08901-1-T

HUMAN PERFORMANCE CENTER

DEPARTMENT OF PSYCHOLOGY

The University of Michigan, Ann Arbor

A Response Bias Explanation of Conservative Human Inference

WESLEY MICHAEL DU CHARME

Technical Report No. 19
December 1969

THE HUMAN PERFORMANCE CENTER

DEPARTMENT OF PSYCHOLOGY

The Human Performance Center is a federation of research programs whose emphasis is on man as a processor of information. Topics under study include perception, attention, verbal learning and behavior, short- and long-term memory, choice and decision processes, and learning and performance in simple and complex skills. The integrating concept is the quantitative description, and theory, of man's performance capabilities and limitations and the ways in which these may be modified by learning, by instruction, and by task design.

The Center issues two series of reports. A Technical Report series includes original reports of experimental or theoretical studies, and integrative reviews of the scientific literature. A Memorandum Report series includes printed versions of papers presented orally at scientific or professional meetings or symposia, methodological notes and documentary materials, apparatus notes, and exploratory studies.

THE UNIVERSITY OF MICHIGAN

COLLEGE OF LITERATURE, SCIENCE AND THE ARTS

DEPARTMENT OF PSYCHOLOGY

A RESPONSE BIAS EXPLANATION OF CONSERVATIVE HUMAN INFERENCE

Wesley Michael Du Charme

HUMAN PERFORMANCE CENTER--TECHNICAL REPORT NO. 19

December, 1969

This research was supported by the National Aeronautics and Space Administration under Grant NGR 23-005-171 to the Engineering Psychology Laboratory, University of Michigan, monitored by the Ames Research Center, NASA.

PREFACE

This report is an independent contribution to the program of research of the Engineering Psychology Laboratory. It was supported by the National Aeronautics and Space Administration under Grant NGR 23-005-171 to the Engineering Psychology Laboratory, University of Michigan, monitored by the Ames Research Center, NASA.

This report was also a dissertation submitted by the author in partial fulfillment of the degree of Doctor of Philosophy (Psychology) in the University of Michigan, 1969. The doctoral dissertation committee was: Associate Professor Cameron R. Peterson, Chairman; Associate Professor Wilfred Kincaid; Associate Professor Richard W. Pew; and Associate Professor Daniel J. Weintraub. I wish to express my sincere gratitude to Dr. Cameron R. Peterson for his warm encouragement and valuable advice not only on this paper but throughout my graduate career.

TABLE OF CONTENTS

ABSTRACT

Conservative human inference has been attributed to misperception or misaggregation of data, but it may be caused by response biases. In the present experiments subjects revised odds estimates about which of two normal distribution data generators were being sampled. An analysis of special sequences and a plot of revised odds against theoretical odds in Experiment I showed a bias in subjects' response functions. They revised odds optimally only over a range of $\pm$ 1.0 log odds. When the experimenter set different levels of prior odds, subjects shifted their response functions so that the optimal range centered around the set prior odds. A second experiment showed that the biased functions remained invariate over changes in data generator familiarity and diagnosticity. Subjects were biased over either cumulative evidence impact or the number system, but within their optimal range they neither misaggregated nor misperceived data.

INTRODUCTION

Of the many kinds of errors men are known to make, one particular kind has been the focus of a considerable number of experiments in the field of decision theory. The error shows up when human opinion revisions are compared to the revisions of the optimal model, Bayes's theorem. The comparison typically shows that humans are conservative, they do not revise their opinions, or probabilities, sufficiently in the light of new information. They change in the right direction, but not enough.

The fact that people revise probabilities in a conservative fashion is of much interest on both theoretical and practical grounds. Theoretically there are two basic questions a person must answer when making any decision; "what's at stake?" and "what are the odds?" In many decision situations the probabilities, or odds, over the various states of the world change as new information comes in. If people cannot accurately keep up with these changes, they will make suboptimal decisions. In practical applications of decision theoretic techniques it has been suggested that men should provide the probabilistic inputs to real world diagnostic systems, e.g., military intelligence systems. But if men's probability revisions are in error, the system may arrive at faulty diagnoses. For these reasons much attention has been directed towards determining the nature of, and causes for, conservative human inference.

Previous Explanations of Conservatism

Of the several explanations of conservatism which have been put forward (for a review see Du Charme, 1969), two have received the most attention. Edwards (1968) has discussed the evidence for these two explanations called misaggregation, and misperception, and some of this presentation will be drawn from Edwards' paper. A brief look at Bayes's theorem will make the discussion somewhat easier to follow. The theorem states that the probability P of a hypothesis H given the occurrence of a datum D is

$$P(H|D) = \frac{P(D|H)\ P(H)}{P(D)} \quad . \tag{1}$$

If we have two mutually exclusive hypotheses H_1 and H_2, and we write out Equation 1 for each hypothesis and then divide one equation by the other the result is:

$$\frac{P(H_1|D)}{P(H_2|D)} = \frac{P(D|H_1)\ P(H_1)}{P(D|H_2)\ P(H_2)} \quad . \tag{2}$$

Equation 2 can be rewritten as

$$\Omega_1 = L\ \ \Omega_0 \tag{3}$$

where the prior odds Ω_0 are equal to $\frac{P(H_1)}{P(H_2)}$, the likelihood ratio L is equal to $\frac{P(D|H_1)}{P(D|H_2)}$ and the posterior odds Ω_1 are equal to $\frac{P(H_1|D)}{P(H_2|D)}$.

In typical probability revision experiments the two hypotheses are binomial populations, and the data consist of binomial events drawn from one or the

other of the populations.

Both the misperception and the misaggregation hypotheses assume that people deviate from the Bayesian model for opinion revision. The misperception hypothesis says that people always fail to use the correct likelihood ratio (L in Equation 3) when they revise their odds or probabilities. They misperceive the diagnostic impact of the datum they are processing. The hypothesis further asserts that these incorrect likelihood ratios are always properly combined or aggregated, i.e., multiplied by the prior odds according to Bayes's theorem. The misaggregation hypothesis asserts just the opposite. It states that people perceive diagnostic impact accurately, the likelihood ratios they use are correct, but that they always fail to properly combine the likelihood ratios with the prior odds. They never aggregate the data and the prior odds according to Bayes's theorem, i.e., multiplicatively. Although these two hypotheses about the cause of conservatism are quite different, it may be, as Edwards (1968) suggests, that both are true. Subjects may both misperceive and misaggregate data. What does the evidence show?

Beach has championed the misperception cause and carried out several experiments testing it. He and others have found a variety of evidence supporting the hypothesis including successful checks of the internal consistency of revised probabilities (Beach, 1966; Peterson, Ulehla, Miller, Bourne & Stilson, 1965); demonstrations that subjective sampling distributions are too flat as predicted by the hypothesis (Peterson, Du Charme & Edwards, 1968; Vlek & Beintema, 1967; Vlek & Van der Heijden, 1967; Wheeler & Beach, 1968); and data which show that the accuracy of Ss' probability revisions increases as their sampling distributions become more accurate (Wheeler & Beach, 1968).

Unfortunately, most of these supporting data can be criticized (as Edwards, 1968 has done) on the grounds that the Ss were dealing with already aggregated samples. For example the experiments generally show that given the Ss' $P(D|H)$ estimates his $P(H|D)$ estimates are those predicted by Bayes's theorem; he correctly aggregates his erroneous likelihood ratios. But the experimenters typically gather $P(D|H)$ estimates for already aggregated samples, e.g., $P(D_1, D_2, D_3|H)$ rather than for individual data, e.g., $P(D_1|H)$, $P(D_2|H)$, $P(D_3|H)$. Thus what the experiments really show, according to Edwards, is that equal amounts of conservatism exist in the two kinds of aggregation, $P(D_i|H)$ and $P(H|D_i)$, which Ss perform.

The misperception hypothesis can also be interpreted to mean that Ss assume different data generators than E in fact uses. Different data generators will lead to different likelihood ratios. Several experiments support the hypothesis that Ss do sometimes assume different data generators (Beach, 1968; Lichtenstein & Feeney, 1968; Vlek & Van der Heijden, 1969).

Support for the misaggregation hypothesis comes from a number of experiments. The evidence consists of having Ss estimate both likelihood ratios and posterior odds (or probabilities), and then showing that a Bayesian aggregation of the estimated likelihood ratios is more optimal than the Ss' own aggregations (Edwards, Phillips, Hays, & Goodman, 1968; Goldstein, Southard & Schum, 1967; Gustafson, 1969; Kaplan & Newman, 1966; Phillips, 1966; Schneider, 1965; Schum, Southard & Wombolt, 1969; Wheeler[1]). Since these

[1]Gloria Wheeler, Personal communication, 1969.

studies took place in settings ranging from simulations of large scale military intelligence systems to medical diagnosis and abstract laboratory tasks, their unanimity is impressive. Misaggregation, either alone or in combination with misperception, would seem to constitute a good explanation of conservatism in human inference. But does it? Certain findings have been made in probability revision experiments which do not seem readily explicable in terms of misaggregation or misperception.

Some Unexplained Facts

There is a whole class of probability revision experiments where findings of conservatism are not made. Whenever Ss' estimates have been gathered for continuous, or near continuous, hypotheses the estimates have been very nearly optimal (Du Charme & Peterson, 1969; Edwards[2]; Peterson & Phillips, 1966). In these three experiments the Ss were faced with a binomial data generator whose unknown proportion could range from 0.0 to 1.0. Their task was to look at a sequence of data generated by this unknown binomial population and to estimate either the central 33% credible interval (Du Charme & Peterson, 1969; Peterson & Phillips, 1966) or the entire probability distribution over the 0.0 to 1.0 range (Edwards, see footnote 2). In all three cases the Ss' estimates were either very close to optimal or extreme. Neither the misperception nor the misaggregation hypothesis give any insight into why Ss perform so nearly optimally in these tasks.

No one has spelled out the misaggregation or misperception hypotheses full enough to account for the fact that the amount of conservatism

[2]Ward Edwards, Personal Communication, 1969.

in Ss' estimates varies as a function of the diagnosticity of the data generators being used (Peterson, Du Charme & Edwards, 1968; Peterson & Miller, 1965; Phillips & Edwards, 1966). That is, no one has explained why conservatism varies with the theoretical value of L in the situation. One has to postulate that Ss misaggregate more, or misperceive more, with diagnostic data than with undiagnostic. Neither argument is very compelling given the nature of the explanations. Even more basically an argument of misaggregation or misperception could just as well lead one to expect Ss' estimates to be too extreme as too conservative. There is nothing in the formulation of either hypothesis which forces a prediction of conservatism.

The inertia effect constitutes another source of confusion in this area. Several experiments have shown that Ss change their estimates more readily when a datum confirms the currently favored hypothesis than when it disconfirms this hypothesis (Geller & Pitz, 1968; Peterson & Du Charme, 1967; Pitz, Downing & Reinhold, 1967). But the reverse of an inertia effect has also been found (Du Charme & Peterson, 1968). Again neither misaggregation nor misperception can explain these findings.

Finally, several experimenters have noted that when Ss' estimates for the first datum in a sequence are separated from later estimates they look distinctly different. In fact the revisions associated with the first datum are very nearly optimal (Du Charme & Peterson, 1968; Peterson & Swensson, 1968, Wheeler, see footnote 1). There seems to be no reason why Ss should correctly perceive the first datum in a sequence and misperceive later data. And obviously the first datum must be aggregated

with the prior odds just as later data are. (It should be noted however that the prior odds in the three experiments just mentioned were always 1:1 so aggregation consisted simply of multiplying the L value by 1, a not very difficult task. Prior odds other than 1:1 should lead to difficulties in aggregation.) In these several instances misaggregation and misperception do not seem to be capable of explaining all the facts. This paper presents arguments for a hypothesis which can encompass these facts and then describes an empirical test of the hypothesis.

Response Bias Explanation

An assumption commonly made in this area of research (Edwards, 1968) is that the numbers Ss estimate represent their subjective probabilities, i.e., their true beliefs. Phillips & Edwards (1966) provide evidence that different subjective probabilities will be inferred depending on the type of response scale Ss use, e.g., odds or probabilities. Such a finding is not consonant with the belief that any number a subject estimates is a valid indicator of subjective probability.

Ramsey recognized this problem many years ago when he said, ". . .the measurement of beliefs is almost certainly an ambiguous process leading to a variable answer depending on how exactly the measurement is conducted." (1964, p. 69). Indeed the same sort of criticism has been leveled at magnitude estimation experiments--another area of research where Ss attach numbers to their subjective feelings. Both Poulton (1968) and Zinnes (1969) have recently discussed evidence for response biases in magnitude estimation experiments. It seems reasonable that the same

issue should be explored in the context of probability revision experiments.

The misaggregation and misperception hypotheses both essentially posit information processing errors on the part of the Ss. The problem, however, may lie in the response system. It may be naive to expect Ss to be so familiar with probability or odds scales that they make no response errors. What sort of a response bias could be hypothesized to explain the known facts? Obviously it would have to be a bias against extreme probabilities or odds. Two points must be made here. First, to state that Ss display a response bias and to state that Ss' probability revisions are conservative is not at all the same thing. The cause of conservatism may be in the information processing system or in the response system, and these two systems can be experimentally distinguished. Second, the term 'extreme' is a rather ambiguous one which obviously needs operational definition.

Given that Ss have a bias against extreme probabilities or odds, what existing data can be accounted for? A response bias will, of course, explain the occurrence of conservatism in general. Anytime a sequence of data points toward a particular hypothesis strongly enough, Ss' responses will appear conservative. They will be loath to estimate numbers as extreme as those called for by the optimal model. More particularly, a response bias of this nature can explain why Ss' conservatism increases when they deal with diagnostic data generators. On the average more diagnostic data generators will produce more diagnostic data. These data will require more extreme responses causing Ss' conservatism to increase. A similar argument explains the optimality of estimates

for the first datum in a sequence. On the average the optimal revision necessary after the first datum will be smaller than the revisions required after more data in the sequence have been seen. A response bias would thus have less chance to come into play on the first datum.

A bias against extreme numbers allows not only for the presence of conservatism in some situations, but for the lack of it in others. When Ss estimate credible intervals, or entire distributions, they are not making point (single number) estimates, and it is reasonable that these estimates might not necessarily reflect the same biases as the point estimates. Extreme responses in these two kinds of situations are quite different.

The reverse inertia effect observed by Du Charme & Peterson (1968) was explained by them as possibly being due to the fact that Ss revised their estimates along a curvilinear odds function. Confirming data moved them along the biased part of the function while disconfirming data moved them back over the linear, unbiased part of the function. This kind of a biased response function would lead to more conservative estimates for confirming than for disconfirming data.

On the face of it a response bais looks like a reasonable alternative explanation for conservatism in probabilistic inference. According to the present argument it can explain everything misaggregation or misperception can explain and somewhat more in addition. How well will the hypothesis do when subjected to an empirical test? The present experiments attempt to answer that question.

EXPERIMENT I

How can one distinguish a response bias from misaggregation or misperception? One way is to look at inferred log likelihood ratios (ILLR) as a function of prior odds (to obtain ILLR one subtracts the log of S's prior odds from the log of his posterior odds). If the response bias argument is correct, the same datum occurring when prior odds are close to 1:1, and when they are not, should lead to different amounts of posterior odds revision and therefore, to different ILLRs. Since the level of the prior odds is not an operative factor in either misperception or misaggregation, they predict no differences in the ILLRs. Accordingly, the first experiment presented Ss with two hypotheses and manipulated sequences of data bearing on these hypotheses so that the same data occurred at both high and low prior odds.

There are two ways that prior odds can be manipulated: they can be set at any desired level by the E before any data are sampled, and they can be changed by sequences of data. In the first half of the experiment the initial odds were always set at 1:1 and prior odds were manipulated by the use of hand picked sequences of data. In the second half of the experiment prior odds were set at various levels by E and then changed by samples of data. The question of interest was whehter this variable would have any effect on Ss' response functions. Are the two kinds of prior odds equivalent?

Although the response bias hypothesis predicts a reverse inertia effect, it also predicts that no 'primacy' effect should be found. A

primacy effect is described by Peterson & Du Charme (1967) as an over-reaction to early data in a sequence. Under a primacy effect Ss given a sequence of data pointing first toward one hypothesis and then toward another will be overinfluenced by the early data and end up with too much probability assigned to the first hypothesis. The response bias hypothesis predicts that Ss will simply move up and down their response functions and be neither more nor less influenced by early data. They should exhibit no primacy effect. To test this prediction, one sequence was included which reached odds of 100:1 in favor of one hypothesis, and then returned to odds of 1:1.

Finally, the response bias hypothesis also predicts that revisions based on Trial 1, or single datum trials, will not be optimal if the datum is so diagnostic that it forces S to respond on the biased part of his response function. To test this prediction, some very diagnostic single datum trials were included.

Method

Stimuli. The experiment used two normal distributions as data generators: the heights of United States men and women. These are the same data generators used by Du Charme & Peterson (1968) except that the standard deviations of the two populations were here assumed to be equal (the value used was 2.64 which is midway between the two reported in Du Charme & Peterson). The mean height for the male population was 68.2 inches and for the female 63.0 inches. For the first part of the experiment where the E-set prior odds were always 1:1, E made up 12 sequences to test the response bias hypothesis. The make up of these sequences

along with the response bias predictions derived for them can be seen in Table 1. The table identifies each sequence, gives its length and special characteristics, and then describes the response bias prediction. Comparisons are made between sequences with the same identifying number. For instance, for sequences 1A and 1B, which differ only in the diagnostic impact of their first datum the response bias hypothesis predicts different amounts of posterior odds revision on Trials 2 and 3 and therefore the difference in ILLRs shown in the table.

These sequences are not very representative of what one would expect from a truly random sampling process. To prevent Ss from becoming aware of this the sampling distribution was made representative by the addition of filler sequences. Three aspects of the sampling distribution were controlled with the aid of the filler sequences. For each trial across sequences (1) the expected value of the cumulative log likelihood ratio (CLLR) was maintained, (2) 66% of the CLLRs were kept within $\pm$ 1 standard deviation of the expected CLLR, and (3) the correct proportion of data items (about 20%) had Ls smaller than 1.0 with reference to the currently favored hypothesis. (Even though the male population is being sampled, about 20% of the time a height will be drawn which is more likely to have come from the female population and vice versa.) There were a total of 40 sequences (half of them single datum sequences) from 1-7 trials long. This yielded a total of 158 trials in the first part of the experiment.

In the second part of the experiment E set the prior odds at 2:1, 5:1, 10:1, and 100:1. At each prior odds level sequences of data

TABLE 1

Special Sequences

Sequence		Characteristics		Predictions
Number	Length			
1A	3	Trial 1, L=1.6	Trials 2 & 3 are exactly	ILLRs for Trials 2 & 3 in 1A >
1B	3	Trial 1, L=99	the same	ILLRs for same trials in 1B
2A	5	Trial 1, L=1.3	Trials 2, 3, 4, & 5 are	ILLRs for Trials 2-5 in 2A >
2B	5	Trial 1, L=99	exactly the same	ILLRs for same trials in 2B
3	6	The odds go to 100:1 for one hypothesis, and then back to 1:1		No primacy effect; odds at Trial 6 should be 1:1
4A	4	The last datum is the same in all three sequences. The prior odds before that datum are: 540:1 in 4A, 100:1 in 4B, and 17:1 in 4C.		ILLR for Trial 4 of 4C > ILLR for Trial 4 of 4B > ILLR for Trial 4 of 4A
4B	4			
4C	4			
5A	4	Four data are put in four arrangements such that each datum appears on Trial 1 in one sequence, Trial 2 in another sequence, etc. The final odds reach 10,000:1.		The ILLR for a datum will always be larger in that sequence in which the datum occurs at lower prior odds. There are 24 such comparisons.
5B	4			
5C	4			
5D	4			

from 1-4 trials in length were presented. These sequences were subjected to the same controls as those in the first part of the experiment. On the average, 21 trials were presented at each prior odds level for a total of 83 trials.

Subjects. Twenty-six paid University of Michigan male students served in groups of three, four, or five each.

Apparatus. A vertical board marked off in feet and inches in units of 1/4 inch was used to display the sequences. Magnetic pointers were attached to a metal strip fastened to one side of the board; each pointer represented a sampled height. The Ss responded on devices which featured an odds scale and a lever which moved along the scale. They set the lever to the point they desired on the odds scale, and then wrote those odds down. The odds scales were spaced logarithmically and went from 1:1 to 1,000,000:1 in six ranges (first range 1:1-10:1, second 10:1-100:1, and so on). The odds scales could be rotated to whichever of the six ranges S wanted to respond on.

Procedure. The experiment was broken up into three parts. In the first part Ss were presented with defining samples for the two populations. For both the second and third part of the experiment Ss revised odds estimates about which of the two populations had generated the random samples of data they were observing. During the second part of the experiment the prior odds at the beginning of each sequence were reset to 1:1. In part three these set prior odds were varied.

The defining samples were made up of 100 male and 100 female heights. The subjects were asked to make three estimates of the mean height of each population; before seeing the sample, midway through, and

at the end of the sample. By the end of each sample the average S could estimate the mean quite accurately.

In the second part of the experiment Ss were instructed that E had chosen one of the two populations by randomly drawing a poker chip from a canvas bag. The bag contained one red chip and one blue chip. If a red chip was drawn, the male population was sampled; if a blue, the female population was sampled. This procedure set the prior odds at 1:1. The Ss then observed a sequence of from 1-7 heights, and after each height revised their odds about which population was being sampled. The heights were cumulatively displayed throughout any given sequence. The order of the sequences was randomized for each group.

The third part of the experiment differed from the second only in that the prior odds were changed and the sequences were from 1-4 heights in length. Prior odds were set at 2:1, 5:1, 10:1, or 100:1 by informing Ss that the contents of the canvas bag had been changed to two red chips and one blue, or five red and one blue, or whatever the appropriate proportions were. The populations favored by the prior odds and the colors of the chips were counterbalanced. Prior odds conditions were randomized, and before each sequence Ss set their sliding levers at the appropriate odds.

Results

Figure 1 plots the Ss' average response function for part two of the experiment; it plots the median log estimated posterior odds as a function of log Bayesian odds. Trial 1 estimates are plotted as crosses; the dots represent estimates for Trials 2-7 and the circles

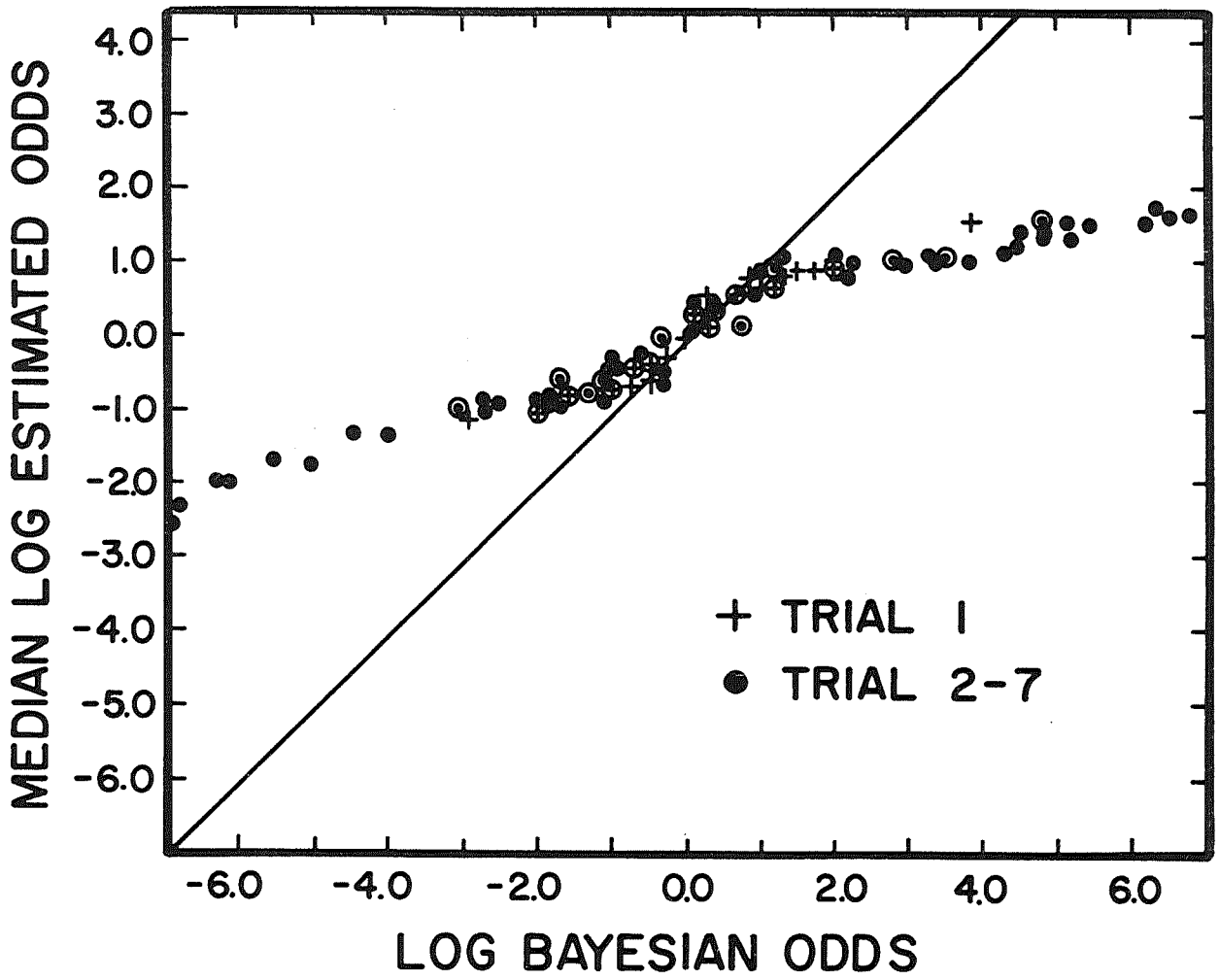

Fig. 1.--Median log estimated odds as a function of log Bayesian odds. Circles represent two data points at the same coordinates.

indicate that two data points fell on the same coordinates. A Bayesian S would revise his odds so they moved up and down the 45° line. Clearly the estimates of the median S are conservative. As in an earlier experiment (Du Charme & Peterson, 1968) Ss perform optimally only in the range of ± 1.0 log odds; outside that range their estimates are conservative. To check the representativeness of the median response function plotted in Fig. 1, a product-moment correlation and regression slope were calculated between the median log estimated odds and each Ss' log odds. The average correlation was .89, and the average slope 1.13 indicating that individual Ss are well represented by the median function.

It is important to note that the crosses representing single datum trials (the first datum in a sequence or a sequence of length one) fall on the same response function as the aggregated odds. This means that Trial 1 estimates are conservative if the datum is diagnostic enough, and that the aggregated odds are not conservative if the evidence is sufficiently undiagnostic. Both these results support the response bias hypothesis.

Further support for the hypothesis comes from an analysis of the special sequences used. The ILLR predictions listed in Table 1 were tested by obtaining ILLR difference scores for each S. For instance, one prediction in Table 1 is that the ILLR for Trial 2 in Sequence 1A will be larger than the ILLR for Trial 2 in Sequence 1B. To test that prediction the difference (D) between the two ILLRs was calculated. Since the response bias hypothesis predicts not only a difference but also the sign of the difference, the D score analysis was set up so that the response bias hypothesis always predicted positive differences. There were 33 D scores for each S, and the average D score across Ss was .16716.

The 95% confidence interval around this average D score is .10636-.22796 which does not include a difference of zero. The antilog of this average D score indicates that the ILLRs for the sequences with high prior odds must be multiplied by 1.47 to equal those for the low prior odds sequences.

The response bias prediction for sequence 3, the primacy sequence, was also borne out. The prediction was that the average estimate would be at 1:1, and the 95% confidence interval around the mean estimate does include this value. The mean estimate itself was 1.9:1 in favor of the hypothesis for which the odds had reached 100:1.

Figure 2 displays the median log estimated odds for the favored hypothesis as a function of the log Bayesian odds for the favored hypothesis in part three of the experiment. The diamond shaped figures represent estimates obtained at prior odds of 2:1, squares represent estimates at 5:1, triangles 10:1 and circles 100:1. Open figures are for Trial 1 and closed figures Trial 2-4. The circumscribed figures indicate two data points at the same coordinates. Again the 45^{o} line represents optimal performance. What the graph shows is that the range over which Ss revisions are optimal changes as a function of the E-set prior odds. This range effect is especially clear in the case of prior odds 100:1 (circles) and 10:1 (triangles). When Ss start with prior odds of 10:1, they revise nearly optimally up to about 100:1, starting from 100:1 they revise well up till nearly 1000:1. The optimal range holds for both single datum and aggregated trials as it does when prior odds are set at 1:1. If Ss are misaggregating, all the single datum trials should be conservative. Another way to look at these

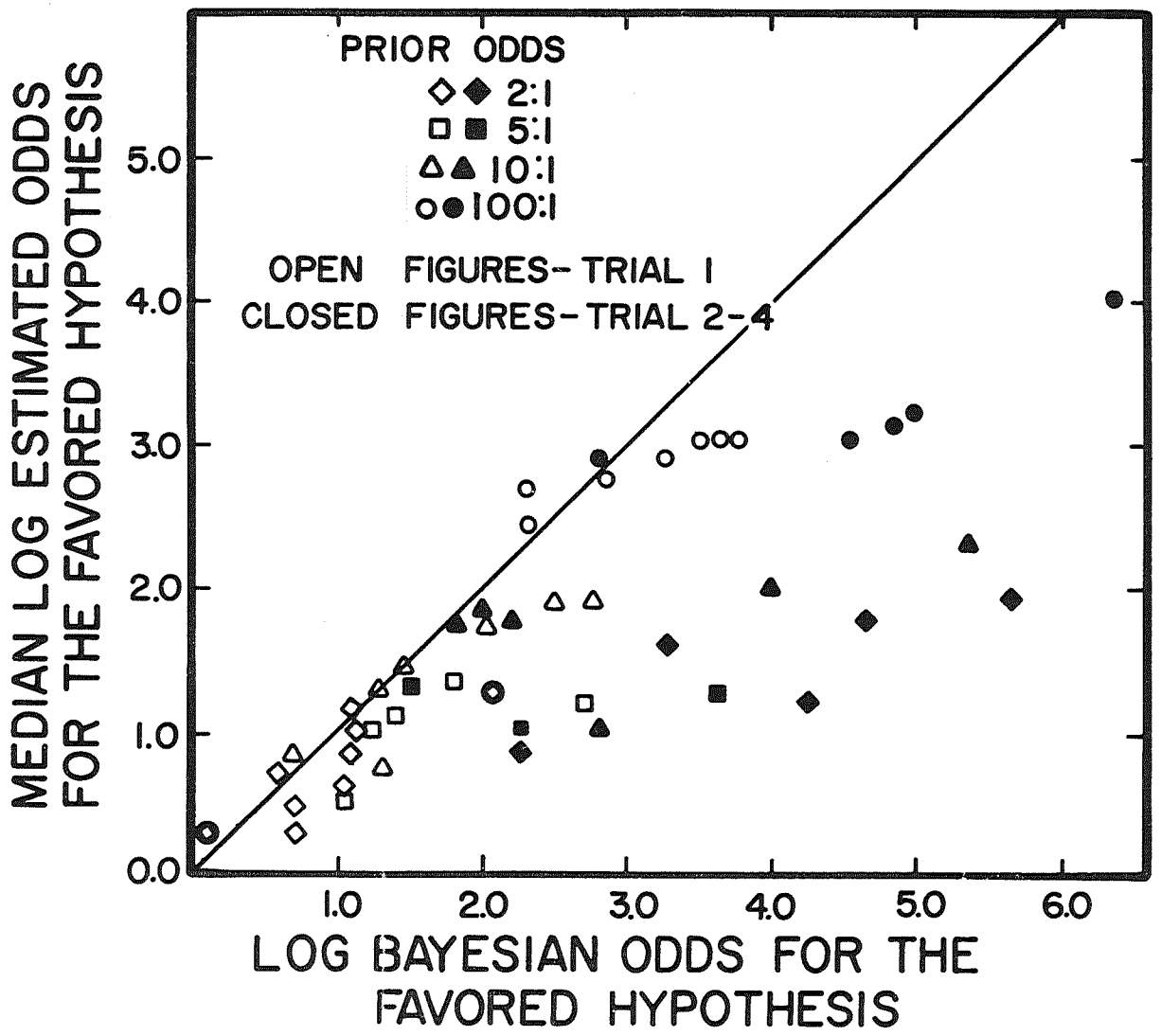

Fig. 2.--Median log estimated odds for the favored hypothesis as a function of log Bayesian odds for the favored hypothesis. Circumscribed figures represent two data points at the same coordinates.

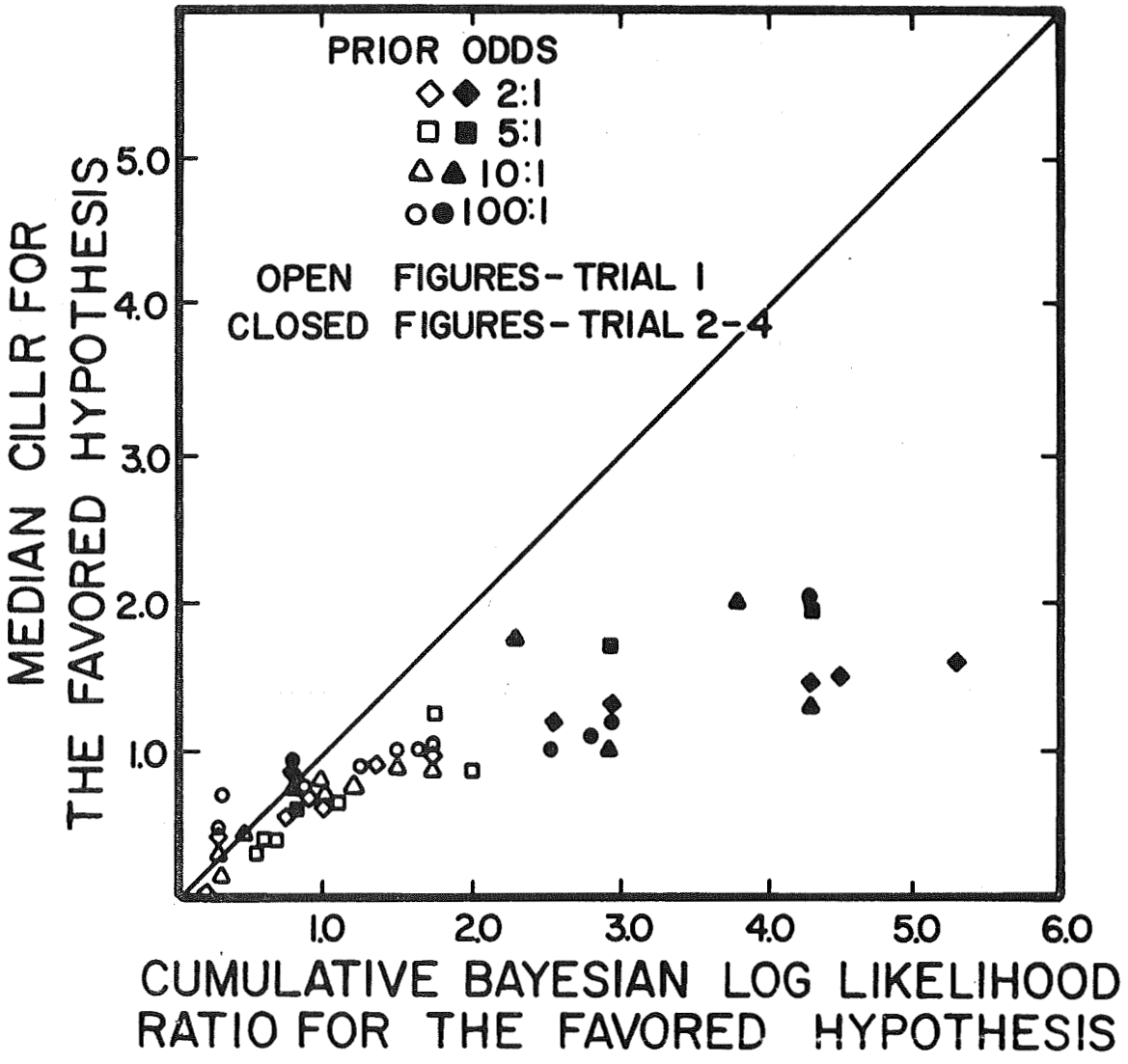

Fig. 3.--Median cumulative inferred log likelihood ratio (CILLR) for the favored hypothesis as a function of cumulative Bayesian log likelihood ratio for the favored hypothesis.

data is to disregard the Ss'starting point and simply look at the cumulative inferred log likelihood ratio (CILLR) function. Figure 3 plots the median CILLR against the cumulative Bayesian log likelihood ratio. The data points for all prior odds conditions appear to fall on the same function. This function looks nonconservative out to a CLLR of 0.8 and then becomes conservative.

Discussion

How much importance should be attached to the multiplicative difference of 1.47 between likelihood ratios inferred for revisions at extreme versus low prior odds? Although this finding is statistically significant, it may seem trivial. The following example will show that it is not. Assume that we have two people who operate on the median response function depicted in Fig. 1. We will let one person estimate posterior odds and the other likelihood ratios which will be machine aggregated. Further, suppose that enough data have been processed so that the odds estimator has just reached that part of his response function where it flattens out, i.e., about 10:1. On the average, after two more data favoring the same hypothesis, the machine aggregated odds of the likelihood ratio estimator will be about twice as great as those of the odds estimator (1.47^2), after four data they will be five times as great (1.47^4), and after six data ten times as great (1.47^6). Clearly the differences are not trivial.

There are several ways of interpreting the biased functions shown in Figs. 1, 2, and 3. The simplest explanation is that Ss are displaying a cumulative log likelihood ratio (CLLR) bias. This can be

seen most clearly in Figs. 1 and 3. Since the prior odds were always 1:1 for the data points plotted in Fig. 1, the cumulative log odds and CLLR plots are identical. Both Fig. 1 and Fig. 3 show a bias appearing in the CLLR function when CLLR reaches about 0.8. Regardless of where E sets the prior odds Ss can correctly interpret and aggregate only a certain amount of information. Once that limit has been exceeded, whether by one datum or by many, the impact of later data is less than it should be. This kind of a bias is neither misaggregation, misperception, nor a simple response bias.

A slightly more complicated explanation is that Ss do indeed display an odds, or number, bias, but not a fixed one. The bias is one against moving too far from the E-set prior odds whatever they are. The Ss shift their biased response function so that it is always centered on the E-set prior odds. Thus if E says the prior odds are 1:1 then 10:1 is an extreme estimate; if he says they are 100:1 then 1000:1 is extreme. It is as though the E-set prior odds provide a stable point from which Ss will willingly revise their odds. They are more reluctant to move toward extremeness from prior odds resulting from their own, perhaps fallacious, aggregation.

The misaggregation hypothesis can also be twisted to fit the data by tacking on the assumption that sometimes Ss aggregate correctly and sometimes they don't. The 'sometimes they don't' can be defined as anytime the CLLR exceeds 0.8 or anytime the cumulative odds exceed the E-set priors by a factor of about 10. Regardless of which way it is stated, the misaggregation hypothesis is clearly less adequate than it once appeared. It is not the number of data to be aggregated that causes

the problem but the amount of revision necessitated by the data. As for the other two interpretations, a certain intuitive appeal and some post experiment questioning of the Ss lead to the tentative conclusion that the odds bias explanation is the correct one.

Regardless of how the biased function is to be interpreted, the important fact is that it is there. Something which can be interpreted as a response bias exists, but how general is it? For instance, will the same response function be obtained if different data generators are used? In particular, does the form of the bias depend in any way on the familiarity most people undoubtedly have with height distributions? An equally important question concerns the effect of expected diagnosticity on the response function. Are Ss' response biases affected by the size of the numbers they expect to estimate? Experiment II sought to answer these questions.

EXPERIMENT II

In the previous experiment the expected log likelihood ratio (ELLR) of a random observation was .86923. This corresponds to a likelihood ratio of 7.4/1 which is also the likelihood ratio of an observation at the mean of either distribution. The ELLR and the likelihood ratio of an observation at the mean will be the same whenever two normal distributions with equal variances are used as data generators. The ELLR of a random observation depends on the distance between the means of the distributions and on their variances. This is exactly the relationship expressed by d'(d' = $\frac{\mu_1 - \mu_2}{\sigma}$) so it can be used as a convenient measure of expected diagnosticity. For the previous experiment d' was

equal to 2.0. If Ss' response functions represent a number bias then it is possible that the bias may change as a function of the numbers they expect to estimate. This hypothesis was tested by varying d' in Exp. II. Three values of d' were used--1.0, 2.0, and 2.77.

There were two other goals for Exp. II. One of them was to see if Ss would perceive two data generators with large mean separation and large variances as more diagnostic then two with small mean separation and small variances. It can be argued that the mean of a distribution is a much more readily inferred quantity than the variance. For that reason Ss might perceive mean separation as a more powerful determinant of diagnosticity. To the extent that they do so d' will not be an adequate measure of diagnosticity. Accordingly two conditions were run in which d' was set at 2.0; in one condition the mean separation and variances were the same as in Experiment I (except that the measurements were in centimeters), in the other condition the means were closer together and the variances were correspondingly smaller.

The final goal was to examine the effect of unfamiliar data generators on the response function. The normal distributions used as data generators in this experiment were not heights of men and women but lengths (measured in centimeters) of fictitious species of fish. Since the species were fictitious the data generators were clearly unfamiliar to the Ss. The purposes of the experiment then were to look at the effects of data generator unfamiliarity and diagnosticity on the response function, and to see if two different ways arriving at the same d' value would appear equivalent to Ss.

Method

Stimuli. Four different pairs of normal distributions were used as data generators. Each pair represented one of the d' conditions. The distributions were of the lengths of fictitious species of fish; each species was identified only by an arbitrary letter of the alphabet. Table 2 lists the relevant characteristics of each pair of distributions. Sequences of from 1-5 data were generated for each pair of species. There were more long sequences in the d' = 1.0 condition and more short ones in

TABLE 2

Distribution Characteristics

Distribution	Mean in cm.	Standard deviation	d'	L of an observation at the mean
A	32.5	3.25	2.0	7.4
B	26	3.25		
C	44	3.25	2.76	46.3
D	35	3.25		
E	56.5	6.5	1.0	1.6
F	50	6.5		
G	73	6.5	2.0	7.4
H	60	6.5		

the d' = 2.76 condition, but for each condition Ss saw 50 data. Relevant aspects of the sampling distribution across sequences were controlled here as they were in Exp. I.

Subjects. Twenty-nine paid University of Michigan male undergraduates served as Ss in groups of three, four, or five. There were seven groups.

Apparatus. The same board and magnetic pointers were used as in Exp. I, but the scale on the board was in centimeters. Four removable graphs displayed histograms or frequency distributions of the defining samples for each species pair. The Ss used the same response levers as in Exp. I.

Procedure. The four d' conditions were presented in four different orders with Group 1 receiving one order, and the other three orders each being given to two groups. At the beginning of each d'condition Ss saw two defining samples of 50 fish lengths, one sample for each species. They made two estimates of the mean for each sample, halfway through and again at the end. The average estimate was very accurate. After the defining sample for each of the two species under consideration had been given, a histogram or frequency distribution graph containing 100 samples for each species was displayed on the board. (Since Ss were familiar with the underlying distributions in Exp. I no such display was used there.) The Ss were told the graphs contained the 100 samples they had seen plus an additional 100 randomly drawn samples. They were told that although the underlying population of lengths undoubtedly looked something like the frequency distribution, the two were by no means the same. As an

example, it was pointed out to the S that if no fish of one particular length turned up in the random sample of 100, it did not follow that there were no fish that length in the population.

The Ss were next informed that one of the two species of fish under consideration had been chosen by tossing a fair coin. They revised their odds about which species had been chosen on the basis of randomly sampled fish lengths from that population. After they had seen all the sequences for one species pair, the frequency distribution graphs were removed and the entire procedure was repeated for the next pair.

Results and Discussion

Figures 4, 5, 6 and 7 display the results of the experiment in increasing order of d'. The figures plot the median log estimated odds against the log Bayesian odds. As in Fig. 1 the crosses represent Trial 1 estimates and the dots later trials. In two important respects these response functions are very similar to the one plotted in Fig. 1. First, single datum trials (the crosses) fall on the same function as aggregated trials, and second, the Ss estimates appear to be Bayesian over the $\pm$ 1.0 range of log odds. Unfamiliar data generators seem to yield the same response functions as familiar ones.

Table 3 presents the average product-moment correlation and regression line slope between the log of each S's estimated odds and the median estimated log odds. The correlation figures reveal that again the median functions are representative of individual Ss.

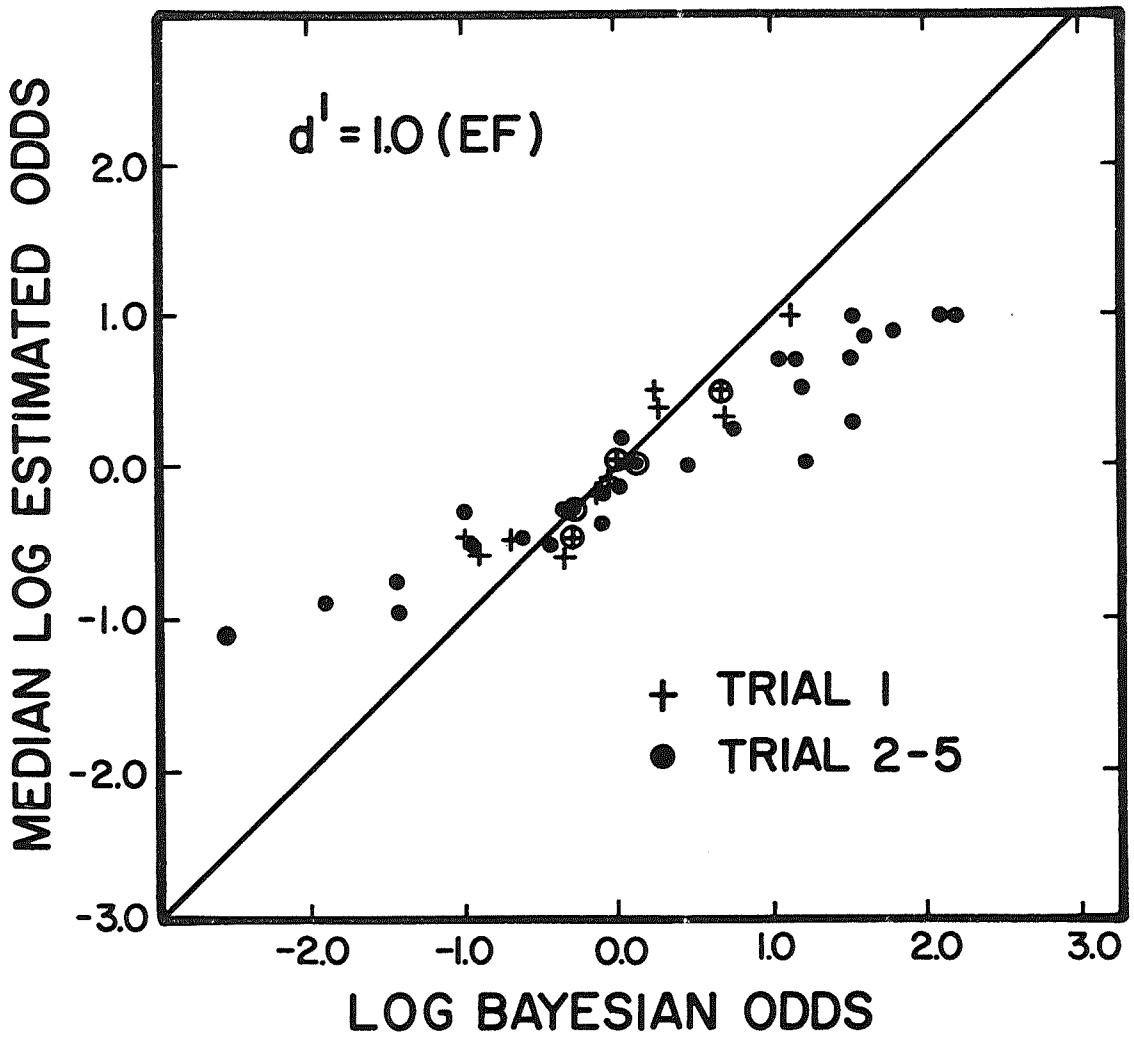

Fig. 4.--Median log estimated odds as a function of log Bayesian odds. Circles represent two data points at the same coordinates.

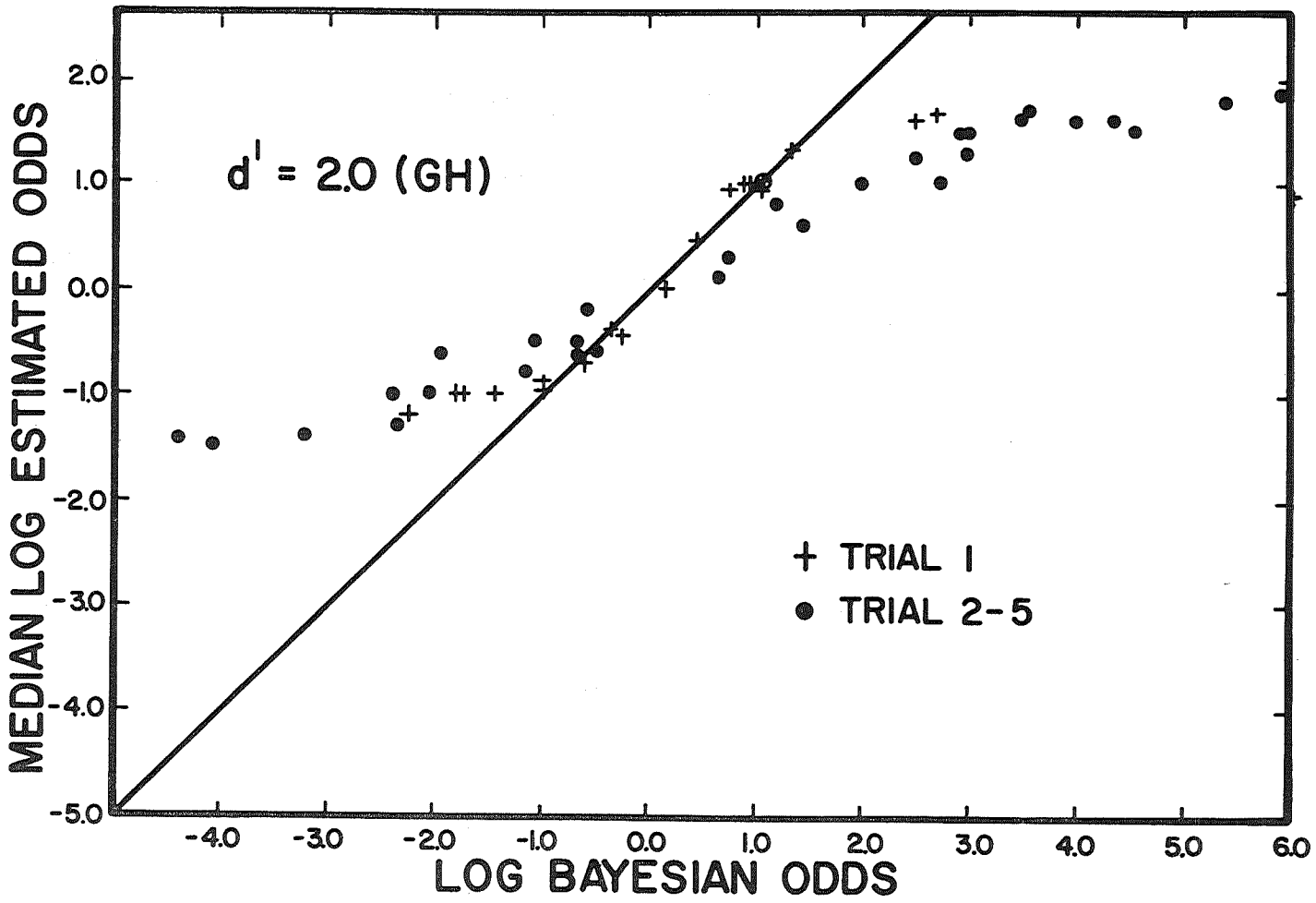

Fig. 5.--Median log estimated odds as a function of log Bayesian odds. Circles represent two data points at the same coordinates.

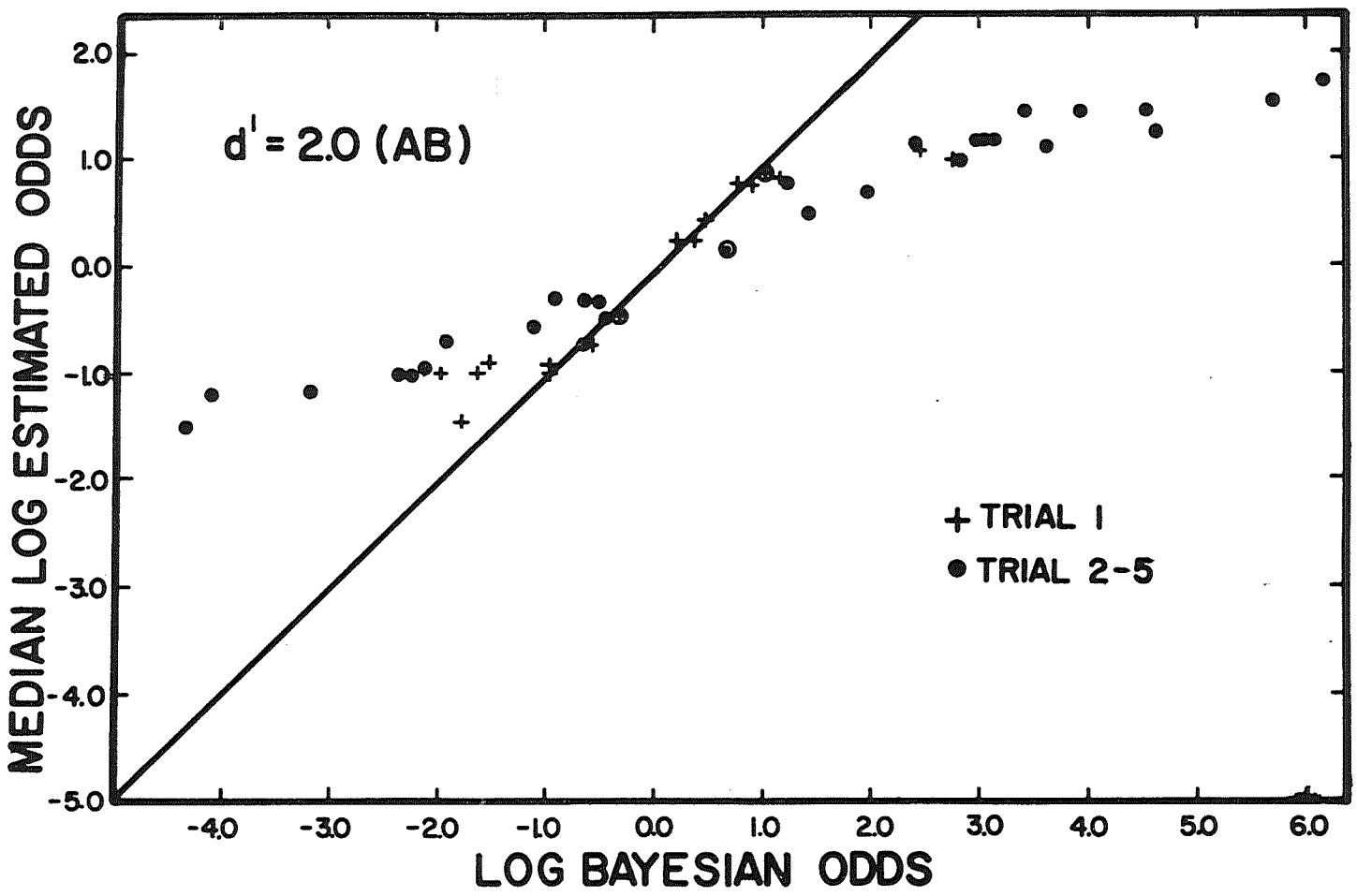

Fig. 6.--Median log estimated odds as a function of log Bayesian odds. Circles represent two data points at the same coordinates.

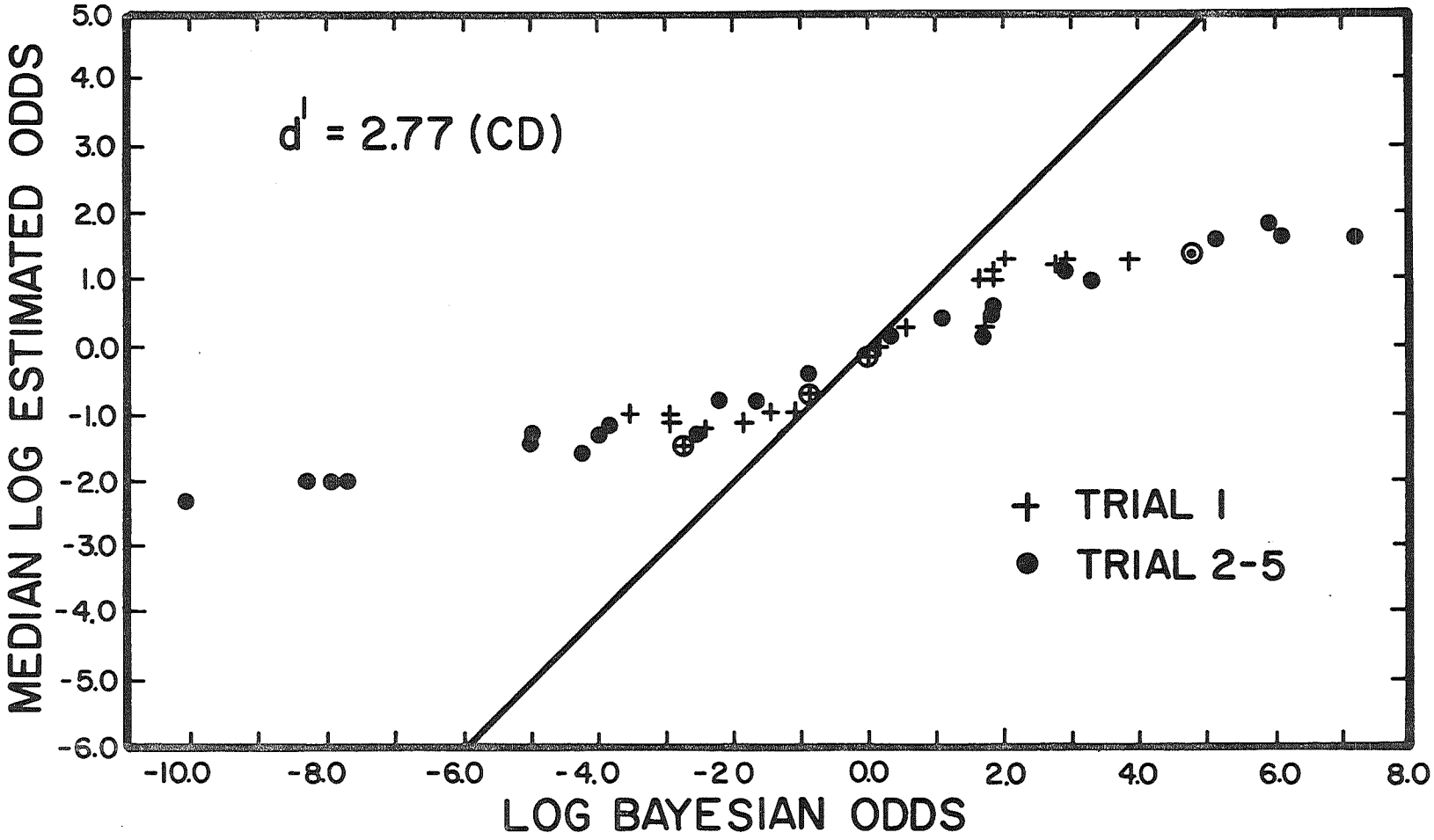

Fig. 7.--Median log estimated odds as a function of log Bayesian odds. Circles represent two data points at the same coordinates.

TABLE 3

Product-Moment Correlations (r) and Regression Slopes (b) Between the Log of Each S's Odds and the Median Log Odds

d' Condition	r	b
1.0 (EF)	.82	1.02
2.0 (GH)	.88	1.00
2.0 (AB)	.91	1.04
2.77 (CD)	.92	.98

Visual inspection of Figs. 5 and 6 indicates that Ss perceived the two conditions in which d' was 2.0 as equally diagnostic. The fact that the two median response functions appear similar means that d' is a useful index of diagnosticity in these tasks. In fact the response functions in all four figures look very similar. To facilitate a comparison of these response functions, running averages of the estimates for all four d' conditions are plotted in Fig. 8. (Except for being noisier, and therefore even more difficult to untangle, a plot of simple averages looks very much the same.) Running averages were calculated by assigning each median log odds estimate a value of one half itself plus one half a linear interpolation between the two es-

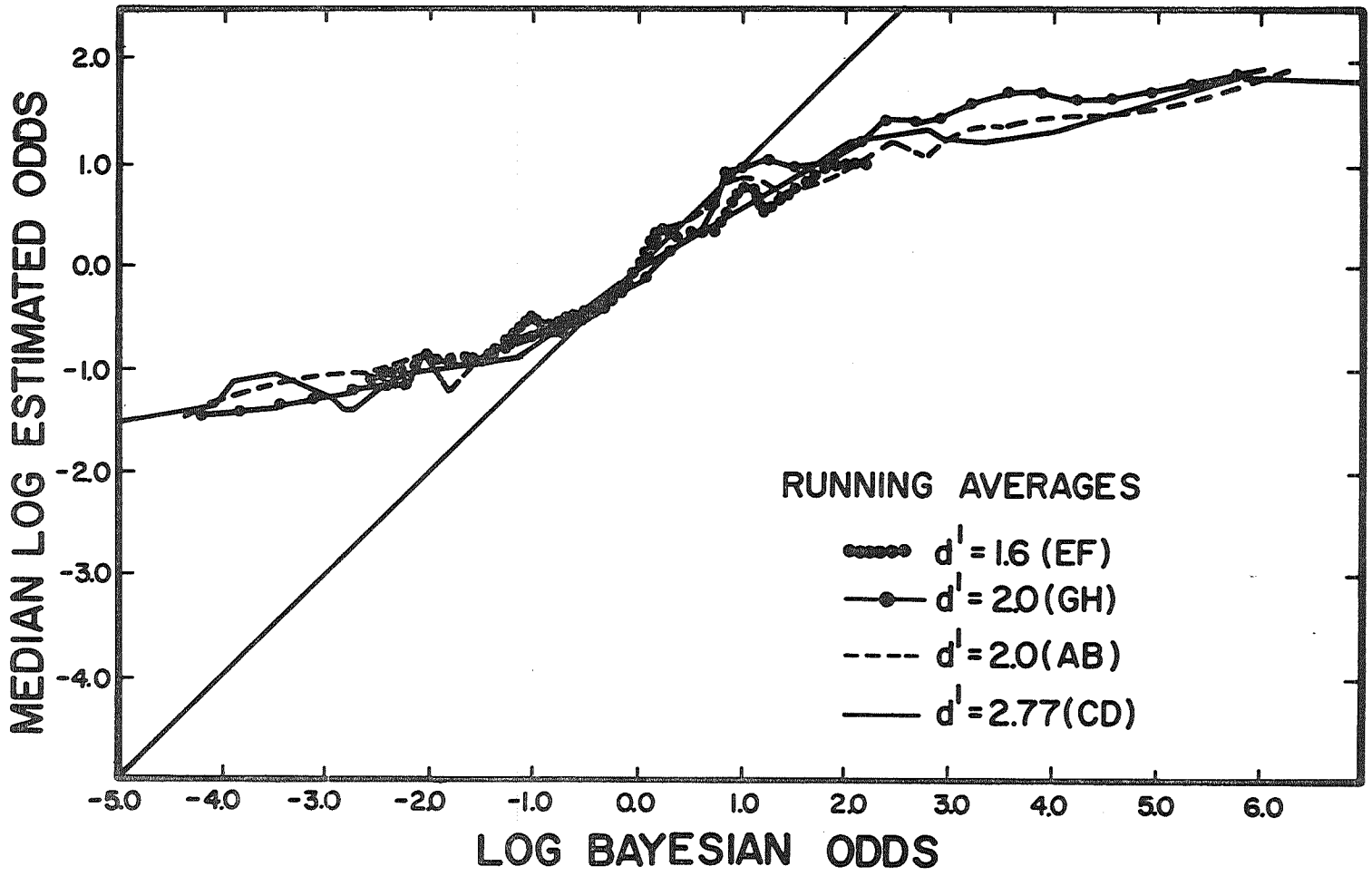

Fig. 3.--Running averages of the median log estimated odds as a function of log Bayesian odds.

timates on either side of it (the extreme ends of the d' = 2.77 function do not appear on the figure). Clearly these functions are similar; they are stable over extreme changes in expected diagnostic impact.

DISCUSSION

The notion that Ss' probabilistic revisions might be distorted is not a new one (Peterson, 1968; Sanders, 1968; Schum, Goldstein, Howell & Southard, 1967), but previously there were no data bearing directly on the hypothesis. The combined impact of the present experiments argues quite strongly for the presence of response biases in human odds estimation. Neither misperception nor misaggregation can easily encompass both the data reviewed earlier and the present data.

Not all questions have been resolved, of course. Experimental evidence on the issue of CLLR biases vs. number biases would be most useful. Further generalizations of the present approach with qualitatively different data generators, e.g., binomial or multinomial, and with different scales, e.g., probabilities, would also be interesting. Another line of extension would be to see if general training on the properties of number systems, odds or probabilities, results in improved revision performance. If the errors Ss presently make are due solely to lack of knowledge about these number systems, one would speculate that such training would be helpful.

More generally, if Ss do have response biases, what can be said about future research in the probability revision area? The obvious moral for experimenters in the area is that response biases may have a confounding effect on the variables they are interested in manipulating. It is necessary

to know how Ss use a particular response scale before any inferences about the effects of other variables can be made. This suggests that experimenters should gather information on Ss' response functions as a basic part of any probability revision experiment.

Probably the main reason the biased nature of Ss' response functions is so apparent in the present experiments is that the numbers they were called on to estimate are quite large. When d' is 2.0, for instance, the expected odds after a sample of six data are about 22,000:1 in favor of the most likely hypothesis. Such high numbers are probably beyond most people's experience with the number system. For applied work, the present results suggest that men should not be put in situations where they have to estimate high numbers in order to be accurate.

Edwards has suggested and tested a system where men estimate likelihood ratios and a computer combines them according to Bayes's theorem (Edwards, Phillips, Hays & Goodman, 1968). The rationale for the system was that people could not accurately process a number of data, they could not aggregate, and therefore it was best to have a machine do the job for them. The present results show that people can aggregate accurately over a limited range of evidence impact so the original rationale is weakened. The strategy is still useful, however, because it avoids the problem of biases in most cases. The exception occurs when very diagnostic data must be processed. Such data would require large likelihood ratio estimates, and thus allow the possibility of biasing. But since very diagnostic data are precisely the kind which rarely or never occur in real world diagnostic systems, Edwards' strategy remains a viable one.

APPENDIX

Data Points For Figures

Figure 1

Trial number	Median log estimated odds	Log Bayesian odds
1	0.3010	0.2039
2	0.4771	0.4892
3	0.8495	1.0189
1	1.0000	1.9971
2	1.0000	2.2825
3	1.0570	2.8121
1	0.3010	0.1222
2	-0.4771	-0.8973
3	0.1761	0.7534
4	0.1761	0.7330
5	-0.3314	-1.0205
1	-1.0880	-1.9971
2	-1.0396	-3.0167
3	-0.7782	-1.3660
4	-0.8451	-1.3863
5	-1.0000	-3.1398
1	0.3891	0.2853
2	0.9031	1.3049
3	1.1276	1.9975
4	1.0085	1.3049
5	0.4771	0.2853
6	0.0303	0.0000
1	-0.4771	-0.6927
2	-0.8741	-1.7930
3	-1.0396	-2.7306
4	-0.9542	-1.7930
1	-0.6990	-0.7746
2	-0.6990	-1.0599
3	-0.9031	-1.9975
4	-0.7386	-1.0599
1	-0.3010	-0.2853
2	-0.4771	-0.3057
3	-0.8741	-1.2433
4	-0.6761	-0.3057
1	0.9031	1.5085
2	1.0396	3.5057
3	1.0207	3.7910
4	1.3495	4.8105
1	0.9031	1.9971
2	1.0966	3.5057
3	1.4375	4.5252
4	1.6215	4.8105
1	0.6990	1.0195

Trial number	Median log estimated odds	Log Bayesian odds
2	0.8741	1.3049
3	1.1276	3.3020
4	1.5880	4.8105
1	0.5396	0.2853
2	0.7782	1.3049
3	1.0603	2.8166
4	1.5000	4.8105
1	0.5396	0.6927
2	-0.3010	-0.4897
3	-0.6021	-1.0194
1	-0.5396	-0.5297
2	-0.9515	-1.7121
3	-0.6990	-1.0194
1	0.6990	1.1824
2	-0.0510	-0.1631
3	0.6990	0.9372
4	1.1505	3.4248
5	1.5880	5.1783
6	1.6505	6.5239
7	1.6505	6.8094
1	0.6021	0.6927
2	-0.6276	-1.7129
3	-0.9515	-2.7324
4	-1.3010	-4.4859
5	-1.6990	-5.5863
6	-2.0106	-6.2789
7	-2.3010	-6.8086
1	0.6990	1.1824
2	-0.0207	-0.3261
3	0.8010	2.1614
4	1.2386	4.4843
5	1.3495	5.1770
6	1.5441	6.1965
1	-0.7386	-1.3456
2	-0.9515	-2.5280
3	-1.3495	-4.0365
4	-1.7559	-5.0560
5	-2.0000	-6.1564
6	-2.5396	-7.0128
1	0.8741	1.7535
2	1.0000	2.9359
3	1.1901	4.2815
4	1.5106	5.4639
5	1.7720	6.3203
1	-1.0880	-2.9765
1	0.3010	0.2039
1	0.3010	0.1222
1	0.4698	0.5297
1	-1.0000	-2.0792
1	1.6021	3.8890
1	-0.8116	-1.2634
1	0.3010	0.2853

Trial number	Median log estimated odds	Log Bayesian odds
1	-0.6761	-1.0195
1	1.0000	1.9971
1	-0.3495	-0.5297
1	0.0000	-0.0204
1	-0.3891	-0.4486
1	0.4771	0.4486
1	-0.7782	-1.1004
1	0.8406	1.2634
1	-0.6990	-0.9376
1	-0.8116	-1.5085
1	0.3891	0.3668
1	0.7386	0.8564
1	-0.8116	-1.6507

Figure 2 (negative log odds indicate sequences in which the data pointed toward the hypothesis not favored by the prior odds)

Trial Number	Median log estimated odds for the favored hypothesis	Log Bayesian odds for the favored hypothesis
	Prior odds 2:1	
1	0.6901	0.5864
2	1.1450	1.1161
1	-0.4771	-0.7185
2	-0.8741	-2.2270
3	-1.1901	-4.2240
1	1.2386	2.0545
2	1.5731	3.2370
3	1.7386	4.5825
4	1.9065	5.6021
1	0.3010	0.0972
1	1.0000	1.1575
1	-0.6021	-1.0445
1	1.2297	2.0545
1	0.8406	1.0756
1	-0.3010	-0.7185
	Prior odds 5:1	
1	1.0000	0.9843
2	1.3010	1.5140
1	-0.5396	-1.0545
2	-1.0000	-2.2369
3	-1.2386	-3.5825
1	1.0000	1.2287

Trial number	Median log estimated odds for the favored hypothesis	Log Bayesian odds for the favored hypothesis
1	1.1127	1.3916
1	1.3495	1.7993
1	1.5731	2.6961
1	0.3010	0.0875
	Prior odds 10:1	
1	1.3010	1.2853
2	1.7386	1.8150
1	0.8741	0.7147
2	-0.7386	-1.2824
3	-1.0000	-2.7910
1	1.9031	2.7535
2	2.0207	3.9359
3	2.3010	5.2815
1	1.4710	1.4456
1	1.6990	2.0195
1	1.7386	2.1824
1	1.8891	2.5085
1	1.8406	2.0195
	Prior odds 100:1	
1	2.6990	2.2853
2	2.9031	2.8150
1	2.9771	3.5085
2	2.9660	4.5281
3	3.1087	4.8134
1	2.9978	3.7535
2	3.2157	4.9359
3	4.0000	6.2815
1	2.4375	2.2853
1	2.7312	2.8564
1	2.9031	3.2634
1	2.9771	3.6507

Figure 3 (negative cumulative log likelihood ratios indicate sequences in which the data pointed toward the hypothesis not favored by the prior odds)

Trial number	Median CILLR for the favored hypothesis	Cumulative Bayesian log likelihood ratio for the favored hypothesis
	Prior odds 2:1	
1	0.3891	0.2853

Trial number	Median CILLR for the favored hypothesis	Cumulative Bayesian log likelihood ratio for the favored hypothesis
2	0.8440	0.8150
1	-0.7782	-1.0195
2	-1.1751	-2.5280
3	-1.4911	-4.5250
1	0.9375	1.7535
2	1.2720	2.9359
3	1.4375	4.2815
4	1.6054	5.3010
1	0.0000	-0.2039
1	0.6990	0.8564
1	-0.9031	-1.3456
1	0.9287	1.7535
1	0.5396	0.7746
1	-0.6021	-1.0195
	Prior odds 5:1	
1	0.3010	0.2853
2	0.6021	0.8150
1	-1.2386	-1.7535
2	-1.6990	-2.9359
3	-1.9375	-4.2815
1	0.3010	0.5297
1	0.4137	0.6927
1	0.6505	1.1004
1	0.8741	1.9971
1	-0.3979	-0.6115
	Prior odds 10:1	
1	0.3010	0.2853
2	0.7386	0.8150
1	-0.1259	-0.2853
2	-1.7386	-2.2824
3	-2.0000	-3.7910
1	0.9031	1.7535
2	1.0207	2.9359
3	1.3010	4.2815
1	0.4710	0.4486
1	0.6990	1.0195
1	0.7386	1.1824
1	0.8891	1.5085
1	0.8406	1.0195
	Prior odds 100:1	
1	0.6990	0.2853
2	0.9031	0.8150

Trial number	Median CILLR for the favored hypothesis	Cumulative Bayesian log likelihood ratio for the favored hypothesis
1	0.9771	1.5085
2	0.9660	2.5281
3	1.1087	2.8134
1	0.9978	1.7535
2	1.2157	2.9359
3	2.000	4.2815
1	0.4375	0.2853
1	0.7312	0.8564
1	0.9031	1.2634
1	0.9771	1.6507

Figure 4

Trial number	Median log estimated odds	Log Bayesian odds
1	0.3979	0.2505
1	-0.4771	-0.2840
1	0.4771	0.2171
1	-0.6021	-0.9187
1	-0.6021	-0.3508
1	-0.4771	-0.6848
2	-0.0414	0.0000
1	0.4771	0.6514
2	0.0000	1.2026
3	0.3010	1.5200
1	-0.0792	-0.0501
2	0.0000	0.0000
3	-0.3010	-1.0189
1	-0.2305	-0.1503
2	0.5315	0.9354
3	0.1761	0.0835
4	0.1761	0.0000
1	-0.4771	-0.2840
2	-0.3010	-0.3007
3	-0.3010	-0.3174
4	-0.5682	-0.4677
1	0.4771	0.6514
2	0.6990	1.1692
3	0.9031	1.8206
4	1.0000	2.2048
1	0.3010	0.7182
2	0.0000	0.1002
3	-0.1139	-0.0167

Trial number	Median log estimated odds	Log Bayesian odds
4	-0.3979	-0.1002
5	0.0000	0.0835
1	-0.4771	-0.3174
2	0.0000	0.4343
3	0.2305	0.7517
4	0.4771	1.2027
5	0.6990	1.4867
1	0.0000	-0.0167
2	-0.3010	-0.3341
3	0.0000	0.0501
4	-0.2788	-0.3675
5	-0.4771	-0.6514
1	1.0000	1.1191
2	0.6990	1.0690
3	1.0000	1.5200
4	1.0000	2.1046
5	0.8451	1.6202
1	-0.4771	-1.0189
2	-0.7634	-1.4365
3	-0.9294	-1.9209
4	-1.0000	-1.9042
5	-1.1139	-2.5556

Figure 5

Trial number	Median log estimated odds	Log Bayesian odds
1	1.0000	0.9354
1	-0.9031	-1.0022
1	1.0000	0.8686
1	-1.1761	-2.2717
1	-0.4771	-0.2673
1	-1.0000	-1.8040
1	1.6021	2.4721
1	-1.0000	-1.7371
2	-0.4771	-1.0690
1	1.0000	1.0690
2	1.6990	3.5411
1	-0.3617	-0.3341
2	-1.0000	-2.0712
1	0.0000	0.1336
2	1.0000	2.7394
3	1.4771	4.5434
1	0.3010	0.3341
2	-0.4771	-0.6681

Trial number	Median log estimated odds	Log Bayesian odds
3	-0.7782	-1.1358
1	0.4771	0.4677
2	0.7782	1.2027
3	1.4771	2.9398
1	-1.0000	-1.4699
2	-1.3979	-3.2217
3	-1.3802	-4.4244
1	-0.9542	-1.0022
2	0.0792	0.6682
3	0.6021	1.4700
1	-0.6990	-0.6013
2	-0.6990	-0.6682
3	-0.6021	-0.4678
4	-1.0000	-2.4054
1	0.9542	1.0690
2	1.2305	2.5122
3	1.6021	3.5144
4	1.5798	3.9821
1	1.0000	1.0690
2	0.3010	0.7349
3	1.0000	2.0044
4	1.4771	3.0066
1	1.6990	2.6726
2	1.3010	3.0067
3	1.6021	4.3430
4	1.8451	5.9465
5	1.7782	5.4120
1	0.9542	0.7350
2	-0.1761	-0.5345
3	-0.6021	-1.9376
4	-1.3010	-2.3385
5	-1.4771	-4.0756

Figure 6

Trial number	Median log estimated odds	Log Bayesian odds
1	0.9031	1.0022
1	-1.0000	-1.0022
1	0.7782	0.8686
1	-1.0000	-2.0249
1	-0.4771	-0.3341
1	-0.9542	-1.8040
1	1.0792	2.4721
1	-1.0000	-1.6704
2	-0.3010	-0.9354
1	0.8451	1.1358

Trial number	Median log estimated odds	Log Bayesian odds
2	1.1761	3.6079
1	-0.4771	-0.3341
2	-0.9542	-2.1381
1	0.2553	0.2004
2	1.0000	2.8061
3	1.3010	4.6101
1	0.2553	0.3341
2	-0.3010	-0.6681
3	-0.5441	-1.1348
1	0.4771	0.4677
2	0.7782	1.2027
3	1.2305	3.0067
1	-0.9031	-1.5367
2	-1.1761	-3.2070
3	-1.4771	-4.3428
1	-0.9542	-1.0022
2	0.1761	0.6682
3	0.4771	1.4032
1	-0.6990	-0.6013
2	-0.6990	-0.6681
3	-0.4771	-0.4677
4	-1.0000	-2.4053
1	0.9031	1.0022
2	1.1761	2.4053
3	1.5052	3.4075
4	1.4771	3.8752
1	0.0931	1.0022
2	0.2041	0.6681
3	0.6990	1.9376
4	1.1761	2.9398
1	1.0000	2.7394
2	1.1761	3.0735
3	1.4771	4.4766
4	1.8129	6.1470
5	1.6021	5.6793
1	0.8062	0.7350
2	-0.3010	-0.5345
3	-0.6990	-1.9376
4	-1.0000	-2.2717
5	-1.1761	-4.0757

Figure 7

Trial number	Median log estimated odds	Log Bayesian odds
1	1.1761	1.8502

Trial number	Median log estimated odds	Log Bayesian odds
1	-1.1761	-1.8502
1	1.0000	1.6652
1	-1.0000	-3.5155
1	1.2305	2.7753
1	-1.1761	-2.9604
1	1.3010	3.8855
1	-1.4771	-2.7754
2	-0.8451	-2.2203
1	1.3802	2.0353
2	1.8751	5.9208
1	-0.7782	-0.9251
2	-1.1761	-3.8855
1	0.3424	0.7401
2	1.4150	4.8106
1	-0.1761	0.0000
2	-0.4771	-0.9251
1	1.0000	1.8502
2	1.6990	6.1058
1	-1.0000	-1.1101
2	-1.3010	-2.5903
1	1.3010	2.9609
2	1.4771	4.8106
1	-1.0414	-2.9609
2	-1.3010	-4.9957
1	-0.7782	-0.9251
2	0.4771	1.8503
1	-1.0792	-1.4802
2	-1.4771	-2.7754
3	-2.0000	-7.7711
1	0.0792	0.1850
2	1.0000	3.3304
3	1.6021	5.1806
1	-1.2305	-2.4053
2	-1.6021	-4.2555
3	-2.0000	-8.3260
4	-2.3010	-10.1762
1	-0.1761	0.0000
2	0.4771	1.1101
3	1.1761	2.9603
4	1.6990	7.2159
1	-0.7782	-0.9251
2	0.1761	0.3701
3	0.1761	0.7401
4	0.6021	1.8502
1	0.3010	0.5551
2	-0.8451	-1.6652
3	-1.3010	-4.0705
4	-1.4624	-4.9956
5	-2.0000	-7.9560

Figure 8 The running averages were found geometrically rather than arithmetically. For each figure 4-7 a running average was plotted geometrically on an overlay. The four running average plots were then transferred to Fig. 8.

REFERENCES

Beach, L. R. Accuracy and consistency in the revision of subjective probabilities. IEEE Transactions on Human Factors in Electronics, 1966, 7, 29-37.

Beach, L. R. Probability magnitudes and conservative revision of subjective probabilities. Journal of Experimental Psychology, 1968, 77, 57-63.

Du Charme, W. M. A review and analysis of the phenomenon of conservatism in human inference. Houston: W. M. Du Charme, 1969 (Mimeo).

Du Charme, W. M. & Peterson, C. R. Intuitive inference about normally distributed populations. Journal of Experimental Psychology, 1968, 78, 269-275.

Du Charme, W. M. & Peterson, C. R. Proportion estimation as a function of proportion and sample size. Journal of Experimental Psychology, 1969, in press.

Edwards, W. Conservatism in human information processing. In B. Kleinmuntz (ed.), Formal Representation of Human Judgment. New York: Wiley, 1968, pp. 17-52.

Edwards, W., Phillips, L. D., Hays, W. L., & Goodman, B. C. Probabilistic information processing systems: design and evaluation. IEEE Transactions on Systems Science and Cybernetics, 1968, SSC-4, 248-265.

Geller, E. S. & Pitz, G. F. Confidence and decision speed in the revision of opinion. Organizational Behavior and Human Performance, 1968, 3, 190-201.

Gustafson, D. H. Evaluation of probabilistic information processing in medical decision making. Organizational Behavior and Human Performance, 1969, 4, 20-34.

Goldstein, I. L, Southard, J. F., & Schum, D. A. Feedback in a complex multiman-machine system. Journal of Applied Psychology, 1967, 51, 346-351.

Kaplan, R. J. & Newman, J. R. Studies in probabilistic information processing. IEEE Transactions on Human Factors in Electronics, 1966, HFE-7, 49-63.

Lichtenstein, S. & Feeney, G. J. The importance of the data-generating model in probability estimation. Organizational Behavior and Human Performance, 1968, 3, 62-67.

Peterson, C. R. Aggregating information about signals and noise. Proceedings of the American Psychological Association Annual Convention, 1968, 123-124.

Peterson, C. R. & Du Charme, W. M. A primacy effect in subjective probability revision. Journal of Experimental Psychology, 1967, 73, 61-65.

Peterson, C. R., Du Charme, W. M., & Edwards, W. Sampling distributions and probability revisions. Journal of Experimental Psychology, 1968, 76, 236-243.

Peterson, C. R. & Miller, A. J. Sensitivity of subjective probability revision. Journal of Experimental Psychology, 1965, 70, 117-121.

Peterson, C. R. & Phillips, L. D. Revision of continuous subjective probability distributions. IEEE Transactions on Human Factors in Electronics, 1966, 7, 19-22.

Peterson, C. R. & Swensson, R. G. Intuitive statistical inferences about diffuse hypotheses. Organizational Behavior and Human Performance, 1968, 3, 1-11.

Peterson, C. R., Ulehla, Z. J., Miller, A. J., Bourne, L. E., Jr., & Stilson, D. W. Internal consistency of subjective probabilities. Journal of Experimental Psychology, 1965, 70, 526-533.

Phillips, L. D. Some components of probabilistic inference. Unpublished doctoral dissertation, University of Michigan, 1966.

Phillips, L. D. & Edwards, W. Conservatism in a simple probability inference task. Journal of Experimental Psychology, 1966, 72, 346-357.

Pitz, G. F., Downing, L., & Reinhold, H. Sequential effects in the revision of subjective probabilities. Canadian Journal of Psychology, 1967, 21, 381-393.

Poulton, E. C. The new psychophysics: six models for magnitude estimation. Psychological Bulletin, 1968, 69, 1-19.

Ramsey, F. P. Truth and probability. In F. P. Ramsey, The Foundations of Mathematics and Other Logical Essays. New York: Harcourt Brace, 1931. Reprinted in H. E. Kyburg, Jr., & H. E. Smokler (Eds.) Studies in Subjective Probabilities. New York: Wiley, 1964.

Sanders, A. F. Choice among bets and revision of opinion. Acta Psychologica, 1968, 28, 76-83.

Schneider, R. J. Optimality of subjective probabilities in clinical inference. Unpublished MA thesis, University of Colorado, 1965.

Schum, D. A., Goldstein, I. L., Howell, W. C., & Southard, J. F. Subjective probability under several cost payoff arrangements. Organizational Behavior and Human Performance, 1967, 2, 84-104.

Schum, D. A., Southard, J. F., & Wombolt, L. Aided human processing of inconclusive evidence in diagnostic systems: A summary of experimental evaluations. Aerospace Medical Research Laboratories Technical Report, 1969, in press.

Vlek, C. A. J. & Beintema, K. A. Subjective likelihoods in posterior probability estimation. Psychological Institute Report No. E 014-67, 1967, University of Leiden, The Netherlands.

Vlek, C. A. J. & Van der Heijden, L. H. C. Subjective likelihood functions and variations in the accuracy of probabilistic information processing. Psychological Institute Report No. E 017-67, 1967, University of Leiden, The Netherlands.

Vlek, C. A. J. & Van der Heijden, L. H. C. Alternative data-generating models for probabilistic information processing. Organizational Behavior and Human Performance, 1969, 4, 43-55.

Wheeler, G. & Beach, L. R. Subjective sampling distributions and conservatism. Organizational Behavior and Human Performance, 1968, 3, 36-46.

Zinnes, J. L. Scaling. Annual Review of Psychology, 1969, 20, 447-512.

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
1. inference						
2. conservatism						
3. probability revision						
4. response bias						
5. Bayes's theorem						
6. optimality						

TECHNICAL REPORTS

1. Phillips, L. D. Some components of probabilistic inference. January 1966.
2. Egeth, H. E. Parallel versus serial processes in multidimensional stimulus discrimination. January 1966.
3. Biederman, I. Human performance in contingent information processing tasks. October 1966.
4. Peterson, C. R. & Beach, L. R. Man as an intuitive statistician. November 1966.
5. Martin, E. & Melton, A. W. Recognition memory for CCC and CVC trigrams of various association values. January 1967.
6. Smith, E. E. Choice reaction time: An analysis of the major theoretical positions. January 1967.
7. Reicher, G. M. Perceptual recognition as a function of meaningfulness of stimulus material. February 1968.
8. Ligon, E. The effects of similarity on very-short-term memory under conditions of maximal information processing demands. May 1968.
9. Triggs, T. J. Capacity sharing and speeded reactions to successive signals. August 1968.
10. Tversky, B. G. Pictorial and verbal encoding in short-term memory. October 1968.
11. Lively, B. L. The Von Restorff effect in very-short-term memory. December 1968.
12. Greeno, J. G. How associations are memorized. December 1968.
13. Swensson, R. G. The elusive tradeoff: Speed versus accuracy in choice reaction tasks with continuous cost for time. December 1968.
14. Bjork, R. A. Repetition and rehearsal mechanisms in models for short-term memory. May 1969.
15. Kamlet, A. S. Processing of sequentially presented signals in information-combining tasks. June 1969.
16. Pollatsek, A. W. Rehearsal, interference, and spacing of practice in short-term memory. July 1969.
17. Garskof, M. H. The effect of spacing and variation of repetition in short-term memory. August 1969.
18. Shulman, H. G. Presentation rate, retention interval, and encoding in short-term memory for homonyms, synonyms, and identical words. September 1969.

MEMORANDUM REPORTS

1. Fitts, P. M. Cognitive factors in information processing. February 1967.
2. Melton, A. W., Sameroff, A., & Schubot, E. D. Short-term recognition memory. May 1967.
3. Martin, E. Responses to stimuli in verbal learning. October 1967.
4. Melton, A. W. First annual report: Human information handling processes. June 1968.
5. Jahnke, J. C. The Ranschburg paradox. July 1968.
6. Greeno, J. G. Theory of graphs on sets with applications to problem solving and understanding. October 1968.
7. Edwards, W. A bibliography of research on behavioral decision processes to 1968. January 1969.
8. Melton, A. W. Second annual report: Human information handling processes. June 1969.
9. Greeno, J. G. A cognitive interpretation of negative transfer and forgetting of paired associates. November 1969.