867 | .733 | .9 | .825 |
| | | | | .925 | .819 |

| $.9 \leq P(S/n) \leq 1.0$ | |
|---|---|
| 1.0 | .975 |
| 1.0 | .975 |

of decreasing, as expected from the new payoff matrix) actually increased; a d' value of 0.5 was achieved, lower than the value for the first payoff matrix; but the criterion, instead of increasing toward the optimal value of 6.00, was only 0.71, and (considering the dynamics of the fishes' behavior) was steadily decreasing.

Testing was discontinued for five days, and then the fishes' performance was rechecked, this time giving the first fifty trials with the payoff matrix set for shock on false-alarms. Fifty trials produced a d' value of only 0.1, about the chance level, with a criterion of 0.93, still far away from the optimal 6.00. A succeeding run of 20 trials without shock on false-alarms produced a d' value of 0.95, better than before, with a loose criterion of 0.58.

All this was viewed with some suspicion, since the possibility of d' reduction by random responses, either spontaneous or induced by the payoff matrix schedule, could not be ignored. As a check, the S/N ratio was lowered to 30 dB, a 10 dB reduction, which (considering the steepness of the psychometric function measured on Fish 3) should have reduced the sometimes marginal d' values found at S/N = 40 dB to nearly zero. The results are tabulated in Table VI, and plotted in Figure 27. No shock was delivered on false-alarms. The gate was used to achieve a range of false-alarm ratios, in order to obtain the ROC curve of Figure 27. It is apparent that

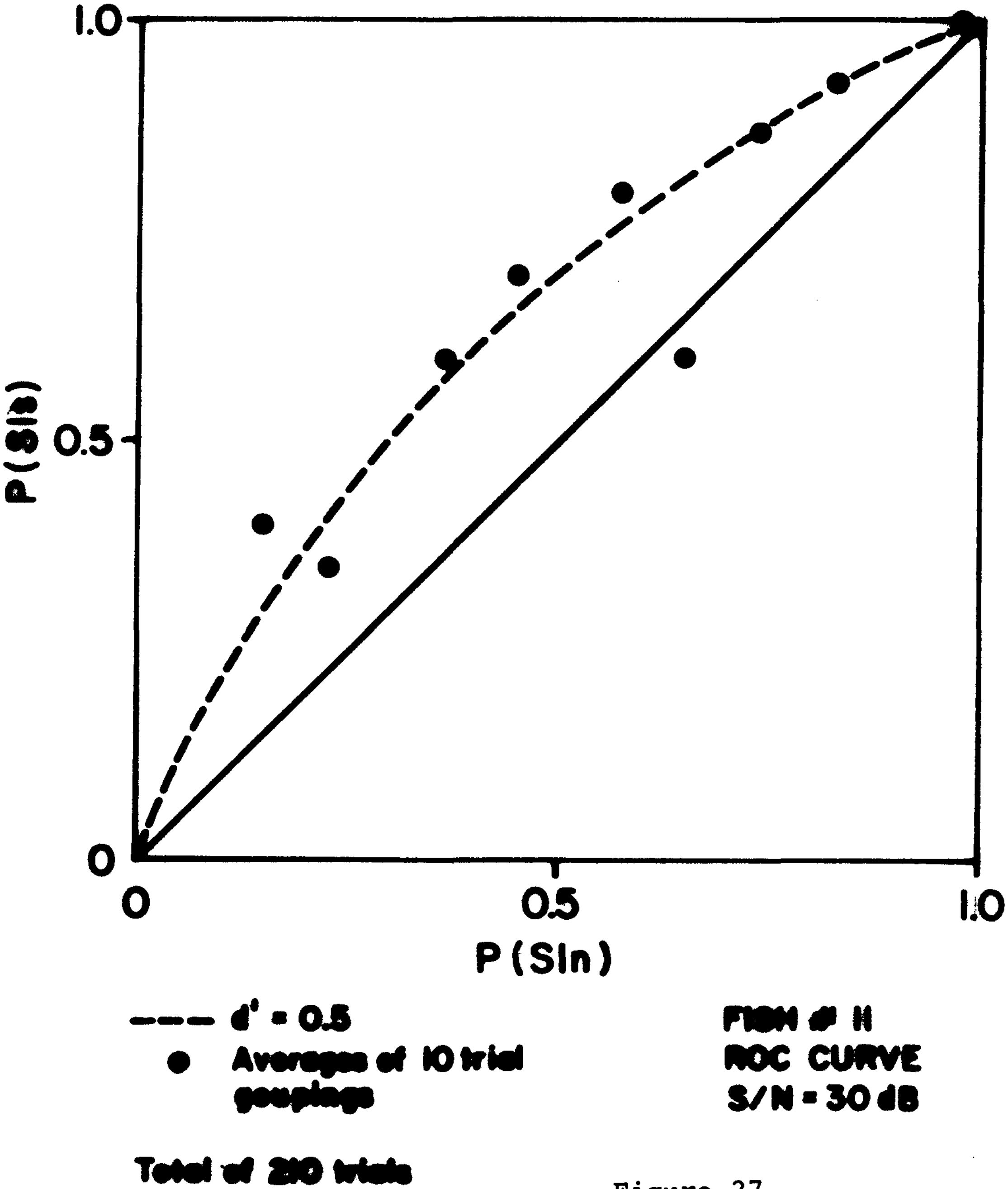

1.0
P(S|s)
0.5
0
0
0.5
1.0
P(S|n)
d' = 0.5
● Averages of 10 trial groupings
Total of 210 trials
FISH # 11
ROC CURVE
S/N = 30 dB

Figure 27

the d' value is low, but definitely above zero; the "eye fit" value of d' = 0.5, with one outlier point, is probably not too far off.

The performances of Fishes 8 and 11 with the payoff matrix set to deliver shock on false-alarms then produced similar results, in that both fish moved their criterion in the opposite direction from that expected from statistical decision theory. However, while Fish 8 showed a higher d' value with the false-alarm-shock matrix, Fish 11 showed a lower value; in the light of the unexpectedly high d' value exhibited by Fish 11 at a 10-dB lowered S/N ratio, this may have been due to an effect of the experimental conditions on random activity like that exhibited by Fishes 3 and 6.

g. Summary of experimental results and analysis. From the eye-fitted ROC curves, it appears that Fishes 2 and 12 were straightforward equal-variance signal detectors, with criterion a random variable for a given payoff matrix; Fish 3 was basically also a signal detector, but with the addition of random responses as described by the Heinemann model, except that the probability of random response was itself a random variable with limited range. Fish 6, highly erratic in performance, generated data which can only be accounted for (among the models tested) by the Heinemann model, with the random-response probability varying over a very wide range indeed.

To check the differences in fit between the high-threshold, equal-variance and small-group models, the calculation shown in Table VII (and described in the Methods section b) were carried out for Fishes 2 and 12, the straight signal detectors. Digital-computer-determined best fit parameter values differ only slightly from the "eye-fit" values, and standard errors of the computer values are also given. Variances of the differences between observed and calculated P(S/s) are also given for the various models. F-ratios were calculated as measures of relative goodness of fit for the various models. For both fish, the fit of the equal-variance model was significantly better than that of the high-threshold model. The fit afforded by the small-group model was better still, significantly so fer Fish 12, better but not significantly so for Fish 2.

Attempts to use the payoff matrix with shock entries for both misses and false-alarms failed, giving a negative result.

**TABLE VII:** Best-Fit Parameter Values, Variances of Data from Models, and F-Ratios for Various Models.

| Model | Best-Fit Parameter Value[1,2] | Variance of Data from Model[3] | F(H-T)[4] | F(E-V)[4] |
|---|---|---|---|---|
| **Fish Number 12** | | | DF = (7,7) | |
| High-Threshold | T=0.351 ± 0.062 | 0.0180 | 1. | |
| Equal Variance | d'=1.17 ± 0.09 | 0.00448 | 4.02* | 1. |
| Small-Group | d'=1.09 | 0.000690 | 26.1* | 6.49* |
| **Fish Number 2** | | | DF = (9,9) | |
| High Threshold | T=0.323 ± 0.053 | 0.0116 | 1. | |
| Equal-Variance | d'=0.866 ± 0.057 | 0.00229 | 5.06* | 1. |
| Small-Group | d'=0.829 ± 0.039 | 0.00110 | 10.5* | 2.08 |

* Significant at 95% confidence level.

1. Standard errors computed using linear model theory.
2. T = hit probability at zero false-alarm probability.
3. Data are from ten-trial sequences grouped by P)S/n).
4. F(H-T) = variance ratio compared to High-Threshold model; F(E-V) = variance ratio compared to Equal-Variance model.

## D. DISCUSSION

### I. Threshold or Signal Detectors?

There seems little doubt that in view of the success of the simple equal-variance Gaussian signal-detection model, with the addition of the Heinemann modification for random activity and the "small-group" correction for varying criteria, that fish function basically as signal detectors. True, they are not as simple signal detectors as humans seem to be: criteria chosen were often far from optimal and sometimes varied considerably, and two of the fish exhibited random responses; these seem to be, however, complications introduced on a basic model which is signal detection, not threshold.

The most convincing data are those for Fishes 2 and 12. Here random activity was absent, and the data were precise enough not only to make a clear goodness-of-fit difference between signal-detection and high-threshold models, but also to show a clearly superior fit for the "small-group" model over the equal-variance signal-detection model.

The data for Fish 3 are not as unambiguous as those for Fishes 2 and 12; it is scarcely possible to do a simple "least-squares" fit of a model when two of the parameters, criterion and random-response probability, are random variables. Fortunately, the model prediction when the random-response probability varies between fairly narrow limits, as is the case with this fish, is that all of the data should be included between two curves with well-defined specific characteristics; the fit of the Heinemann model here carries some conviction, the more so since the model was derived before the data were taken, thus acquiring predictive value.

The data for Fish 6 would make little sense in the absence of the data for Fish 3. True, only the Heinemann model can account for such data, unless extremely high and variable errors were present in the data; that seems unlikely, in view of the precision with which the data for Fishes 2 and 12 are fitted by the theoretical curves, indicating that the procedures for fish testing and data collection were capable of yielding data with little error. With Fish 3 to fill the gap between Fishes 2 and 12, with virtually no random responses, and Fish 6, with random-response probabilities ranging nearly to unity, a pattern of data emerges which encompasses all of the ROC curves obtained.

The one model of those tested which is capable of describing all the ROC curve data obtained for Fishes 2,3, 6,11 and 12, is the equal-variance signal-detection model, with varying criterion for a constant payoff matrix, and with superimposed random-response probability which is itself a random variable, distributed roughly uniformly between zero and some upper limit peculiar to each fish. This upper limit may vary from zero to unity between fish.

No pattern was discerned for the session-to-session variations in random-response rate. There does seem to be an overall pattern for criterion variation, shown by the Fish 3 data plotted in Figure 9; in general, the criterion seems to vary from loose to tight. No reason for this is here adduced. Max A. Woodbury (personal communication), reviewing these data, suggests that by using the method of payoff matrix entry evaluation described in Section III of this Discussion such patterned criterion variation may be approached as a learning process, during which the fish passes from one state to another, each state representing a different strategy.

There are, of course, more complex threshold models than the simple high-threshold model referred to above. Luce's two-state low-threshold model, at the expense of one more parameter than the equal-variance signal-detection model, yields ROC curves somewhat close to those here reported than the high-threshold model; but in spite of its loss of parsimony, it does not fit the data. While one can obviously do better with two straight-line segments than with one, the observed ROC curves are not straight-line segments, but are curvilinear. By extending the number of thresholds involved, one can come up with as complex an ROC curve as desired, literally shaped to one's heart's desire. While such models cannot be disproved, they would seem to be presently ruled out by Occam's razor. Certainly, from the viewpoint of the working biologist anxious to obtain economical sensitivity indices, the simplest model which works is desirable.

There remain for consideration the neural-quantum models. In addition to the loss of parsimony which such models entail, there exists definite evidence that auditory stimuli, when sensation is internally limited, compete with internal auditory stimuli, rather than with synaptic thresholds. Saxon and Dahle (1970) have shown that auditory thresholds increase as heart rate goes up, and decrease as heart rate goes down; further, during the heart cycle itself, Saxon (1970) has shown that during four phases of the cardiac cycle, the threshold is a maximum during QRS. This strongly suggests that internal noise due to blood flow is the limiting factor here, rather than neural events.

For fish, then, the most desirable point of departure in sensory response models seems to be the equal-variance Gaussian signal-detection model, with provision for random responses and variable criteria. The simple high-threshold models do not fit the data. More complex threshold models are not justified biologically, and entail gross loss of parsimony.

Having said that none of the above threshold models is tenable, let us return to Fechner's definition: the threshold is that liminal value of the stimulus at which sensation vanishes.

Replacement of Fechner's definition by Guilford's

50% detectibility point was based largely on methodological grounds. The various threshold models earlier discussed were largely attempts to rationalize the fact that stimuli sometimes fail to evoke responses, and responses sometimes occur without apparent stimuli. These models, like the signal-detection models, are primarily concerned with the description and measurement of "sensation", rather than with the basic threshold concept itself, and are in essence concerned with questions raised by methodology. Swets' (1961) challenge to threshold theory, being based on the success of signal-detection theory in rationalizing the incidence of hits and false-alarms (as compared to equally parsimonious threshold models) also deals with questions of definition and measurement of sensation, and not with the basic threshold concept itself.

If signal-detection theory is accepted and the "threshold" theories rejected, we have simply substituted one index of sensation for another; the threshold concept can now be re-examined, using d' as an index of sensation. We may now ask -- "Is there some non-zero stimulus value at which d' vanishes?" This question has not been examined here; fundamentally, the subject of inquiry has not been threshold theory, but threshold methodology.

## II. Comparison of Man, Pigeon and Fish as Signal Detectors.

Judging from the many smooth ROC curves in Green and Swets (1966), man seems to function as a strict signal detector, with few or no random responses, and good criterion stability. The data for actual and optimal criteria indicate that man's choice of criterion is at least monotonic with the theoretical optimum, yielding payoffs which differ little from the optimal (Green and Swets, 1966, p. 92: Green, 1960). That a subjective element may be present in human decision making is obvious to any gambler; but, to quote Green and Swets (1966, p. 93) "Thus, the well-known fact that human decision makers commonly introduce a subjective transformation of real probabilities and of real decision values is not necessarily salient in the detection setting." However, the smoothness of the ROC curves in Green and Swets (1966) is not necessarily evidence of good criterion stability; many of the ROC curves are obtained by the rating method, which automatically guarantees monotonicity of the ROC curves.

For pigeons, the data of Blough (1967) and Hobson (1970) indicate that the pigeon may perform much like man in signal detection. However, Heinemann, Avin, Sullivan and Chase (1969) found that pigeons may exhibit an appreciable amount of random activity, which remains (for a given set of experimental conditions) fairly constant, and which increases with increasing difficulty of task.

For fish, the work reported here indicates that the random activity reported in pigeons by Heinemann et al (1969) may be extreme or nearly zero, and that its probability is itself a random variable. Similarly, the criterion employed by the fish could be fairly constant or could be a rapidly-varying random variable. There is a slight hint that the random-response probability may, as in the pigeon, sometimes be related to the difficulty of the task; the few points for Fish 3 taken at S/N = 30 dB showed no random activity, and those taken at 25 dB showed quite appreciable random activity. This problem, of random activity when the probability of a random response is itself a random variable, is difficult; possibly it is best handled, for sensitivity measurements, by careful training and selection of subjects. Although it may be avoided, this variable random-response rate is apparently a behaviorial characteristic of the goldfish. The second problem, that of rapid criterion variation, would appear to be satisfactorily handled by the "small-group" method here presented, with appropriate modification to the shape of the ROC curves to be fitted. The objection which might be raised to the formulae as presented, the uniform prior distribution for P(S/n), can be mitigated by choosing a sufficiently long sequence (if the criterion does not vary too quickly) or by including in the calculations the observed distribution of false-alarms.

Tavolga (1970) has given six levels of interaction among organisms and their environment; vegetative, tonic, phasic, signal, symbolic and verbal. In auditory learned tasks, function would seem to be at Level 4 or higher.

Man is capable of function at the highest Level 6. Both bird and fish are capable of function at Level 4, but not Level 5; there are, however, behavioral as well as anatomical differences between them of a qualitative nature (Bitterman, 1965; Behrend, Powers and Bitterman, 1970). The tests reported here also show that fish differ in some respects from pigeons in their detection responses.

With some reservations, particularly in regard to criterion choice, the following appears to characterize the signal-detection performance of man, pigeon and fish:

Man. Relatively stable criteria, with payoffs not far from optimal; little or no random responses.

Pigeon. Relatively small criterion variability; random responses may be present with stable probability related to task difficulty.

Fish. Variable and sometimes highly suboptimal criteria; random responses may be present, with variable probability.

## III. Implications for Information Processing by the Subject.

It is difficult to draw profound conclusions from lab experiments such as these about the way in which an animal normally processes complex signal and noise stimuli in his natural environment to enhance survival. Even that foremost exponent of species interchangeability and generalization from lab tests, B. F. Skinner, has said (1971): "It seems clear that a feature of the environment can be conducted into a stimulus that elicits responses characteristic of the phylogenetic endowment of the species... the effect is quite different from operant conditioning, even though both processes generate responses having similar topographies."

Konishi (1971) briefly reviews the status of the synthesis of behavioral and neurophysiological outlooks; he concludes that a new science, Neurobiology, is needed adequately to deal with the questions to be answered. Nevertheless, some speculation is possible.

First, the latencies in response involved, and the time to learn the task, indicate that high centers in the brain are involved. This in turn suggests the speculation that the differences we have just noted between man, pigeon and fish may be phylogenetic in origin, and related to differences in brain structure and function. Of interest here is the increasingly evident random activity as one descends the phylogenetic scale.

Secondly, the inability of the fish tested to learn the more complex payoff matrix with shock on false-alarms <u>may</u> be associated with their relative lack of cerebral cortex, as may be the occasionally very suboptimal criterion choice. In the conventional avoidance-conditioning scheme, a shock is always something that can be escaped by responding; in the more complex situation with shock on false-alarms, a shock sometimes means "Respond!" and sometimes "Don't Respond!". The fish tested seemed unable to deal with this dual meaning in the time given for testing. Whether they could have learned the task given more time is problematical, since it is doubtful they could have survived many more shocks. Patently, the decision to cross was not made on the basis of optimal payoff.

Even though the payoff matrix for the conventional avoidance paradigm does not lend itself to a direct calculation of the optimal criterion unless one ignores hurdle-crossing difficulty, one can work out <u>post hoc</u> the weight which the subject attaches to hurdle-crossing compared to shock on the basis of optimum criterion calculation. Assuming a penalty of 1 for a miss, i.e. a shock, and a hurdle-crossing penalty of v, we now have Payoff Matrix 3 of Figure 1. The value of v for a given criterion choice can be worked out as:

$$v = \beta /(P(n)/P(s) + \beta )$$

For $P(n)/P(s) = 6$ as in the present tests, and a not-untypical $\beta$ value of 1, we obtain $v = 0.143$, indicating that in this case the fish weights hurdle-crossing difficulty about one-seventh as heavily as shock. Now, consider limiting values. For $\beta = 0$, or an "always cross" criterion, $v = 0$, so that hurdle crossing is viewed as negligible compared to shock. In the other extreme, with $\beta = \infty$, the value of v obtained is 1; if the fish weights hurdle-crossing difficulty equally with shock, a criterion of "never cross" should be (and very occasionally is) adopted.

(The effect of multiple shocks is here ignored.) It is then conceivable that when not confronted with dual meanings for shock, the fish can use rough decision theory in criterion determination; the subjectivity attached to the hurdle-crossing penalty is not unusual even among humans (Green and Swets, 1966, p. 106). In a typical "threshold-testing" situation with $P(n)/P(s) = 6$, hit probability $P(S/s) = 0.5$, and one inter-trial crossing per minute so that $P(S/n) = 0.167$, we obtain a $d'$ value of 1.0; a criterion $\beta$ of 1.6; and a $v$ value of 0.21, indicating that in this typical situation hurdle-crossing is viewed as about one-fifth as much of a problem as shock.

The finding by Hillyard, Squires, Bauer and Lindsay (1971) that the scalp evoked potential in a human subject during a detection task is different for a hit than for a false-alarm, and that evoked potentials for false-alarms, misses and correct rejections are indistinguishable, strikes at the roots of signal detection theory, in terms of the neural processing involved. Critical to signal detection theory is the indistinguishability of hits and false-alarms. This finding has not, at this writing, been replicated; if it is, a lot of re-thinking will have to be done.

## IV. Why Does Threshold Theory Work?

Given that the simple threshold theory on which such test methods as the up-down technique are based is untenable, why can reproducible threshold measurements be achieved by different investigators employed different paradigms? Already cited in the Introduction is the agreement on goldfish audiograms between Jacobs and Tavolga (1967), Fay (1969) and Enger (1967), employing totally different paradigms; even more impressive is the agreement on goldfish frequency discrimination limens between Jacobs and Tavolga (1968) and Fay (1970), again employing very different paradigms. I believe the answer is found in the explicit and implicit restrictions placed on the training and testing procedures, which result in yielding thresholds at very similar d' values for very similar signal-to-noise ratios. Let us consider these restrictions.

First, false-alarm probabilities are reduced by training to below an absolute permitted maximum. In the case of the often-employed method of Tavolga and Wodinsky (1963), this is about one false-alarm per minute with a ten-second CS-US interval, giving a P(S/n) of 0.17 at the 50% threshold hit rate P(S/s) = 0.5. This immediately yields a d' value very close to 1.

Secondly, false-alarm rates very much lower than this are (implicitly) not tolerated, the tests being rejected on some such basis as "the fish just isn't working today". Clearly, as already pointed out by Steiner (1971), false-alarm rates should always be recorded in threshold testing.

Thirdly, the regularity required of the reversals in up-down testing eliminates from testing fish with very variable criteria or with high random-response rates.

Finally, the steepness of the psychometric function helps greatly to achieve reproducible results in lower animal tests, when agreement within 5 dB is considered good.

In short, in spite of the invalidity of the high-threshold theory, test methods based on it may also be valid within signal detection theory, at least to the extent of yielding reproducible results at a d' of about 1; however, this reproducibility may involve the suppression of a great deal of inter- and intra-subject variability. The existence of highly-variable criteria and random-response rates may offer a partial explanation as to why certain animals "don't work".

## V. Summary.

Fish function basically as signal detectors, and not according to simple threshold theory. Threshold methods can nevertheless yield reproducible results, under certain explicit and implicit restrictions normally attached to the threshold method. Fish also exhibit random responses observed in pigeons but not in man; they may also exhibit spontaneously-varying and highly-suboptimal criteria to a much greater extent than humans or pigeons. This is consistent with the position of fish on the phylogenetic scale. Much more work is needed before the neurophysiological correlates of the behavior of fishes in a detection task can be defined with any certainty.

## E. REFERENCES

Atkinson, R. C. 1963. A variable-sensitivity theory of signal detection. Psychol. Rev. 70: 91-106.

Banks, W. P. 1970. Signal detection theory and human memory. Psychol. Bull. 74: 81-99.

Behrend, E. A., A. S. Powers and M. E. Bitterman. 1970. Interference and-forgetting in bird and fish. Science 167: 389-390.

Bekesy, G. von. 1930. Ueber das Fechnersche Gesetz und seine Bedeutung fur die Theorie der akustichen Beobachtungsfehler und die Theorie des Horens. Ann. d. Phys. 7: 329-359.

__________. 1960. Experiments in Hearing (E. G. Wever, ed. and translator). McGraw-Hill, New York.

Bitterman, M. E. 1965. Phyletic differences in learning. Am Psychol. 20: 396-410.

Blackwell, H. R. 1953. Neural theories of simple visual discrimination. Bull. Eng. Res. Univ. Mich. 36.

Blough, D. S. 1967. Stimulus generalization as signal detection in pigeons. Science 158: 940-941.

Boneau, C. A., M. K. Holland and W. M. Baker. 1965. Color discrimination performance of pigeons: effects of reward. Science 149: 448-455.

Cahn, P. H., W. Siler and J. Wodinsky 1969. Acousticolateralis system of fishes: tests of pressure and particle velocity sensitivity in Grunts, Haemulon sciurus and Haemulon parrai. J. Acoust. Soc. Am. 46: 1572-1578.

__________, W. Siler and M. Fujiya. 1972. Sensory factors in response of fish to environmental change. In Response of Fish to Environmental Changes (W. Chavin, ed.). Charles Thomas, Springfield, Ill. In press.

Chernoff, H. 1969. Sequential designs. In The Design of Computer Simulation Experiments (T. H. Taylor, ed.). Duke Univ. Press, Durham, N. C.

Corso, J. F. 1956. The neural quantum theory of sensory discrimination. Psychol. Bull. 53: 371-393.

__________. 1963. A theoretico-historical review of the threshold concept. Psychol. Bull. 68: 371-393.

Enger, P. S. 1967. Effect of the acoustic near field on the sound threshold in fishes. In Lateral Line Detectors (P. H. Cahn, ed.). Indiana Univ. Press, Bloomington, Ind.

Fay, R. 1969. Behavioral audiogram for the goldfish. J. Aud. Res. 9: 112-121.

__________. 1970. Auditory discrimination in the goldfish. J. Comp. Physiol Psychol. 73: 175-180.

Fechner, G. T. 1860. Elemente der Psychophysik. Breitkopf und Hartel, Leipsig, Germany.

Fisher, R. A. 1925. Statistical Methods for Research Workers. Fourteenth edition, 1970: Hafner, Darien, Conn.

Green, D. M. 1960. Psychoacoustics and detection theory. J. Acoust. Soc. Am. 33: 897-903.

__________ and J. A. Swets. 1966. Signal Detection Theory and Psychophysics. John Wiley, New York.

Guilford, J. P. 1954. Psychometric Methods, 2nd. ed. (1st. ed. 1936). McGraw-Hill, New York.

Guttman, N. 1972. Book review: Animal Psychophysics (W. C. Stebbins, ed.). Science 175: 508.

Hack, M. H. 1963. Signal detection in the rat. Science 139: 758-759.

Heinemann, E. G., E. Avin, M. A. Sullivan and S. Chase. 1969. Analysis of stimulus generalization with a psychophysical method. J. Exp. Psychol. 80: 215-224.

Herbart, J. F. 1824. Psychologie als Wissenschaft, neu gegrundet auf Erfahrung, Metaphysik und Mathematik. Unzer, Konigsburg, Germany.

Hillyard, S. A., K. C. Squires, J. W. Bauer and P. H. Lindsay. 1971. Evoked potential correlates of auditory signal detection. Science 172: 1357-1360.

Hobson, S. L. 1970. Discrimination of fixed-ratio schedules by pigeons: sensitivity and bias. Doctoral dissertation. Columbia Univ., New York.

Jacobs, D. W. and W. N. Tavolga. 1967. Acoustic intensity limens in goldfish. An. Behav. 15: 324-335.

Jacobs, D. W. and W. N. Tavolga. 1968. Acoustic frequency limens in goldifsh. An. Behav. 15: 324-335.

Konishi, M. 1971. Ethology and neurobiology. Am. Scientist 59: 56-63.

Leibnitz, G. W. 1714. Principles de la nature et de la grace. Republished, 1890, In The Philosophical Works of Leibnitz (G. M. Duncan, ed. and translator). Tuttle, Morehouse and Taylor, New Haven, Conn.

Luce, R. D. 1963. A threshold theory for simple detection experiments. Psychol Rev. 70: 61-79.

Morgenstern, O. and J. von Neumann. 1953. The Theory of Games and Economic Behavior. Princeton Univ. Press, Princeton, N. J.

Nevin, J. A. 1964. A method for the determination of psycho-physical functions in the rat. J. Exp. Anal. Behav. 77: 169.

__________. 1965. Decision theory in studies of discrimination in animals. Science 150: 1057.

Rilling, M. and C. McDiarmid. 1965. Signal detection in fixed-ratio schedules. Science 148: 526-527.

Saxon, S. A. 1970. Detection of near threshold signals during four phases of cardiac cycle. Alabama J. Med. Sci. 7: 427-430.

__________ and A. J. Dahle. 1970. Auditory thresholds during periods of induced high and low heart rates. Soc. Psychophysio. Res. 8: 23-29.

Shannon, C. E. and W. Weaver. 1949. The Mathematical Theory of Communication. Univ. Ill. Press, Urbana, Ill.

Siler, W. and J. Lubowsky. 1969. A laboratory computer system for on-line processing of physiological data. In Digest for 1969 Computer Group Conference, 173-178. I.E.E.E., New York.

Skinner, B. F. 1971. Autoshaping. Science 173: 752.

Stebbins, W. C. (ed.). 1970. Animal Psychophysics. Appleton-Century-Crofts, New York.

Steiner, G. 1971. Stimulus control of avoidance learning in fish. J. Comp. Physiol. Psychol. 74: 52-58.

Stevens, S. S. 1961a. To honor Fechner and repeal his law. Science 133: 80-86.

Stevens, S. S. 1961b. Is there a quantal threshold? In Sensory Communication (W. A. Rosenblith, ed.). Wiley, New York.

Subowski, M. D. and A. A. Spevack. 1968. Signal detection analysis of instrumental discrimination conditioning. Canad. J. Psychol 22: 26-34.

Swets, J. A., W. P. Tanner Jr. and T. G. Birdsall. 1955. The evidence for a decision-making theory of visual detection. Technical Report No. 40, Electronic Defense Group. University of Michigan.

__________. 1961. Is there a sensory threshold? Science 134: 168-177.

__________,W. P. Tanner Jr. and T. G. Birdsall. 1961. Decision processes in perception. Psychol. Rev. 68: 301-340.

Tanner, W. P. Jr. and J. A. Swets. 1953. A new theory of visual detection. Technical Report No. 18, Electronic Defense Group. University of Michigan.

__________, and J. A. Swets. 1954. A decision-making theory of visual detection. Psychol Rev. 61: 401-409.

__________, J. A. Swets and D. M. Green. 1956. Some general properties of the hearing mechanism. Technical Report No. 30, Electronic Defense Group. University of Michigan.

Tavolga, W. N. and J. Wodinsky. 1963. Auditory capacity in fishes. Bull. Am. Mus. Nat. Hist. 126: 177-240.

__________. 1970. Levels of interaction in animal communication. In Development and Evolution of Behavior: Essays in Memory of T. C. Schneirla. (L. R. Aronson, E. Tobach, D. S. Lehrman and J. S. Rosenblatt, eds.). Freeman, San Francisco.

Thurstone, J. L. 1927. Psychophysical analysis. Am. J. Psychol. 38: 368-389.

Wald, A. 1950. Statistical Decision Functions. Wiley, New York.

Weber, E. H. 1834. De tactu: annotacionices et physiologicae. Leipsig, Germany.

Weiss, B. 1966. Auditory sensitivity in goldfish. J. Aud. Res. 6: 321-335.

Wiener, N. 1949. Extrapolation, Interpolation and Smoothing of Stationary Time Series. M.I.T. Press, Cambridge, Mass.

Woodbury, M. A. 1970. On-line experimental design program. In Picture Processing and Psychopictorics (B. S. Lipkin and A. Rosenfeld, eds.). Academic Press, New York.

Yarczower, M. and R. Switalski. 1969. Stimulus control in the goldfish after massed extinction. J. Exp. Anal.. Behav. 12: 565-570.

## APPENDIX: DERIVATION OF "SMALL-GROUP" MODEL FORMULA

The problem of inferring properties of populations from those of samples has been the subject of controversy for over two centuries. The mathematical theory of "Inverse Probability" was originated by an English mathematician, Thomas Bayes (1702-1761); his main works, including the celebrated theorem which bears his name, appeared in the Philosophical Transactions of 1763 after Bayes' death. A basic proof by Bayes illustrates the method, and also its Achilles heel: if m:n is the relative frequency of an event on n independent occasions, then m:n is also the most probable value of the event's probability, provided that any value of this probability is _a priori_ as probable as any other value. The problem here is the determination of the _a priori_ or prior probabilities; these require a knowledge of that same population about which one wishes to infer properties.

The great mathematician Laplace rederived the above formula in 1774, and made extensive use of the principle. The above assumption that all probabilities are equally probable was more generally stated as the "principle of equal distribution of ignorance" by another great mathematician, Jacques Bernoulli. But R. A. Fisher (1925) harshly rejected the entire argument, saying; "This is not the place to enter into the subleties of a prolonged controversy; it will be sufficient in this general outline of the scope of Statistical Science to reaffirm my personal conviction, which I have sustained elsewhere, that the theory of inverse probability is founded upon an error, and must be wholly rejected." The controversy continues to the present day; it was the author's pleasure to hear, at the 1965 meeting of the International Federation of Information Processing Societies, a paper delivered by R. S. Ledley on "Application of Bayes' Theorem to Medical Diagnosis", followed immediately by M. A. Woodbury's paper on "The Inapplicability of Bayes' Theorem to Medical Diagnosis".

The derivation now to be given of the "Small Group" model for ROC curves employs the principle of equal distribution of ignorance; however, as will be shown, the ROC curves obtained tend to be insensitive to violations of this assumption, the insensitivity increasing with increasing group size. Another point: although Bayes' theorem is classically associated with the theory of inverse probability, it is not here so employed. Fisher (1925) says: "The deduction of inferences respecting samples, from assumptions respecting the populations from which they are drawn, shows us the position in Statistics of the classical Theory of Probability." Bayes theorem is here used in that manner: from assumptions regarding the population of measurements of hit and false-alarms, a conclusion is drawn regarding the expected values of hit and false-alarm ratios for samples drawn from that population, the samples being of finite size and drawn in a specified way.

First, a proof will be given for a relationship needed in the model derivation.

To prove: $$E(Y(X)/z) = \frac{\int Y(X)P(z/X)p(X)\,dX}{\int P(z/X)p(X)\,dX}$$

Consider the discrete case. Then, by Bayes' theorem,

$$P(X_i/z_j) = \frac{P(z_j/X_i)P(X_i)}{\sum_i P(z_j/X_i)P(X_i)}$$

in which X and z are discrete variables. (To simplify notation, drop the z subscript.) Now consider X to be a continuous random variable. Then, using p to denote a probability density function, while P denotes a probability, the counterpart of Bayes' theorem is:

$$p(X/z) = \frac{P(z/X)p(X)}{\int P(z/X)p(X)\,dX}$$

By definition,

$$E(Y/z) = \int Yp(Y/z)\,dY$$

and, if $Y = Y(X)$,

$$p(X) = p(Y)\ dY/dX$$

Now, employing the continuous version of Bayes' theorem given above,

$$E(Y/z) = \int \frac{Y\ P(z/X)\ p(X)\ dX}{\int P(z/X)\ /(X)\ dX}$$

Since the integral in the denominator is a definite integral, being carried out over the range of X, it is a constant, and the last equation simplifies to:

$$E(Y/z) = \frac{\int Y(X)\ P(z/X)\ p(X)\ dX}{\int P(x/X)\ p(X)\ dX} \qquad (1)$$

the desired result.

We now wish to apply (1) to the problem of determination of ROC curves experimentally when the subject's criterion is not constant.

Given a large number of trials of a subject, with hit and false-alarm probabilities observed to vary widely; and given that these trials are grouped in small sequences by the observed false-alarm probabilities P(S/n), the corresponding hit probabilities P(S/s) being averaged, we desire the expected values of P(S/s) for the various false-alarm groupings, assuming the Gaussian signal-detection model to hold.

To define symbols:

X = true (and unknown) P(S/n), a random variable.
z = observed P(S/n), a constant for any one group.
Y(X) = the hit probability P(S/s) corresponding to the false-alarm probability X under the equal-variance Gaussian model.

$P(z/X)$ = conditional probability of observing z given the true false-alarm probability X.

$p(X)$ = prior probability density function for the true false-alarm probability X.

With these definitions, formula (1) above will yield the desired expectation for the hit probabilities. However, as in many Bayesian procedures, the prior density function p(X) is unknown. Lacking any information about p(X), let us assume it uniform, i.e. $p(X) = 1$. Formula (1) then simplifies to:

$$E(Y(X)/z) = \frac{\int Y(X)\ P(z/X)\ dX}{\int P(z/X)\ dX} \tag{2}$$

Y(X) follows from the equal-variance Gaussian signal-detection model, given only the sensitivity index d' .

To obtain P(z/X), we assume that the subject has M opportunities to false-alarm within the sequence grouped, and that N false-alarms are actually observed. Then:

$$z = N/M = P(S/n)$$

and, from straightforward probability theory,

$$P(z/X) = \binom{M}{N} X^{N} (1-X)^{M-N} \quad . \tag{3}$$

The evaluation of the expectations given by formula (2) can be carried out by digital computer, with results as shown in Figures 10, 11 and 12 of the text.

The arbitrary nature of the prior distribution assumed for P(S/n) casts some doubt on the validity of the above formulation. Fortunately, the behavior of P(z/X) approaches a Dirac delta function as the number of trials N included in the sequence from which z is determined increases without limit, and even for a moderate number of trials sufficient to give reasonably good resolution for z (such as 40, used throughout this work), P(z/X) is fairly peaky. Measure of this are given in Table A-1 below. In this table are given values for the width at one-half height of the curve of P(z/X) plotted against X.

### TABLE A-1

Widths at 1/2 height of P(z/X) vs. X curves.

| | Number of trials N in sequence: | | | | |
|---|---|---|---|---|---|
| | 4 | 10 | 20 | 40 | 100 |
| Minimum: | 0.16 | 0.07 | 0.04 | 0.02 | 0.01 |
| Maximum: | 0.54 | 0.36 | 0.26 | 0.19 | 0.12 |
| Average: | 0.36 | 0.27 | 0.16 | 0.12 | 0.07 |

In the absence of knowledge about the prior distribution of p(X), then, the "small group" model is useful when the false-alarm rate varies (the criterion of the subject not being stable), and sufficient trials are grouped in a sequence for determining the observed false-alarm rate z so that the peakiness of the P(z/X) curve makes the formulas insensitive to p(X). Forty opportunities for false-alarms, as used in this work, seems a reasonable compromise value, especially since at low values for z, where the greatest variation in criterion was observed, the width at half height of the P(z/X) curve is a minimum.

## REFERENCE

Fisher, R. A. (1925). Statistical Methods for Research Workers. Hafner, Darien, Conn. Fourteenth Edition, pp. 9 and 10.