DOCUMENT RESUME

ED 045 695 TM 000 171

AUTHOR	Lewis, Charles
TITLE	The Countback Method for Analyzing Sensitivity Data.
INSTITUTION	Educational Testing Service, Princeton, N.J.
SPONS AGENCY	National Science Foundation, Washington, D.C.; Office of Naval Research, Washington, D.C.
REPORT NO	RB-70-30
PUB DATE	May 70
NOTE	82p.
EDRS PRICE	EDRS Price MF-$0.50 HC-$4.20
DESCRIPTORS	Perception, Perceptual Development, Probability, Reactive Behavior, *Research Methodology, *Response Mode, Simulation, *Statistical Analysis, *Statistics, *Stimulus Behavior
IDENTIFIERS	*Countback Method, Sensitivity Data

ABSTRACT

Sensitivity data is defined as involving two response categories, with responses observed at different levels of some variable. The responses are taken to indicate sensitivity to the variable and may be labeled "positive" or "negative." The countback method offers confidence limits for the 50% point, the level of the variable for which the two responses are equally likely, in the special case when only one response is observed at each level. A lower bound on the confidence coefficient for these limits can be given, assuming only independent responses and the existence of something like a 50% point. No specific model of response probabilities is required, and the variable need not even be continuous. To gain a greater understanding of the countback method, it is used here to analyze sensitivity data obtained in an experiment on the perception of depth in pictures, and to study two artificial situations where the response probabilities can be specified exactly. The countback method is compared with the minimum Chi-square and the maximum likelihood estimation methods and its extension to data with more than one response per level is discussed. The results suggest that the countback method has a significant place among previous methods for analyzing sensitivity data. Appendices include historical, statistical, simulation, and experimental details. (Author/CK)

RB-70-30

THE COUNTBACK METHOD FOR ANALYZING SENSITIVITY DATA

Charles Lewis
Princeton University

Prepared in connection with research done under
Office of Naval Research Contract N 00014-67-A-0151-0006
Project No. NR 006-04-01 and
National Science Foundation Grant GB-3402
Continuation of G-22889
with Princeton University, Princeton, New Jersey
Project on "Mathematical Techniques in Psychology"
Principal Investigator: HAROLD GULLIKSEN

Educational Testing Service
Princeton, New Jersey
May 1970

THE COUNTBACK METHOD FOR ANALYZING SENSITIVITY DATA

by

Charles Lewis

A DISSERTATION

PRESENTED TO THE

FACULTY OF PRINCETON UNIVERSITY

IN CANDIDACY FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE BY THE

DEPARTMENT OF STATISTICS

May 1970

THE COUNTBACK METHOD FOR ANALYZING

SENSITIVITY DATA

Abstract

Sensitivity data are defined as involving two response categories, with responses observed at different levels of some variable of interest. The responses are taken to indicate sensitivity to the variable, and, in this regard, may be labeled "positive" and "negative." Of particular interest in analyzing sensitivity data is the estimation of the 50% point, the level of the variable for which the two responses are equally likely.

The countback method provides confidence limits for the 50% point in the special (but important) case when only one response is observed at each level. A lower bound on the confidence coefficient for these limits can be given, assuming only independent responses and the existence of something like a 50% point. No specific model of response probabilities is required, and the variable of interest need not even be continuous.

To gain a greater understanding of the countback method, we use it to analyze sensitivity data obtained in an experiment on the perception of depth in pictures. Its performance is also studied in two artificial situations, where the response probabilities can be specified exactly. In the first of these, the probability of a positive response at different levels is given by the cumulative logistic model. The second situation is highly asymmetric, using the logistic model below the 50%

point and an errorless model (the positive response probability set equal to 1) above it.

For purposes of comparison, two asymptotically efficient estimation methods, namely maximum likelihood and minimum χ^2, are also applied to the cases described above. In the two artificial situations, where actual confidence coefficients can be determined, both of these methods fall short of their asymptotic performance levels.

The countback confidence limits, by contrast, perform at a substantially higher level than that given as a lower bound by the theory. Moreover, the ability of conservatively chosen countback limits to reject null hypotheses is quite adequate for the purposes of the experiment mentioned earlier. Finally, a simple extension of the countback method to data with more than one response per level is discussed.

On the basis of these results, the countback method is seen as taking a significant place among the previously existing methods for analyzing sensitivity data.

Appendices are provided for those interested in statistical, simulation, experimental, and historical details.

Acknowledgments

I would like to thank John W. Tukey for the organization and direction he has given this work. In addition, I am indebted to Professor Tukey for his careful readings of the manuscript at various stages and his invariably helpful and illuminating suggestions and comments. Of course, problems, ambiguities, and errors which remain are entirely my responsibility.

Among the many others who have provided me with both relevant training and encouragement, I am especially grateful to Harold Gulliksen, Melvin R. Novick, and Hans Wallach.

I also greatly appreciate the efforts of many individuals at Educational Testing Service responsible for putting the work in its final form, particularly Ann King for supervising these efforts, Betty Clausen for the final typing, and Marie Davis and Frances Shaffer for proofreading the manuscript.

Much of the initial research was conducted while I was an Educational Testing Service Psychometric Fellow and was supported by Educational Testing Service; the Office of Naval Research, Contract N 00014-67-A-0151-0006, administered through Princeton University; and the National Science Foundation, Grant GB-3402, continuation of G-22889, administered through Princeton University, Project on "Mathematical Techniques in Psychology," Harold Gulliksen, Principal Investigator.

Considerable use was made of the facilities of the Princeton University Computer Center and of the Kiewit Computation Center of Dartmouth College.

Finally, I wish to express my gratitude to Subject 11, for being perfect.

List of Tables

List of Tables (Contd)

List of Figures

THE COUNTBACK METHOD FOR ANALYZING

SENSITIVITY DATA

Consider the following experiments. (1) A human subject is asked to observe very dim flashes of light. The intensity of the flash is varied and the subject reports, for each flash, whether or not he was able to see it. (2) A weight is dropped from different heights onto several samples of the same explosive. In each case, it is recorded if the sample exploded. (3) Insects of a given species are exposed to different concentrations of an insecticide. Each is observed to see if it dies as a result. (4) Several different dilutions are made from a sample of soil. The presence or absence of a species of protozoan in a fixed amount of each dilution is determined after exposure to a suitable medium.

These are examples of what may be called the sensitivity experiment. This is most clearly an apt label for the first three. They are aimed, respectively, at determining the sensitivity of the eye to light, of explosives to impact, and of insects to insecticide. In each, a two-category response is examined to see the effect of different levels of the variable of interest. Each has the goal of estimating the value of the variable for which the two responses will occur with equal probability. This value (which will be referred to as the 50% point) is taken in these examples to be a good indication of sensitivity.

It should be noted that this probabilistic approach is an essential feature of the sensitivity experiment, and for good reason. Only rarely do sensitivity data indicate the existence of a critical level, with one response always observed below and the other always observed above it. Consequently, the simple notion of a critical level had to be generalized, and the 50% point is the most widely accepted generalization.

In the fourth experiment, a model is used to estimate directly the number of organisms in the original sample, but this value is a simple function of the dilution at which 50% of the exposed mediums will be sterile. In the same way, the first experiment may be concerned with the minimum number of photons necessary for vision, rather than the intensity which produces a visible flash half the time. Again, these two are directly related when an appropriate model is introduced.

Thus, the general sensitivity problem, as it has been called, is to estimate a 50% point, or some function of it, from two-category response data.

There is some latitude available in designing a sensitivity experiment. In many cases, for instance, the experimenter is free to choose the levels of the variable as well as the number of observations at each level. He may even be free to make these decisions as the experiment progresses.

In general, the choice of design is closely related to the contemplated analysis. Consider three different methods of analysis. One widely adopted method, known as probit analysis (Finney, 1952), works well when a large number of observations is made at each of a limited number of closely spaced levels, particularly if these happen to cover the 50% point. Fisher's (1922) approach to dilution series is suggested for use with a moderate number of dilutions and relatively few observations at each. It requires, however, that successive dilutions be in a constant ratio throughout the series. The up-and-down method of Dixon and Mood (1948) requires the levels to be chosen in advance, but lets the number per level be determined by the experiment. As they describe it for testing the sensitivity of explosives to impact, if a sample tested at one level explodes, a new sample should be tested at the next lower level. If there is no explosion, the new test should be made at the next higher level. (Examples of the use of these methods are listed following the Bibliography.) Finney (1964, esp. pp. 551-553) provides recommendations on additional methods, not mentioned above.

There will, of course, be many legitimate examples of the sensitivity problem for which none of the above experimental designs is desirable, or even possible. A particularly striking case is a variation of the insecticide experiment described above. If the researcher is concerned with the effective level of insecticide involved, he may measure the quantity in each insect at the end of the experiment. Thus, he has no more than a crude control over the levels of the variable, and only a small chance of making more than one observation at any of the levels attained. In such a case, one of the methods mentioned above, namely probit analysis, can still be applied (see Bliss, 1938; Finney, 1947, 1952, sec. 43), but the results should be treated with caution, as we will see later.

More generally, a sensitivity experiment of a preliminary nature may require wide coverage with a minimal number of observations involved. Consequently, a method of analysis specifically designed to deal with one observation at a level and allowing arbitrarily spaced levels would fill a significant gap left open by the collection of techniques referred to above.

There is another limitation on the previously mentioned approaches, namely that the variable of interest is assumed to be continuous. This makes it easy to assume a continuous effect of the variable on the response probabilities, and, with a little more effort, the existence of an exact (and unique) 50% point. There is, however, nothing in the design of the sensitivity experiment that prohibits the use of an essentially discontinuous variable. Since only discrete values are used, all that is required is for the variable to have a progressive effect on response probabilities. One example from psychology is an experiment on memory span. The subject hears a list of digits read and is asked to repeat the list. The variable here is the length of the list. The two possible categories of response are a repetition and an error. In this experiment, it is very unlikely that a 50% point exists. Nonetheless, something like a 50% point would be very useful in summarizing the data.

Since the lengths are labeled numerically, it would be easy to pretend that a continuous variable exists, and then apply some standard technique. Statements such as "the subject was able to repeat a list of 6.76 digits with probability .5" could then be expected to follow routinely. There is, however, an approach which does less violence to one's sensibilities, namely, the use of confidence limits.

An important part of the sensitivity problem is indicating the quality of the estimate. An accepted means of so indicating is with confidence limits. Confidence limits are calculated from the data in a way which guarantees the value being estimated will fall between them with some given probability (.95 is a favorite value) upon successive repetitions of the experiment. This does not mean that the true value has different locations with different probabilities. It is considered to be fixed. Instead, it means that the confidence limits change with each repetition of the experiment and that 95% of the time (for 95% limits) the true value will fall between them. Moreover, this guarantee holds regardless of the location of the true value.

Even when there is no true value, as in the last example, it may still be possible to find modified confidence limits. Suppose, on the basis of memory span data we have collected, we choose two list lengths, L and U . L is chosen so that it is likely the subject could repeat lists of length L or shorter more than half the time. U is chosen so that it is unlikely the subject could repeat lists of length U or longer more than half the time. If "likely" and "unlikely" were formalized satisfactorily, L and U could then be considered as confidence limits for the 50% point, even though no 50% point can be realized. A relevant statement about the experiment then might be: "With probability .95, errors predominate for lists of ten or more digits, while a repetition is the more likely event with lists of three or fewer digits."

A framework has now been provided for a new approach to the sensitivity problem. This approach, to be described below, is designed to handle one observed response per level of a possibly discrete variable, and to deal directly in terms of confidence limits for the 50% point. For reasons which will become clear, we call it the countback method.

Returning to the first example, suppose that one response has been obtained for each of several intensities of the light flash. The first step is to single out the lowest level at which the flash was seen and the highest level at which it was not. These will be referred to as the low and high levels.

At this stage, we could take the average of these two levels and use the result as an estimate of the 50% point. This technique was first described by Gaddum (1933). It has since been called the method of extreme effective doses. As he proposed it, however, there are drawbacks which severely limit its use. Most importantly for this discussion, there is no way given to describe the quality of the estimate. In addition, the estimate may be badly biased unless the levels used are symmetrically related to the response probabilities. These facts led Finney (1964) to advise that the method of extreme effective doses never be used.

While rejecting the estimate, one should not ignore the importance of the low and high levels. As discussed above, there is some value in bypassing the estimation of the 50% point entirely and going directly to a treatment of quality, via confidence limits. It is in this approach that the low and high levels play a more appropriate role. In fact, for any integer c, we may take L to be c levels below the low level and U, c levels above the high level. (Fig. 1 provides an example of this procedure with $c = 4$.)

There is one exception to this rule which is worth mentioning now. This is the case of perfectly consistent data. Here there is no overlap between the two responses, so that the high level occurs one step below the low level. In this situation, we take extra care and use the level $c + 1$ steps below the low level for L, as well as $c + 1$ steps above the high level for U. To put it another way, L should be the level c steps below the <u>high</u> level and U, c steps above the <u>low</u> level.

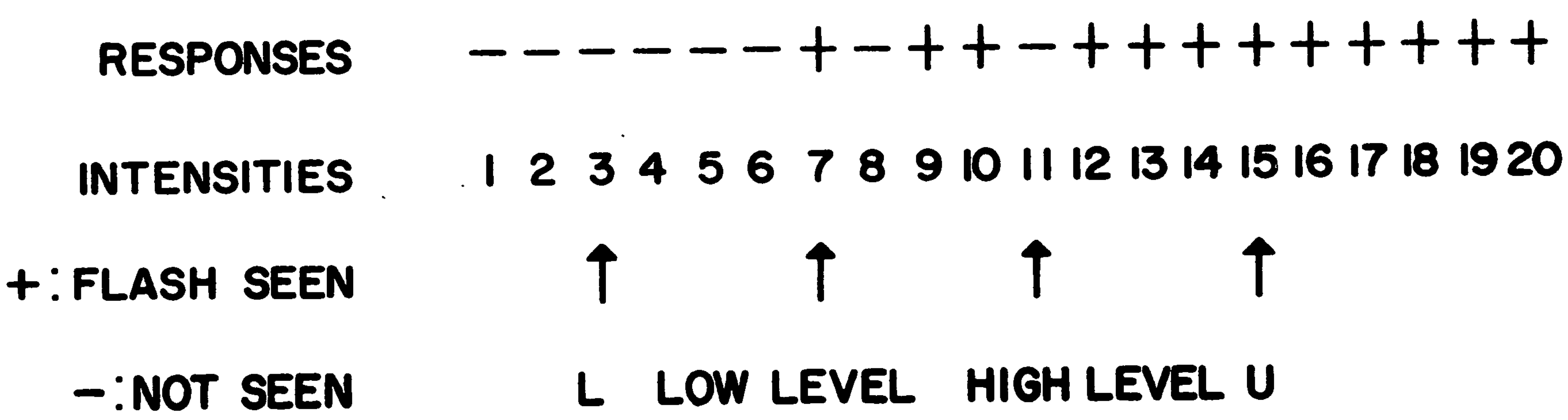

Fig. 1. Application of the countback method to imaginary data from a brightness threshold experiment. Found with $c = 4$, the confidence limits L and U have an associated confidence guaranteed to be greater than $1 - 2^{-4} = .9375$.

These are the confidence limits as determined by the countback method. Now, what is it we're confident of as regards these limits? We are confident that, for intensities of level L and below, the probability that the subject will see a flash is less than .5; while, for those of U and higher, this probability is greater than .5. Next, what proportion of the time will our confidence be misplaced? The degree of confidence associated with L and U cannot be specified exactly without introducing restrictive assumptions. With very modest prerequisites, however, it can be proved that this value will always be greater than $1 - 2^{-c}$. Thus, as noted in Figure 1, counting four levels away from the low and high levels guarantees a confidence of more than $1 - 2^{-4} = 15/16 = .9375$. Table 1 shows how this value increases with c.

While the prerequisites for this result may be modest, it is important to make them explicit. The first requirement is that the observations be statistically independent. This is a standard assumption which allows us to write the probability of any set of responses as the product of the individual response probabilities. Intuitively, it means that response probabilities are not influenced by previously observed responses, and that no common influence acts on any group of responses.

The second requirement for the countback method is that there be something like a unique 50% point. Specifically, for all levels of the variable above some point, one response (henceforth referred to as the positive response) should predominate. For all levels below, the other (negative) response should be more likely. It is not, of course, necessary for the experimenter to know where this point is, but merely

Table 1

Lower Bound on Confidence for Countback Limits as a Function of c , the Number of Levels Counted Back

Levels Counted Back c	Lower Bound $1 - 2^{-c}$
1	.5000
2	.7500
3	.8750
4	.9375
5	.9688
6	.9844
7	.9922
8	.9961
9	.9980
10	.9990

that such a point exists. Intuitively, one may see this as requiring that the levels of the variable be ordered in a way which is meaningful for the response being studied and that the variable exert a significant influence on the response.

Intuition has been made explicit for both these prerequisites because it is contemplated that they should be satisfied at an intuitive level. A satisfactory test of either assumption could probably not be made without considerably more data collection than the basic experiment would require. Thus, it is the responsibility of the experimenter who uses the countback method (or any of the others mentioned, for that matter) to design his experiment in such a way that these two assumptions may reasonably be made in his case.

Taken together, the levels of the variable actually used in an experiment may be referred to as the experimental series. Once a potential user of the countback method has satisfied the two general prerequisites discussed above, he must still select an experimental series. The series should have two characteristics for satisfactory results using this technique. First, levels at the high end of the series should evoke a positive response almost invariably, while those at the low end should be similarly related to negative responses. Another way of saying this is that the experimental series should cover the range for which either response is a distinct possibility.

Second, the experimental series should cover this range reasonably well. In particular, once a confidence level (and the corresponding c) has been chosen, it would be foolish to use an experimental series whose length was less than $2c$, and a length of $5c$, as in Figure 1, is acceptable, but not at all extreme. (Here, length refers to the number of levels in a series.)

Both of these recommendations are acceptable at an intuitive level. There is, however, a technical consideration which makes them important. Going back to Figure 1, suppose that a positive response has been obtained for intensity number 3. Then, with $c = 4$, we would run out of levels while counting down to locate L. In such a situation, the theory stipulates that L be set at an artificial absolute lower limit for the variable, designated, for instance, by $-\infty$. The idea is that we have insufficient experimental information to locate L and we indicate this by writing $L = -\infty$. At this value, L is considered to be a perfect (though useless) lower confidence limit. Similar remarks apply to U and the arbitrary upper limit, $+\infty$.

This strange situation is a result of the fact that practically nothing has been assumed about the response probabilities at different levels. Consequently, only a reasonable amount of consistent data can establish nontrivial locations for L and U. The two recommendations made earlier simply put this observation in concrete terms.

This is the countback method for constructing confidence limits. Now that we know how to use it, as well as the minimum to expect from it, the next step is to see how the countback method actually works. For this purpose, a few interesting examples have been selected and will be studied in some detail.

Earlier it was noted that other approaches were not designed to handle data taking the form of one observation per level. Nonetheless, one of these methods, probit analysis, can be used with such data. More generally, the principle underlying probit analysis, namely maximum likelihood, can be applied to the case of one observation per level. As a consequence, some form of comparison between the countback method and traditional approaches is possible. This comparison is valuable for putting the countback method in perspective and is made in the forthcoming examples.

To use maximum likelihood, a specific model relating response probabilities to levels of the variable must be assumed. A model which is often used (in probit analysis and the up-and-down method, for instance) is the Gaussian, or normal cumulative distribution. In the present case, the similar, but computationally simpler, logistic cumulative distribution was used instead. Formally, then, the probability of a positive response at level x was assumed to be $(1 + \exp(-1.7a(x - b)))^{-1}$. (This form was adapted from that used by Lord and Novick, 1968, p. 400.) Here $\underline{a}$ and $\underline{b}$ are parameters to be estimated (via maximum likelihood) from the data. One of the parameters, $\underline{b}$, is the 50% point and is of primary interest while the other, $\underline{a}$, is related to the slope of the function when $x = \underline{b}$ and is used mainly to provide a better fit to the data.

One of the primary virtues of maximum likelihood estimates is that they are asymptotically efficient and normally distributed. In practical terms, asymptotic efficiency means that no other estimates are better, given that the sample of data used is sufficiently large.

Moreover, normally distributed estimates allow for a straightforward way of obtaining confidence limits for the parameter they are estimating. Upper and lower 95% limits, for example, are found by adding to and subtracting from the estimate 1.96 times the standard deviation of the estimate. (In our case, we use the standard deviation of an efficient estimate, and this gives us, asymptotically, the best confidence limits for the parameter.)

These limits, of course, are only valid asymptotically and their actual confidence is not specified for small samples. This is an important part of the reason that maximum likelihood has not been considered applicable to sensitivity data with one observation per level. Consequently, we should not be surprised by the behavior of these limits in the examples to be considered.

In comparing maximum likelihood and the countback method, two artificial situations and one informal experiment were used. The experiment was one in which perception of depth in pictures was studied. It should be noted that this experiment was performed primarily to obtain realistic sample data. Many of the controls essential to a serious study of the problem were ignored and the results should be seen in that light.

Subjects were shown pictures (such as the one in Fig. 2) of two rectangular panels at different angles to the picture plane and were asked to indicate which of the panels appeared wider. The design of the experiment provided that each subject would see four experimental series (randomly interspersed) and make one observation per level in each. A series was comprised of pictures showing a standard panel, with fixed width and slant, and a comparison panel with fixed slant and one of 20 widths. Thus, each series had 20 levels. The 50% point of such a series is the width at which the comparison appears as wide as the standard. Having four series allowed estimation of the 50% points for four different

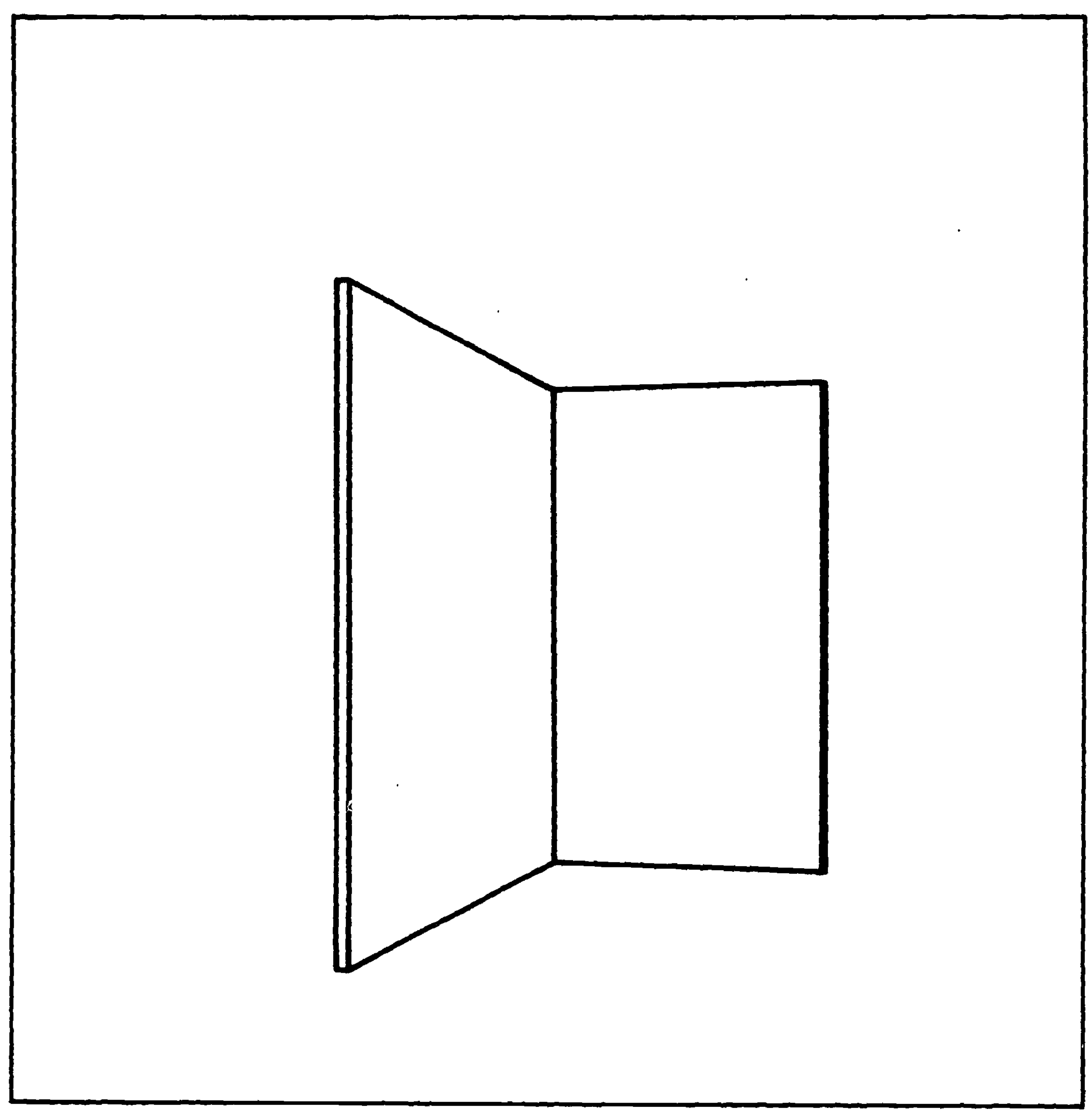

Fig. 2. Example of the pictures viewed by subjects in the experiment.

pairs of slants with each subject. There were 17 subjects, two of whom were used twice, so this experiment provided a total of 76 data sets of 20 observations each on which to apply the two methods of analysis.

Using maximum likelihood, nominal 95% and 99% confidence limits were obtained for each data set. Using the countback method, limits were found for $c = 1, 2$, and 3. These limits are guaranteed to contain the 50% point with probabilities greater than .5, .75, and .875, respectively.

One way of comparing these results is to look at the lengths of the confidence intervals obtained. These are given by $U - L$. To give some idea of the scale, it is worth noting that adjacent levels (widths measured on the plane of the picture) differed by .05 inches throughout all four series used. The median length of all 95% confidence intervals provided by maximum likelihood is .1721 inches, comparable to the median length of intervals from the countback method with $c = 1$, which is .15 inches. Moreover, 99% maximum likelihood intervals have a median length of .2262 inches, which compares with .25 inches for the countback method with $c = 2$. The median length for $c = 3$ is .35 inches (see Table 2).

In evaluating these comparisons, it should be realized that most of the data from this experiment were quite "good." That is, most of the data sets were either perfectly consistent or involved only a single reversal of positive and negative responses. Thus, our conclusions should be limited to this extent. For more complicated patterns of responses, maximum likelihood generally improves somewhat, relative to the countback method. For instance, consider the data in Figure 1. When $c = 1$, L is 6 and U is 12. The 95% limits are just above 6 and just below 11. For $c = 2$, L becomes 5 and U, 13. The 99% limits are

Table 2

Median Lengths of the Confidence Intervals in the Experiment, as Obtained by Maximum Likelihood and the Countback Method

Maximum Likelihood		Countback Method	
95%	.1721	c = 1	.1500
99%	.2262	c = 2	.2500
		c = 3	.3500

just above 5 and just below 12. Thus, in this example, maximum likelihood intervals are shorter by about one step in each case.

This improvement is particularly marked, of course, when counting back goes beyond the range of the experimental series. On the other hand, when maximum likelihood provides limits outside this range, strong assumptions are necessary to justify these limits.

The data from this experiment also provide another possibility for comparing the two methods. When the experiment was designed, the hypothesis in mind was that most people see some depth in pictures, but not to the extent that the cues of linear perspective would predict. Thus, it was hoped that two hypotheses, which I shall call 2D and 3D, could each be rejected for most of the subjects. In terms of the data, these hypotheses predicted the location of 50% points for each of the four series used. The 3D 50% points fell outside the range for all four series, while the 2D 50% points were just outside (on the other side) for two, and well within for the other two. The ranges were chosen this way to provide an efficient use of the observations, on the assumption that most of the people would produce 50% points quite different from those for either 2D or 3D.

The hypotheses were meant to apply to the overall behavior of subjects, rather than to their responses on each of the series separately. Thus a criterion had to be chosen for deciding if either hypothesis applied to a given subject. The hypotheses predict 50% points and the data give rise to confidence intervals. If at least three of the four predictions lay outside the corresponding intervals, the relevant hypothesis was rejected. This was felt to be fairly stringent, while allowing for some odd responses.

Applying this criterion to the data and using the four sets of confidence limits mentioned earlier gave the following results (pretending that there were 19 subjects, rather than 17 with two repeats). The 95% limits of maximum likelihood rejected both 2D and 3D for 17 subjects, with one subject classified as a possible 2D and one as a possible 3D. Expanding the limits to 99% resulted in two more subjects classified as possibly 2D.

The countback method, with $c = 1$, gave results almost identical to those of 95% maximum likelihood. The only difference is that one additional subject was possibly 3D. When c was set equal to 2, only 13 subjects were neither 2D nor 3D, with three possibly 2D and three possibly 3D. The two 3D subjects not so classified by the 99% limits were actually the same subject. Consequently, one could say there is agreement to within one subject, both between 95% and $c = 1$ and between 99% and $c = 2$, maximum likelihood being more powerful than the countback method in each case (see Table 3).

We can now look at two artificial situations for added information about the two methods. The experiment allowed a study of the relative lengths of the confidence intervals, as well as of the relative ability to reject alternative hypotheses. An important question remains, however, as to the actual confidence coefficients associated with the limits these two methods provide. (We have spoken of confidence coefficients before, but without giving them a name. The confidence coefficient for a set of limits is the probability that these limits contain the true value: .95, for example, for actual 95% limits.) The countback method, it will be remembered, states only an absolute lower bound for its

Table 3

Results of Testing 2D and 3D Hypotheses with Maximum Likelihood (ML) and Countback (CB) Confidence Limits

Method	Number of Subjects		
	Possibly 2D	Not 2D or 3D	Possibly 3D
95% ML	1	17	1
CB, c = 1	1	16	2
99% ML	3	15	1
CB, c = 2	3	13	3
CB, c = 3	3	12	4

confidence coefficients. Maximum likelihood, on the other hand, gives coefficients based on large samples drawn from the population specified by its model.

For purposes of analysis, the main thing that distinguishes artificial situations from actual experimental ones is that, in the former, we choose the response probabilities and, thus, know them exactly. Consequently, when we apply a method of analysis to an artificial situation, we know whether or not the assumptions of the method are being met.

One way of studying an artificial situation is using it to generate artificial data. These data can be analyzed, confidence limits obtained, and the limits can be checked to see if they actually contain the 50% point (since we know what the 50% point is in an artificial situation). Repeating this procedure for many sets of data generated in the same way eventually provides a good estimate of how well the confidence limits are working. Specifically, the proportion of times the limits contain the 50% point is an estimate of the confidence coefficient for these limits in this artificial situation. If the limits claim to have 95% confidence, we hope the proportion would be close to .95.

Although the above approach is quite straightforward, there are other ways of studying an artificial situation which are more satisfactory when they can be used. For discussion of these and related topics, the interested reader is referred to the Simulation Appendix.

The first artificial situation was designed to see how well each method would fare when all the assumptions (except large sample size) of our version of the maximum likelihood approach were met. Consequently, the response probabilities for this situation were chosen according to the logistic model. In order to make the results of this analysis easily comparable with those of the experiment, the same 20 levels of the variable as appeared in two of the experimental series were used. Specifically, the values were .50, .55, .60, ... , 1.40, 1.45. For good measure, the 50% point was placed in the middle of this series, at .975. The performance of the maximum likelihood and countback confidence limits (i.e., how often they contained .975) was then ascertained, as the steepness of the logistic curve varied. (Fig. 3a, b, c, and d shows curves with central slopes proportional to 2.5, 5, 10, and 20, respectively. They comprise most of the range over which the methods were tested.)

In discussing the results, it is important to emphasize that the true confidence coefficients for both methods varied with the slope of the underlying logistic curve. These coefficients tended to be lowest for a middle range of slopes, improving as the slopes got either very small or very large.

The so-called 95% confidence limits given by maximum likelihood had a confidence coefficient below .91 when the logistic curve had a slope of 10. The coefficient was still below .95 for slopes of 5 and 15, but rose to .967 when the slope increased to 20. The 99% limits produced a coefficient less than .98 for a slope of 5, but were up well above .99 for slopes of 15 or more.

With c chosen as 1, the countback method guarantees only a confidence coefficient of .5. In this situation, however, it did considerably better, and was almost identical to 95% maximum likelihood. When the slope was 8, its true coefficient reached a minimum of .879. Taking

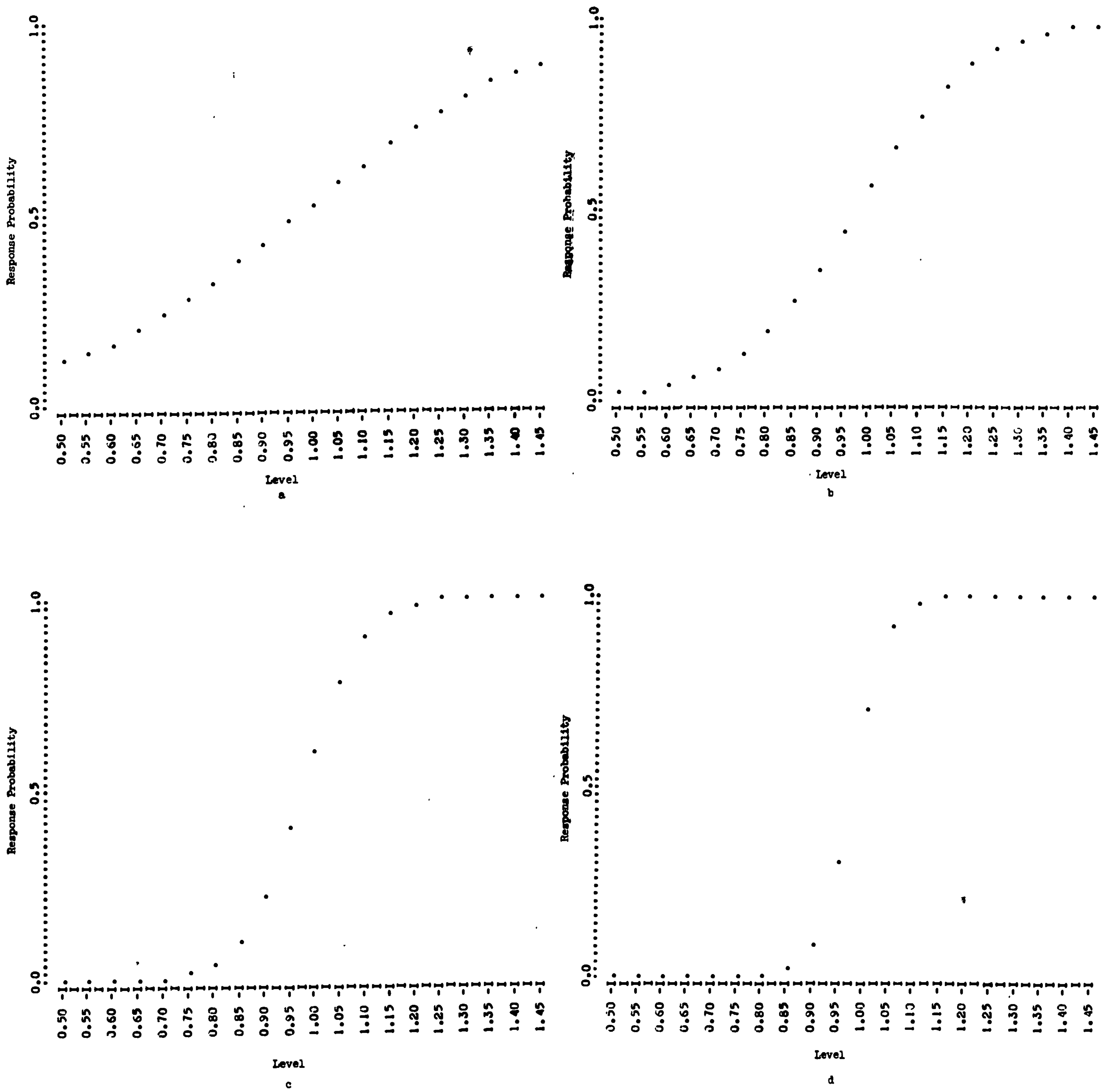

Fig. 3. Positive response probabilities for different levels as given by the logistic model, with slopes of 2.5, 5, 10, and 20.

$c = 2$ (a guaranteed confidence coefficient of .75), the minimum coefficient was .973, attained at a slope of 5. The coefficient increased to above .99 for slopes greater than 11. Finally, the coefficient for $c = 3$ never dropped below .994 (see the first part of Table 5).

It is worth summarizing and discussing these results. First, consider maximum likelihood. At worst, this approach underestimates the probability that its limits do not contain the 50% point by at least a factor of two. To see whether this difficulty was related to some special feature of maximum likelihood estimation, another method, namely minimum X^2, was used. This method is not as widely used as maximum likelihood, but has been earnestly advocated (see Berkson, 1949).

As with maximum likelihood, one must assume a specific probability model to use minimum X^2, and, for consistency, the logistic was again the model chosen. Minimum X^2 is also guaranteed to produce asymptotically efficient and normally distributed estimates as the sample size gets large, but here the relationship to maximum likelihood ends. It is an open question how well the two methods will agree for any specific small sample situation. In fact, the minimum X^2 results were virtually identical to those of maximum likelihood for the 95% limits (giving a minimum true confidence coefficient of less than .91), and only slightly better for the 99% limits (giving a minimum of less than .985).

Now let us consider the countback method. In contrast to maximum likelihood, it did much better than its claims. This fact led to the development of another theoretical result about countback confidence limits. If response probabilities are symmetric[1] around the 50% point (as those of the logistic model are), the minimum confidence coefficient for the countback is $1 - 2^{-2c}$ (instead of $1 - 2^{-c}$). Thus, for $c = 1$, symmetry guarantees a confidence of at least .75 and, for $c = 2$, at least .9375 (see Table 4 for more values).

This artificial situation has given us a better picture of the countback method than we had from the theory and experiment alone. In fact, there is more information to be gained. For each value of the slope, $\underline{a}$ (between 0 and 20), it was possible to find the median length of the countback confidence intervals. We can thus see how this length changes relative to $\underline{a}$. When $c = 3$ (the only case studied), the median interval is infinitely long when $\underline{a}$ is less than 3. For $\underline{a}$ between 3 and 8, the median length is approximately $3/a$. Above 8, the median length is the minimum possible for $c = 3$ and a spacing between levels of .05, namely .35.

Evidently, the countback intervals are most efficient for $\underline{a}$ somewhere between 3 and 8. For smaller values of $\underline{a}$, the responses are not being obtained over a wide enough range of the variable of interest. For larger values, it seems that the range is more than adequate and a finer coverage (i.e., closer spacing between levels) is desirable instead.

Now let us turn to a consideration of the second artificial situation. The confidence limits of the countback method are at their worst when there is an asymmetry in the response probabilities. Such an

---

[1]Here, symmetric means that the probability of a positive response k levels below the 50% point equals the probability of a negative response k levels above the 50% point, for $k = 1,2,3,\ldots$.

Table 4

Lower Bound on Confidence Coefficient for Countback Limits When the Response Probabilities Are Symmetric Around the 50% Point

Levels Counted Back c	Lower Bound $1 - 2^{-2c}$
1	.7500
2	.9375
3	.9844
4	.9961
5	.9990

asymmetry could involve a slow decrease in positive response probability below the 50% point and a rapid increase above this point. An extreme case of asymmetry occurs when the probability is .5 for all levels up to a point, above which the probability is 1. In this situation, the confidence coefficient for the countback method is just above $1 - 2^{-c}$.

In the second artificial situation the same levels were used as before, but the 50% point was chosen to be between the sixth and seventh levels, at .79 on the scale used in the experiment. All levels above the sixth were assigned a positive response probability of 1. The remaining response probabilities were specified with the lower half of a cumulative logistic curve. The slope of this curve varied from 0 to 20 on the same scale used before. The resulting situations, for slopes of 2.5, 5, 10, and 20, are shown in Figure 4a, b, c, and d. When the slope is 0, we have the extreme case mentioned above, which results in a step function.

In general, the confidence coefficients for all methods studied (countback, maximum likelihood, and minimum X^2) increased as the slope of the lower half of the probability curve increased. Unlike the previous situation, the confidence coefficients in this case could be evaluated exactly for all methods. Consequently, the results can be easily summarized in graphical form. (See Fig. 5 on p. 50. Table 5 also gives the results, along with those for the first situation.)

There are several things to say in connection with these results. The countback with $c = 1$ differs from 95% maximum likelihood only for small values of the slope. Maximum likelihood's 99% limits are effectively identical to the 95% limits attained by minimum X^2 , hence the

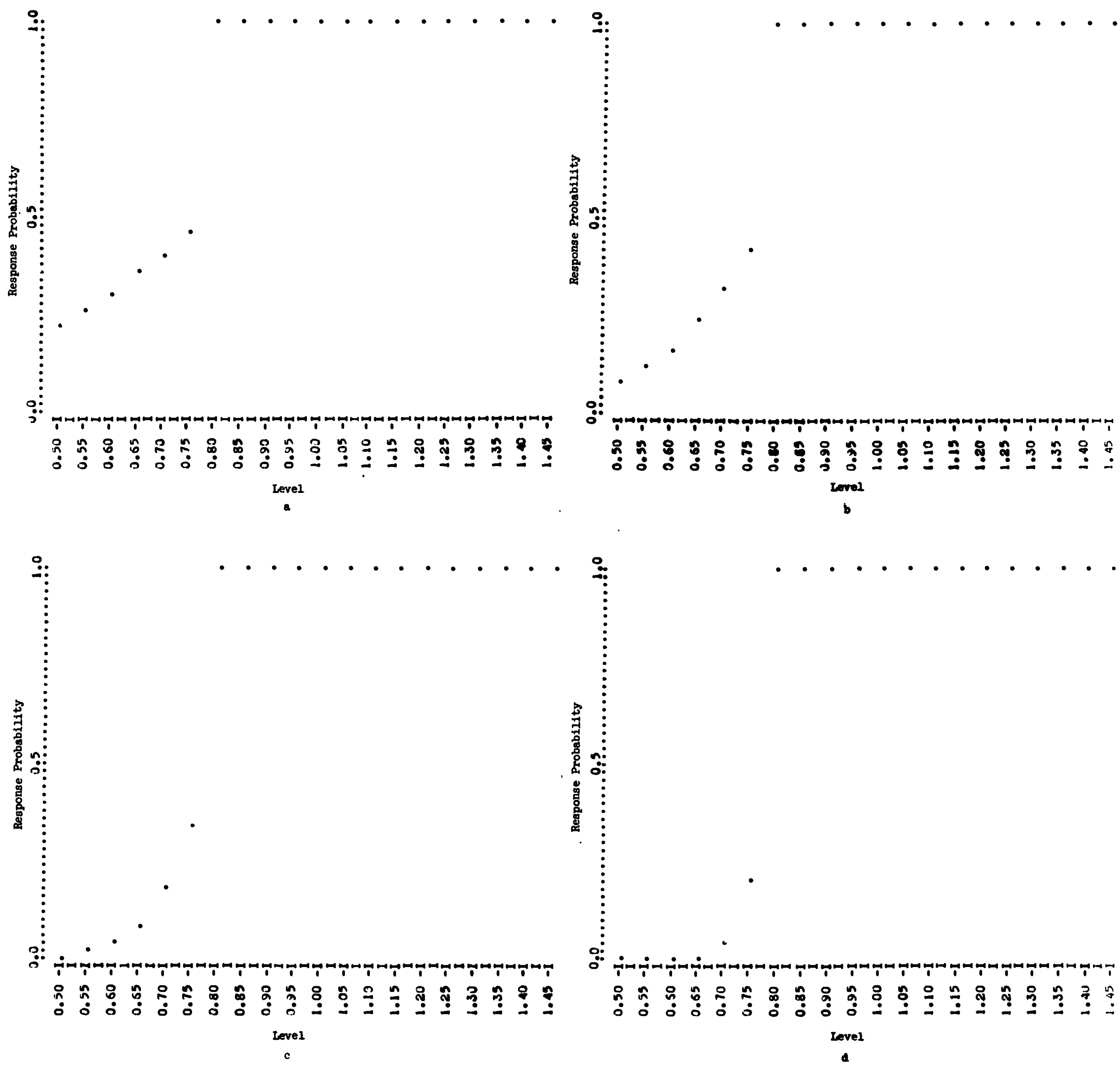

Fig. 4. Positive response probabilities for different levels as given by an asymmetric model, with lower slopes of 2.5, 5, 10, and 20.

Table 5

Confidence Coefficients for the Countback Method (CB), Maximum Likelihood (ML), and Minimum χ^2 (MC) Confidence Limits in the Two Artificial Situations (Logistic and Asymmetric)

Method	Confidence Coefficients					
	Logistic Slope			Asymmetric Lower Slope		
	5	10	15	5	10	15
CB, c = 1	.90	.89	.93	.73	.90	.967
95% ML	<.93	<.91	<.94	.73	.90	.965
95% MC	<.95	<.91	<.94	.82	.92	.969
CB, c = 2	.973	.988	.997	.94	.991	.999
99% ML	<.979	<.988	<.997	.82	.92	.969
99% MC	<.986	<.990	<.998	.958	.993	.999
CB, c = 3	.995	>.999	>.999	.990	>.999	>.999

two labels for the same curve in Figure 5. The 99% limits of minimum χ^2 are better than those of the countback with $c = 2$, but worse than with $c = 3$.

Based on the comparisons of the previous situations, the most surprising finding is the relative failure of the 99% maximum likelihood confidence limits. This reinforces the observation made earlier that maximum likelihood and minimum χ^2 need not give the same results for small samples. In this case, the minimum χ^2 , with its generally longer confidence intervals, seems to have the advantage. For example, with a slope of 9, the confidence coefficient for the 99% minimum χ^2 limits is indeed .99, while the 99% maximum likelihood limits have a coefficient of only .90.

Finally, of all the limits studied, only those of the countback with $c = 3$ come close to being "safe" for an experimenter interested in confidence levels of at least .95. Again, it should be emphasized that the present example is an extreme case, and it may be considered likely that it represents the methods at something near their worst.

Now we have seen how the countback method and two asymptotically efficient approaches (maximum likelihood and minimum χ^2) work in a few special cases. As was emphasized earlier, the latter two were designed to deal with data involving many observations per level and following a narrowly specified model of response probabilities. Consequently, the lack of correspondence between the confidence coefficients derived from these assumptions and those obtained in our cases should not be surprising. This lack of correspondence should, however, strengthen our reservations about naively applying maximum likelihood or minimum χ^2 to situations similar to those we have examined.

The countback method gains additional support in these examples from the fact that its true confidence levels exceed the theoretical minimum values by a substantial margin. Thus it is comforting to know that one does not need to count back 5 steps in order to be reasonably sure of having 95% limits, or 7 steps for 99% limits. Instead, it appears that taking $c = 2$ for roughly symmetric situations, or $c = 3$ in more extreme cases will easily provide 95% limits, and increasing c by 1 will increase the confidence to above 99%.

Finally, from the point of view of design, using as few as 20 levels (hence 20 observations) seems to be satisfactory. Moreover, if there is some knowledge of the location of the 50% point, there is no need for extremely wide coverage with these levels.

Thus a method starting with minimal assumptions about the data is able to operate effectively for an experimental design where traditional approaches are largely inappropriate. It is hoped that this development will encourage wider use of the sensitivity experiment with one observation per level and its application to new research problems.

We have studied the countback method only as it applies to data with one observation per level. There is, however, no reason why the same approach could not be used when more than one observation per level has been made. The high and low levels could still be the highest with a negative response and the lowest with a positive response, respectively. For an integer c, the lower limit L could still be c levels below the low level and the upper limit U, c levels above the high level.

The restrictions for perfect data and for artificial absolute limits could still be made.

The only difference, in fact, would be in the lower bound for the confidence coefficient associated with L and U. Suppose, for simplicity, that the same number of observations, n, has been made at each level. Then the lower bound becomes $1 - 2^{-nc}$. Table 6 gives actual values of this bound for different choices of n and c.

Table 6

Lower Bound on Confidence Coefficient

for Countback Limits with More Than

One Observation Per Level

Levels Counted Back c	Lower Bound $1 - 2^{-nc}$ $n = 2$	$n = 3$	$n = 4$
1	.7500	.8750	.9375
2	.9375	.9844	.9961
3	.9844	.9980	.9998
4	.9961	.9998	.99998
5	.9990	.99997	.999999

Statistical Appendix

This appendix is divided into two sections. The first gives the essentials of the theory associated with the countback method. The second section deals with the maximum likelihood and minimum χ^2 approaches as they apply to the logistic model and one observation per level.

The Countback Method

The main purpose of this section is to prove and expand upon assertions made earlier about the countback method of constructing confidence limits for a 50% point in sensitivity experiments. It will be remembered that these limits were defined in terms of the lowest level at which a positive response was observed and the highest level at which a negative response was observed. Once a positive integer c is chosen, the lower limit, L, is defined as the level c steps below the low level. Similarly, the upper limit, U, is located c steps above the high level.

If there are not enough levels in the experiment to allow L or U to be defined in this way, the artificial absolute limits, $-\infty$ or $+\infty$, may be used for L or U, respectively. These are assumed to be absolute lower and upper limits. They are used to keep the confidence associated with L and U within reasonable bounds, even when the data itself is unrewarding in at least one direction for the determination of a 50% point.

The other exception to the general countback rule occurs with perfect data, i.e., a string of negative responses followed by a string of positive responses. Here one counts back $c + 1$ steps rather than c. (Or, to put it another way, one counts down c steps from the high level for L and up c steps from the low level for U.) Briefly, the reason for this modification is that perfect data need not be as good as they look. Consequently, it is best to treat them with a little extra care. (This subject comes up again in the second section, where the problem is considerably more serious than at present.)

There are two assumptions associated with the countback method. The first of these is that a level of the variable under study exists with the following property: the probability of a positive response at all higher levels is greater than .5 and less than .5 at all lower levels. (This level will occasionally be loosely referred to as the 50% point.) The second assumption is that the observations are statistically independent.

With these two prerequisites, it is possible to establish absolute lower bounds on the confidence associated with L and U, with these bounds depending only on c.

In the discussion which follows, the experimental levels of the variable of interest will be designated $1,2,\ldots,N$, from the lowest level to the highest. The probability of a positive response at level i will be abbreviated as P_i, with Q_i $(= 1 - P_i)$ referring to the probability of a negative response at level i. It will also be assumed that the 50% point falls well within the series, simply because the other cases are trivial extensions of this main one and add nothing new.

Finally, the modification for perfect data will be ignored until later in the discussion.

If the 50% point lies between experimental levels $i - 1$ and i, then the assumption is that $P_j < .5$ for $j < i$ and $P_j > .5$ for $j \geq i$. Now, in order for the limits L and U to be valid, L must be one of levels $-\infty, 1, 2, \ldots$, or $i - 1$, and U must be one of levels $i, i + 1, \ldots, N$, or $+\infty$. Put another way, L and U will not include the 50% point if L is any level greater than $i - 1$, or if U is any level less than i. This latter point of view turns out to provide easier mathematics, so it is the one which has been adopted.

L will be level i if the lowest positive response occurs at level $i + c$. More generally, L will be above level $i - 1$ if and only if the lowest level with a positive response is $i + c$ or higher. This last is equivalent to saying that only negative responses occur at levels below $i + c$. The probability of such an event (by the independence assumption) is $\prod_{j=1}^{i+c-1} Q_j$. Similarly, U will be below level i if and only if positive responses occur at all levels above $i - c - 1$. The probability of this event is $\prod_{j=i-c}^{N} P_j$.

The two events just described are mutually exclusive (i.e., it is impossible to have L above the 50% point and U below it simultaneously). As a consequence, the probability that L and U will fail to include the 50% point is just the sum of the individual probabilities: $\prod_{j=1}^{i+c-1} Q_j + \prod_{j=i-c}^{N} P_j$. Subtracting this quantity from 1 gives the confidence coefficient associated with L and U.

The next step is to find the minimum value this confidence coefficient can attain, given the constraints on the P_j. This is equivalent to maximizing the above sum. Clearly, the first term of the sum is greatest when each of the Q_j is maximized $(j = 1, \ldots, i + c - 1)$. This means setting $Q_j = 1$ for $j < i$ and $Q_j = .5$ for $j \geq i$, according to the constraints. The resulting maximum is $.5^c$ since there are c levels above $i - 1$ in the product. Incidentally, the second term is 0 under this choice of probabilities since setting $Q_j = 1$ is equivalent to setting $P_j = 0$.

In the same way, the second term can be maximized at $.5^c$ by taking maximal values for the P_j, and this choice makes the first term 0. Thus the question remains as to whether the sum can attain a value greater than $.5^c$ for some compromise choice of the P_j. In answering this, it is important to note that the total probability of L and U failing is a linear function of each of the P_j. This means that the maximum value of this probability will be attained when each P_j is set at ones of its extremes.

Now let us proceed by a process of elimination. Clearly, any choice of P_j which makes either term 0 is unacceptable in trying to make the sum exceed $.5^c$. Consequently, all P_j with $j = i - c, \ldots, i + c - 1$ must be set at .5. This is because all of these are represented in both terms, once as P_j and once as $1 - P_j$ $(= Q_j)$. For lower values of j, only Q_j appears (so for these, P_j may be set at 0). Similarly, for higher levels only P_j appears, and P_j may be taken as 1. Since there are $2c$ levels for which P_j has been set at .5, the total probability becomes: $.5^{2c} + .5^{2c} = .5^{2c-1}$. This value is equal to $.5^c$ when $c = 1$ and less than $.5^c$ when $c > 1$.

The above proof is the justification for the claim that the count-back confidence coefficient is at least $1 - .5^c$ when one counts back c steps.

Now we are ready to see what happens when the modification for perfect data is made. Before, a statement was made that L will be above $i - 1$ if and only if only negative responses occur below $i + c$. This is no longer true. If only negative responses occur below $i + c$ and all higher responses are positive, L will be located at $i - 1$. Consequently, the probability of this particular set of perfect data must be subtracted from our original value to describe the new situation. The result is

$$\prod_{j=1}^{i+c-1} Q_j - \left(\prod_{j=1}^{i+c-1} Q_j\right)\left(\prod_{j=i+c}^{N} P_j\right) .$$

In the same way, the probability of U being below the 50% point is reduced to

$$\prod_{j=i-c}^{N} P_j - \left(\prod_{j=1}^{i-c-1} Q_j\right)\left(\prod_{j=i-c}^{N} P_j\right) .$$

Without going through any proof, one can see that this change can only increase the minimum confidence coefficient. The amount by which it is increased (namely $.5^N$) is hardly worth mentioning, and $1 - .5^c$ remains the practical lower bound.

It is worth noting that the perfect data modification was made to help the method in typical rather than extreme cases. The minimum confidence coefficient, after all, is attained when there is perfect discrimination on one side of the 50% point, and perfect confusion on the other side. This may be thought of as the maximally asymmetric case. Consequently, we may be led to ask what happens when the response probabilities are symmetric about the 50% point. In one context, this may be taken to mean that $Q_{i-1} = P_i$, $Q_{i-2} = P_{i+1}$, and so on. This, in turn, implies that

$$\prod_{j=i-c}^{i+c-1} Q_j = \prod_{j=i-c}^{i+c-1} P_j \ ,$$

given that $c + 1 \leq i \leq N - c + 1$. For ease of notation, let us set

$$\prod_{j=1}^{i-c-1} Q_j = A \ , \quad \prod_{j=i-c}^{i+c-1} Q_j = B \ , \text{ and } \quad \prod_{j=i+c}^{N} P_j = C \ .$$

In these terms, the probability that L and U will fail to include the 50% point becomes AB + BC without the perfect data modification and AB - ABC + BC - ABC with it. These can be rewritten as B(A + C) and B(A + C - 2AC) , respectively. Because of the symmetry, we always have $B \leq .5^{2c}$. In addition, A + C can be as great as 2, while A + C - 2AC must always be less than 1. Consequently, when the 50% point is well within the experimental series $(c + 1 \leq i \leq N - c + 1)$, symmetry increases the lower bound for the confidence coefficient from $1 - .5^c$ to $1 - .5^{2c-1}$ in the unmodified case and to $1 - .5^{2c}$ after the modification. These limits are listed in Table 7.

Finally, when the 50% point is near either end of the series, symmetry becomes a less important constraint and the bound drops accordingly. In the extreme, when the 50% point is outside the series, symmetry has no value and the bound returns to $1 - .5^c$.

Maximum Likelihood and Minimum χ^2 Logistic Analysis

These approaches differ from the countback method in that they require a specific model of response probabilities. In the present work, the cumulative logistic distribution was taken as the model. If x is

Table 7

Effect of Perfect-Data Modification on

Lower Bound for Confidence Coefficient

in the Symmetric Case

Levels Counted Back	Lower Bound	
	Without Modification	With Modification
c	$1 - .5^{2c-1}$	$1 - .5^{2c}$
1	.5000	.7500
2	.8750	.9375
3	.9688	.9844
4	.9922	.9961
5	.9980	.9990

the value of the variable at any level, it provides a value for the probability of a positive response at that level by $P = (1 + \exp(-1.7a(x - b)))^{-1}$. In this context, maximum likelihood and minimum χ^2 are two ways of fitting this function to data by providing suitable choices for *a* and *b* .

Let us first concentrate on maximum likelihood. Suppose we define $r = 1$ for a positive response at any level and $r = 0$ for a negative response. Then the probability of a given data set is $\Pi P^r(1 - P)^{1-r} = L$, where the product is taken over all experimental levels. Maximum likelihood estimation finds *a* and *b* which maximize L . Usually (and our case is no exception) this means solving the equations $\partial \log L/\partial a = 0$ and $\partial \log L/\partial b = 0$. (Taking logarithms does not change the location of the maximum and makes things mathematically simpler as well as numerically better behaved.)

After some rearrangement, these equations become $-1.7\,\Sigma(r - P)(x - b) = 0$ and $1.7a\Sigma(r - P) = 0$, which are equivalent to $\Sigma r = \Sigma P$ and $\Sigma rx = \Sigma Px$. These equations were solved numerically with the help of a modified Newton-Raphson method. The choice of starting values for *a* and *b* was important, but, given a good choice, a wide range of data could be fitted to the limits of computational accuracy, usually within six iterations.

For the minimum χ^2 method, a function other than the likelihood is important. As implied by the name, this approach finds *a* and *b* which minimize the usual measure of goodness of fit, namely

$$\chi^2 = \Sigma\left[\frac{(r - P)^2}{P} + \frac{((1 - r) - (1 - P))^2}{(1 - P)}\right] = \Sigma\,\frac{(r - P)^2}{P(1 - P)} \quad .$$

In a manner analogous to maximum likelihood, this means solving the equations $\partial \log \chi^2/\partial a = 0$ and $\partial \log \chi^2/\partial b = 0$ for _a_ and _b_. Written out and simplified, these equations are

$$(1.7/\chi^2)[\sum_{r=0} (x - b)e^{y} - \sum_{r=1} (x - b)e^{-y}] = 0$$

and

$$(-1.7a/\chi^2)[\sum_{r=0} e^{y} - \sum_{r=1} e^{-y}] = 0 \quad ,$$

where $y = 1.7a(x - b)$ and χ^2 simplifies to $\sum_{r=0} e^{y} + \sum_{r=1} e^{-y}$. Again, a modified Newton-Raphson procedure was applied, with much the same success as before.

Having obtained estimates of the 50% point (_b_ in our logistic model), the next concern should be the quality of these estimates. Both minimum χ^2 and maximum likelihood estimates can be shown to be asymptotically normally distributed around the true values with a variance-covariance matrix given by $-D^{-1}$ (Kendall & Stuart, 1967). Here D is the expected value of the matrix of second-order derivatives of the log likelihood, evaluated at the true values of _a_ and _b_. Thus, for instance, the element $d_{12} = E(\partial^2 \log L/\partial a \partial b)$.

To pass from this information to nominal confidence limits requires quite an act of faith. Nonetheless, it is possible to evaluate the second-order derivatives at the estimated values of _a_ and _b_ and, from them, follow the theory to a matrix, S, analogous to the asymptotic variance-covariance matrix. Then s_{22} is analogous to the variance of our estimate for _b_ and, for instance, "95%" confidence limits can be generated by subtracting and adding $1.96\sqrt{s_{22}}$ to the estimate. It is, of course, a completely open question what the actual confidence coefficient will be for these limits in any given situation. It was partly to study this question that the work described in the Simulation Appendix was carried out.

Examples of computer output for selected data are found in Table 8. Here, as in the simulations and the experiment, there were 20 levels of the variable, from .50 to 1.45 in steps of .05. The output titled MAXILIK gives the results of applying maximum likelihood to the data, while MINICHI gives the results of applying minimum χ^2 . Data appear following question marks, 0 for a negative response and 1 for a positive response. Below each data set are listed EST., the estimate, S.E., the standard error of estimate ($\sqrt{s_{22}}$, based on the asymptotic results), and DER., the final derivative of the relevant function (log likelihood or log χ^2), for the slope (A) and 50% point (B). Finally, "95%" and "99%" confidence intervals (C.I.) for the 50% point are given in parentheses.

Comparing the results of the two programs for identical data suggests some generalizations. First and foremost, estimates for <u>b</u> , the 50% point, agree very closely. Second, minimum χ^2 provides a smaller estimate of the slope. (See Little, 1968, for a theoretical treatment of this point.) Third, maximum likelihood gives shorter confidence intervals. (Considering the simulation results, this should be seen as a failing rather than a virtue.)

One more problem needs to be discussed. What happens to these methods with perfect data (i.e., data consisting of a string of negative responses, followed by a string of positive responses)? Both maximum likelihood and minimum χ^2 find the best (indeed,perfect) fit to

Table 8

Examples of Output from Computer Programs Used to Find Maximum Likelihood (MAXILIK) and Minimum χ^2 (MINICHI) Confidence Limits

MAXILIK

```
? 0,0,0,0,0,0,0,0,0,1,0,1,1,1,1,1,1,1,1,1
   EST.      S.E.           DER.
A 15.413     9.72732        -1.77666 E-7
B 0.975      4.36839 E-2    -4.5755 E-8
C.I. FOR B: ( 0.88938 , 1.06062 )  ( 0.86247 , 1.08753 )

? 0,0,0,0,0,1,0,1,1,1,1,1,1,1,1,1,1,1,1,1
   EST.      S.E.           DER.
A 15.398     9.75583        -1.60502 E-6
B 0.774986   4.37133 E-2     1.29816 E-6
C.I. FOR B: ( 0.689308 , 0.860664 )  ( 0.662381 , 0.887591 )

? 0,0,0,0,0,0,0,1,0,0,1,1,1,1,1,1,1,1,1,1
   EST.      S.E.           DER.
A 10.7731    5.70744        -1.53066 E-6
B 0.924988   5.22584 E-2    -1.7147 E-6
C.I. FOR B: ( 0.822562 , 1.02741 )  ( 0.790371 , 1.05961 )

? 0,0,0,0,0,0,0,0,0,0,1,0,1,0,1,1,1,1,1,1
   EST.      S.E.           DER.
A 8.72058    4.23208        -6.28525 E-9
B 1.07517    5.81569 E-2    -4.24733 E-6
C.I. FOR B: ( 0.961179 , 1.18915 )  ( 0.925354 , 1.22498 )

? 0,0,0,0,0,1,1,0,1,0,1,1,1,1,1,1,1,1,1,1
   EST.      S.E.           DER.
A 6.55659    2.96189        -6.9328 E-7
B 0.823254   6.77866 E-2     1.60414 E-6
C.I. FOR B: ( 0.690392 , 0.956116 )  ( 0.648636 , 0.997872 )
```

Table 8 (Contd)

MINICHI

? 0,0,0,0,0,0,0,0,0,1,0,1,1,1,1,1,1,1,1,1

	EST.	S.E.	DER.
A	11.3192	6.13501	-6.04948 E-8
B	0.975	5.09778 E-2	-1.16612 E-6

C.I. FOR B: (0.875083 , 1.07492) (0.843681 , 1.10632)

? 0,0,0,0,0,1,0,1,1,1,1,1,1,1,1,1,1,1,1,1

	EST.	S.E.	DER.
A	11.2691	6.22063	-3.25245 E-10
B	0.774958	5.11698 E-2	-1.27135 E-6

C.I. FOR B: (0.674665 , 0.875251) (0.643145 , 0.906771)

? 0,0,0,0,0,0,0,1,0,0,1,1,1,1,1,1,1,1,1,1

	EST.	S.E.	DER.
A	7.39583	3.3522	-2.95945 E-9
B	0.913384	6.32176 E-2	-7.53317 E-7

C.I. FOR B: (0.789478 , 1.03729) (0.750536 , 1.07623)

? 0,0,0,0,0,0,0,0,0,0,1,0,1,0,1,1,1,1,1,1

	EST.	S.E.	DER.
A	5.84153	2.51411	-2.72263 E-10
B	1.07545	7.17866 E-2	-4.23514 E-7

C.I. FOR B: (0.934751 , 1.21615) (0.89053 , 1.26037)

? 0,0,0,0,0,1,1,0,1,0,1,1,1,1,1,1,1,1,1,1

	EST.	S.E.	DER.
A	4.32084	1.86446	-1.90648 E-9
B	0.831057	8.67783 E-2	-1.40864 E-7

C.I. FOR B: (0.660972 , 1.00114) (0.607517 , 1.0546) perfect data when $1/a = 0$. If $1/a$ really were 0, only perfect data centered at the 50% point would have any chance of occurring. Thus the variance of the estimate of the 50% point would certainly be 0. This, however, leads us to the unacceptable conclusion that the length of the confidence intervals for perfect data should be 0: that we should have complete confidence in our estimate. Clearly, another approach must be taken.

On intuitive grounds, one might want the confidence interval for perfect data to be at least as short as that for any other data which give the same estimate for $\underline{b}$. In fact, the nonperfect data with the shortest associated confidence interval have a single reversal in the response pattern. Explicitly, single reversal data have a string of negative responses, followed by a positive response, a negative, and a string of positives, in that order.

Consequently, a conservative procedure in line with the intuition expressed above is to treat perfect data as though they were single reversal data, centered at the same place. For lack of any better idea, this is what has been done throughout the present study, both in the simulations and for the experimental data. It cannot be emphasized too strongly, however, that this is an unsolved problem, and an important one because of the relatively high likelihood of obtaining perfect data. (See Table 9 for the probability of perfect data with the probit model used in the simulation.)

Table 9

The Probability of Obtaining Perfect Data with the Logistic Model Used in the First Simulation (20 Equally Spaced Levels, Centered on the 50% Point)

Slope of Logistic a	Probability of Perfect Data
2.5	.0071
5.0	.0872
7.5	.2455
10.0	.4069
12.5	.5405
15.0	.6443
17.5	.7236
20.0	.7842

Simulation Appendix

Before discussing the details of the simulations, it would be worthwhile to describe the models used. For the standard situation, the logistic model provided values for P, the probability of a positive response at a given level of the variable: $P = (1 + \exp(-1.7a(x - b)))^{-1}$. In this equation, x is the value of the variable, $\underline{b}$ is the 50% point, and $\underline{a}$ is related to the slope of the logistic curve when $x = b$. (The constant factor 1.7 has been added to make $\underline{a}$ roughly comparable to $1/\sigma$ in the normal distribution function.)

To allow for comparisons with the accompanying experiment, twenty levels were used, with x varying from .50 to 1.45 in steps of .05. In the standard situation, $\underline{b}$ was placed in the middle of this series, at .975, and $\underline{a}$ was varied between 2.5 and 20.

The extreme situation differed from this in only two respects. First, $\underline{b}$ was changed to .79. Second, all values of P corresponding to $x > .79$ were taken to be 1, instead of the value given by the logistic. In other words, this situation operates on a logistic model below the 50% point, and on a deterministic (or errorless) one above it.

Three kinds of simulation were applied to the above situations. In the first kind, sets of 20 independent values (each uniformly distributed between 0 and 1) were obtained from a random number generator. If the i-th random number in a set was less than P for level i, it was taken to represent a positive response (and otherwise a negative one).

In this way, sets of artificial data could be produced based on either situation and any choice of $\underline{a}$. Moreover, these data could be analyzed by any of the three techniques to see whether or not _b_ lay within the resulting confidence intervals.

In fact, this approach was used only for the standard situation and then only with the maximum likelihood method. Five values of _a_ were chosen (2.5, 5, 10, 15, and 20) and 1,000 sets of numbers were generated for each value. The performance of both "95%" and "99%" confidence limits was observed for each of the samples. Finally, 99% confidence limits based on the binomial distribution (Wilks, 1962) were calculated for each estimate of the confidence coefficients. (The results are given in Table 10.)

The second kind of simulation involved a partial or complete enumeration of the possible response sets. In the extreme situation there are only 64 possible sets, since $P = 1$ for all but the six lowest levels. Thus, these 64 could each be analyzed, and those sets giving rise to intervals not including the 50% point could be singled out. For any choice of _a_, the probability of obtaining each of these sets could be calculated directly. Thus, an exact value could be obtained for the confidence coefficient of limits generated by any of the three techniques. In fact, this approach was used only for maximum likelihood and minimum χ^2. The results are shown in Figure 5. (Note that the levels for "99%" maximum likelihood limits are identical to those for "95%" minimum χ^2 limits. This is a consequence of the fact that these two sets of limits failed to include the 50% point for precisely the same sets of data.)

Partial enumeration was used for the standard situation. The lowest five responses were fixed as negative and the highest five fixed as positive. All combinations of the middle 10 responses were then taken,

Table 10

Estimates and 99% Confidence Limits for Maximum Likelihood (ML) Confidence Coefficients, Based on Samples of 1000 Data Sets, in the Standard Situation (Logistic Model)

Slope	Confidence Coefficient					
a	"95%" ML			"99%" ML		
	Lower Limit	Esti-mate	Upper Limit	Lower Limit	Esti-mate	Upper Limit
2.5	.9249	.9450	.9619	.9720	.9840	.9924
5.0	.8798	.9050	.9275	.9618	.9760	.9867
10.0	.8742	.9000	.9231	.9774	.9880	.9950
15.0	.9086	.9310	.9501	.9844	.9930	.99796
20.0	.9520	.9680	.9806	.9947	.9990	.999995

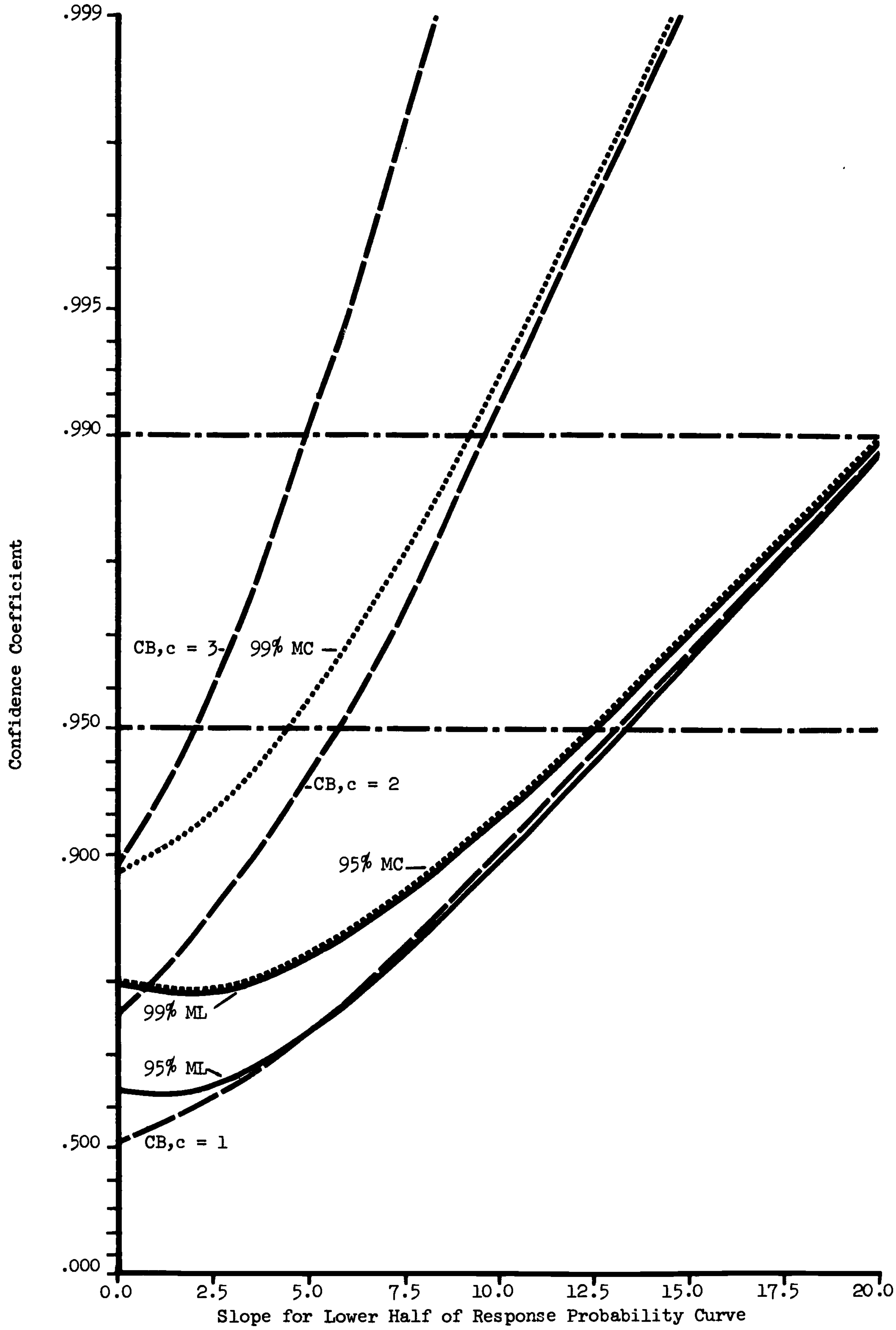

Fig. 5. Confidence coefficients for the countback method (CB), maximum likelihood (ML), and minimum χ^2 (MC) in the asymmetric situation, as a function of the lower slope.

and the resulting 1024 data sets analyzed, both by maximum likelihood and minimum χ^2 . Again, the total probability of those sets whose intervals did not include the 50% point was calculated for different values of a .

Not all possible data were considered in this evaluation. The two extremes for the remainder are that none of the associated intervals include the 50% point and that all of them do. Based on these extremes, absolute bounds for the confidence coefficients were found.[2] These bounds are quite good when a is large and almost useless when a is small (see Table 11).

We can combine the results of Tables 10 and 11 to provide a composite picture of the maximum likelihood confidence coefficients for the standard situation. To do this, we simply take the better bounds on the confidence coefficients at each value of a . The results are given a natural representation (namely bands, within which the coefficients almost certainly lie) in Figure 6, where they may be compared with the countback coefficients.

The third kind of simulation is best described as analytic, and was used exclusively with the countback method. As shown in the Statistical Appendix, exact formulas are available for the countback confidence coefficients in terms of P . These were used with several values of a , and for c (the number of levels counted back) equal to 1, 2,

[2] The derivation of the absolute bounds is quite simple. Suppose p_1 is the probability associated with the data examined whose intervals contain the 50% point. Also suppose p_2 is the probability associated with the data not examined. Then the bounds on the true confidence coefficient are p_1 and $p_1 + p_2$.

Table 11

Estimates and Absolute Bounds for Maximum Likelihood (ML) and Minimum X^2 (MC) Confidence Coefficients, Based on Partial Enumeration of the Possible Data Sets, in the Standard Situation (Logistic Model)

Slope <u>a</u>	Method	"95%" Lower Bound	"95%" Esti-mate	"95%" Upper Bound	"99%" Lower Bound	"99%" Esti-mate	"99%" Upper Bound
2.5	ML	.1336	.9052	.9860	.1423	.9638	.9947
	MC	.1391	.9421	.9914	.1445	.9785	.9968
5.0	ML	.5546	.8927	.9333	.6000	.9658	.9787
	MC	.5665	.9118	.9452	.6068	.9767	.9855
7.5	ML	.7918	.8935	.9056	.8656	.9768	.9794
	MC	.7975	.9000	.9114	.8708	.9827	.9846
10.0	ML	.8760	.9045	.9075	.9562	.9873	.9877
	MC	.8776	.9062	.9091	.9586	.9898	.9901
12.5	ML	.9132	.9213	.9220	.9852	.9939	.99396
	MC	.9136	.9217	.9224	.9861	.9948	.99485
15.0	ML	.9367	.9390	.9392	.9948	.99728	.99728
	MC	.9368	.9391	.9393	.9951	.99757	.99758
17.5	ML	.9540	.9547	.95474	.9981	.99884	.99884
	MC	.9540	.9547	.95476	.9982	.99892	.99892
20.0	ML	.9671	.9673	.96731	.9993	.99951	.99951
	MC	.9671	.9673	.96731	.9993	.99954	.99954

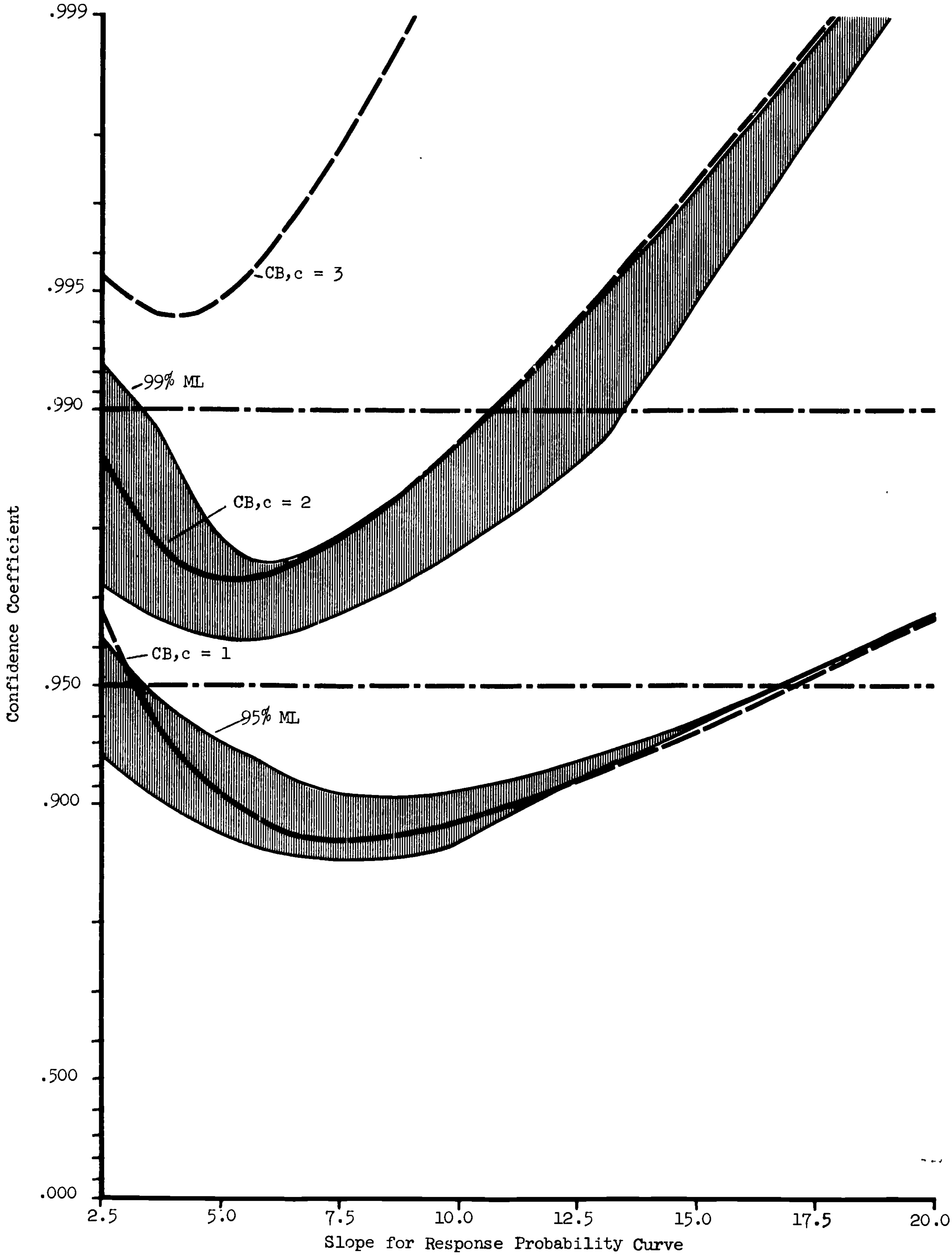

Fig. 6. Confidence coefficients for the countback method (CB) and bounds on the confidence coefficients for maximum likelihood (ML), based on Tables 10 and 11, for the logistic situation as a function of the slope.

and 3 in both the extreme and the standard situations. The results are summarized in Figures 5 and 6.

Finally, median lengths of countback intervals were found through analytic simulation. The possible data for the standard situation were divided into equivalence sets, according to the length of the interval they would lead to. Due to the simplicity of the countback method, the probability for each set is expressible in an exact formula.

To find the median interval length, one orders the sets according to length, begins with the smallest (for convenience) and accumulates the associated probabilities until the total reaches .5. The common length for the set of data last considered is the median of the lengths for all the data. The results of applying this procedure with $c = 3$ are shown in Figure 7. Here a was varied between 2 and 20 but this has been reinterpreted as varying the spacing between levels with a fixed. Both spacing and lengths are given in units of $1/a$.

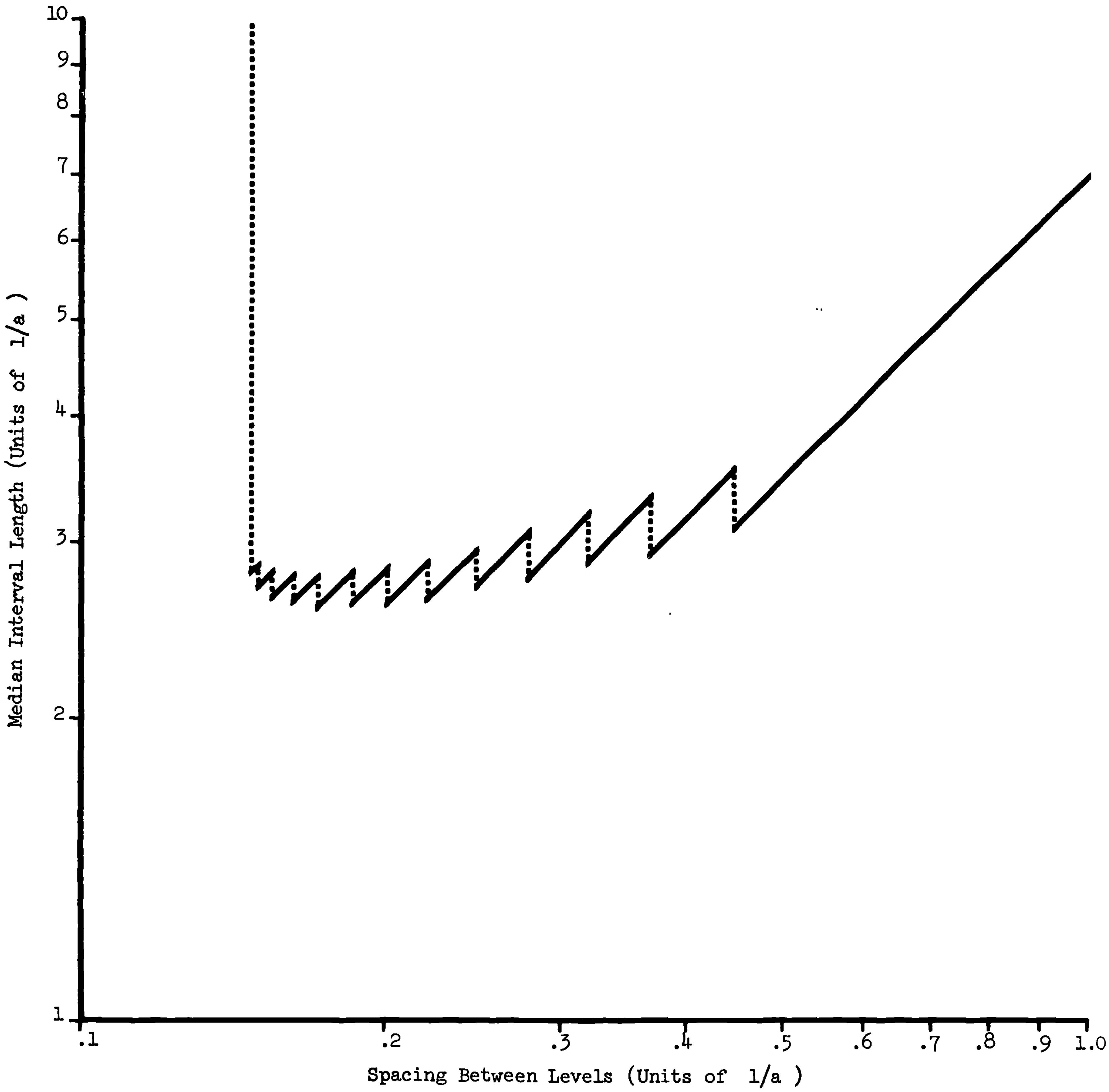

Fig. 7. Median length of countback confidence intervals with c = 3 , for the logistic situation.

Historical Appendix

Rather than provide a comprehensive history of approaches to the sensitivity problem, this section will attempt to give a general framework, so that the present development may be seen in proper perspective. In addition, more detailed discussions of various approaches will be referred to for the benefit of the interested reader.

By far the most widely adopted method of dealing with sensitivity data is maximum likelihood estimation. (A small sample of recent applications is provided in a list immediately following the Bibliography.) Most appropriate in situations where a specific probability model arises from theoretical considerations, it appeared in this context in a paper by Fisher (1922). In trying to estimate the density of organisms in a suspension, the dilution series experiment described at the beginning of the main section is usually employed. Fisher assumed that the number of organisms in any sample would have a Poisson distribution. (This would follow, as an adequately close approximation, from the assumption that the organisms are distributed randomly in the suspension.) On this basis, the probability that a given sample will have no organisms (resulting in a negative response) can be given explicitly in terms of the density of organisms in the suspension and the dilution of the sample.

As a consequence, the probability of a given set of data can be explicitly stated. The value of the density which maximizes this probability is the maximum likelihood estimate of the true density. Fisher worked out a way of obtaining this estimate and later provided tables (Fisher & Yates, 1963) to simplify the process. A good expository treatment appears in a paper by Cochran (1950), and applications as well as statistical modifications continue to appear in large numbers.

Unfortunately, most sensitivity problems do not come complete with a convenient and appropriate probability model. As a consequence, advocates of maximum likelihood estimation (e.g., Finney, 1952) have suggested the use of a standardized model, usually the cumulative normal distribution, with the provision that the variable under study be transformed, if necessary, to make the model more reasonable.

To actually find the maximum likelihood estimates, the standard suggestion has been to convert the problem into a linear one and use weighted least squares. The conversion is made by transforming the observed proportion of positive responses at each level into its equivalent normal deviate (using the inverse of the cumulative normal function). As a convenience, the normal function usually used has a mean of 5 and a standard deviation of 1. In this case, the transformed proportions have been called probits (Bliss, 1934), and the name probit analysis has been applied to the technique as a whole.

First correctly treated by Gaddum (1933) and Bliss (1934), probit analysis has been discussed in detail by Finney (1952, 1964), who includes a history of the method in his 1952 monograph. Another excellent review can be found in a collection of papers in Vol. 52 of the Annals of the New York Academy of Sciences. Of particular interest in this collection is a paper by Miller (1950) in which he discusses and compares varieties of probit analysis, related shortcut techniques, transformations other than the normal, and nonparametric methods. (These last will be mentioned later.)

The applications of probit analysis are extremely widespread, mostly within biology and pharmacology but, to an increasing extent, in psychology as well. Statistical treatment and modification also continue at a healthy pace. Much of this work can be found in Biometrics and some in Psychometrika.

It may have been observed that probit analysis (as I described it) has an important failing which is highly relevant to the main subject of this paper. This is that the probit corresponding to the proportion 0 is $-\infty$ and, corresponding to 1, $+\infty$. This is a problem, no matter how many responses are obtained at a level, but becomes acute when there is only one response per level, so that all proportions are either 0 or 1. Both Bliss (1938) and Finney (1947) have discussed this problem and offered reasonable solutions in terms of what Finney calls minimum and maximum working probits. Typically, this approach would involve several iterations of the weighted least squares solution. The alternative which I adopted is to maximize the likelihood function directly (by numerical techniques) thus working only with the proportions. Needless to say, this alternative is facilitated by the existence of high-speed computers, a resource not available at the time probit analysis was developed.

Another method using maximum likelihood and the normal model is the "up-and-down" approach of Dixon and Mood (1948). Here the main novelty is that the number of responses per level is determined as the experiment progresses. This has the important advantage of concentrating observations near the 50% point, thus improving the precision of the estimate. It has the disadvantage that responses must be obtained on an individual basis, with the next level only chosen after the previous response is known.

This has limited its applications, particularly in some of the biological work discussed earlier.

Hopefully, the foregoing remarks have provided some feeling for the extensive history maximum likelihood solutions to the sensitivity problem have enjoyed. To the nonstatistical reader, a word of explanation is in order. An extensive theory has been built up around maximum likelihood estimation. Among other things, this theory says that, asymptotically (i.e., as the sample of data becomes infinitely large), maximum likelihood estimates have minimum variance among all (asymptotically) unbiased estimates.

It is a fact, however, that maximum likelihood is not unique in this regard. For instance, estimates obtained with minimum X^2 techniques also have minimum variance as the sample becomes infinite. As the name implies, minimum X^2 involves finding estimates which minimize the usual goodness-of-fit measure for the data with respect to some model (such as the normal cumulative distribution).

Berkson is the main advocate of this approach for sensitivity data. He has discussed several variations of the method with both normal and logistic models, and compared it with maximum likelihood, in numerous papers (1946, 1949, 1953, 1955a, 1955b, 1956). Applications of minimum X^2 (as well as maximum likelihood) to psychological scaling problems are given a thorough treatment in a recent book by Bock and Jones (1968). Theoretical developments, as with maximum likelihood, are found mainly in Biometrics.

Finally, there are several methods which are related to the count-back, in that no strict assumptions are made concerning the response probabilities (although, unlike the countback, symmetry is always required). The best discussion of these approaches occurs in Finney (1964), where extensive comparisons are made among them. A common difficulty with these methods is that they emphasize estimation of the 50% point without providing much information about the quality of the results.

The sole exception to this generalization is the Spearman-Kärber method (discussed by Finney, 1964, and Brown, 1961). It estimates the 50% point as a linear function of the number of positive responses at each level. Since the number of positive responses at any level is a binomial random variable, the variance of the estimate is a linear function of binomial variances.

There has been no discussion on how to use this variance to construct confidence intervals for small samples, and the difficulty is acute for one observation per level, where one is not even sure how to estimate the variance itself. Nonetheless, Finney (1964) recommends the Spearman-Kärber method for some cases of the sensitivity problem.

The other method in this group which Finney recommends estimates the 50% point using a moving average (of the number of positive responses per level) and linear interpolation. Approximate expressions for the variance of this estimate have been developed (Harris, 1959), but none of these appears satisfactory for small samples. This moving average method is particularly unsuited for use with one observation per level. Not only is estimation of the variance problematical, there is also a real chance that its estimate of the 50% point will not be uniquely defined in this case.

Experimental Appendix

Introduction

The conflict between the two- and three-dimensional aspects of a picture has long been recognized, particularly by painters. The experiment discussed in the main section was designed to study this conflict systematically. A door or window, when photographed from an angle, may give rise to an image in the form of a trapezoid whose bases are vertical lines. To reconstruct the width of the door or window from its image, one must consider several things. The two which will be considered here are the distance between the bases (the width of the image) and the angles between the sides and the bases (the linear perspective). If we simplify the picture to include only this trapezoid, then its two- and three-dimensional aspects will be the two features just described. By obtaining judgments about the width of the original object as these features are varied, something may be learned about the nature of the conflict mentioned above.

General Description

Each subject was seated about two feet in front of a rear-projection screen. These were the instructions read to him:

> In this experiment, you will be shown a series of computer-generated pictures. Each of these shows two rectangular panels which meet at the center of the picture. You will be asked to indicate which of these panels appears wider to you by saying, "left" or "right." The panels are never of the same width, though they may be very similar. Also, no two pictures are identical, although some are very similar.

The first slide each subject saw is shown in Figure 2. The entire collection of 80 slides actually consisted of four subsets. These were constructed by choosing two standard panels, two comparison series, and combining each of the standards with each of the comparisons. The standards were both relatively flat in the picture plane and of moderate width, while the comparisons were more sharply angled toward the subject with widths varying systematically. (Thus, in Figure 2, the panel on the right is the standard, while that on the left is the comparison.)

Since more or less independent judgments were desired, members of the same series were not shown in immediate succession and the side on which the standard panels appeared was varied. Once an order of presentation had been chosen, it was held constant for all subjects. In a more rigorous experiment, the order would have been varied randomly for different subjects.

Equipment and Physical Specifications

A Kodak Carousel projector with zoom lens and remotely controlled advance was used to show the slides. The rear-projection unit was manufactured by the Radiant Corporation. It has a 14-inch square mylar screen, a sliding tray to hold the projector below and directly behind the screen, and two mirrors, one of them adjustable, to direct the picture from projector to screen.

During the experiment, the room was completely dark except for a six-inch square lighted area in the center of the screen. The brightness of this area was brought to a comfortable viewing level by operating the projector at "low" and by using two pieces of Polaroid glass placed in front of the projection lens as a variable filter. After this adjustment had been made, a Gossen footcandle meter was used to measure the illuminance of the screen at 25 inches, the approximate viewing distance of the subjects. The value arrived at in this way was 1.1 footcandles.

From slide to slide, the length of the central vertical line (see Fig. 2) is kept constant. When the stimuli were produced, the ratio of this length to the viewing distance was taken to be .1. With the length of this vertical line at 2.5 inches in the projected image, the appropriate viewing distance is 25 inches.

In this kind of experiment, the usual procedure would be to provide a small viewing aperture 25 inches away from the screen and directly in front of it and to otherwise block the subject's forward view. This arrangement would have the additional advantage of restricting the subject to monocular vision. In this way, the subject would have no binocular information about the flatness of the screen to contradict the impression of depth provided by the linear perspective.

Nonetheless, we allowed the subjects a free binocular view of the stimuli. They were seated in chairs which put them at approximately the correct distance and orientation, but no attempt was made to restrict their movements or compensate for their varying heights. In part, this was due to the nonrigorous nature of the experiment. It was also due to the fact that this is the way we normally look at pictures.

Production of the Stimuli

As indicated in the instructions to the subjects, the slides used for this experiment were generated with the help of a computer, the IBM 7094 at Princeton University. It operates a cathode ray tube and a

35 mm. camera together so that images appearing on the former are photographed on microfilm by the latter. The limits of image resolution in the finished product are fixed by the CRT rather than the film. All images must be built up on the face of the tube from a 1024 by 1024 square matrix of points. This is far short of ideal, but seemed satisfactory in the present case.

The Geometry of Perspective

To represent an object in perspective, the position of the object must be established with respect to a picture plane and a viewing point not in the plane. Imagine a line passing through the viewing point and any point in the object. The intersection of this line with the picture plane locates the representation of this part of the object.

In terms of a three dimensional coordinate system, it is convenient to use the locus of points for which $z = 0$ (in other words, the $x - y$ plane) as the picture plane, and to place the viewing point at $(0, 0, p)$, along the z axis. If the object point is (x, y, z), its representation is $(x', y', 0)$, where $x' = (p/(p - z))x$ and $y' = (p/(p - z))y$. This obviously excludes any part of the object for which $z = p$.

The situation is represented in Figure 8. The values for x' and y' are found with the help of similar triangles: y' and y are the lengths of one set of corresponding sides, while x' and x are the lengths of another set.

There are two opposing properties of perspective representation which we may note at this point. The first is that any change in the position of the object will, in general, result in a different representation. Moreover, this new representation will not, in general, even be

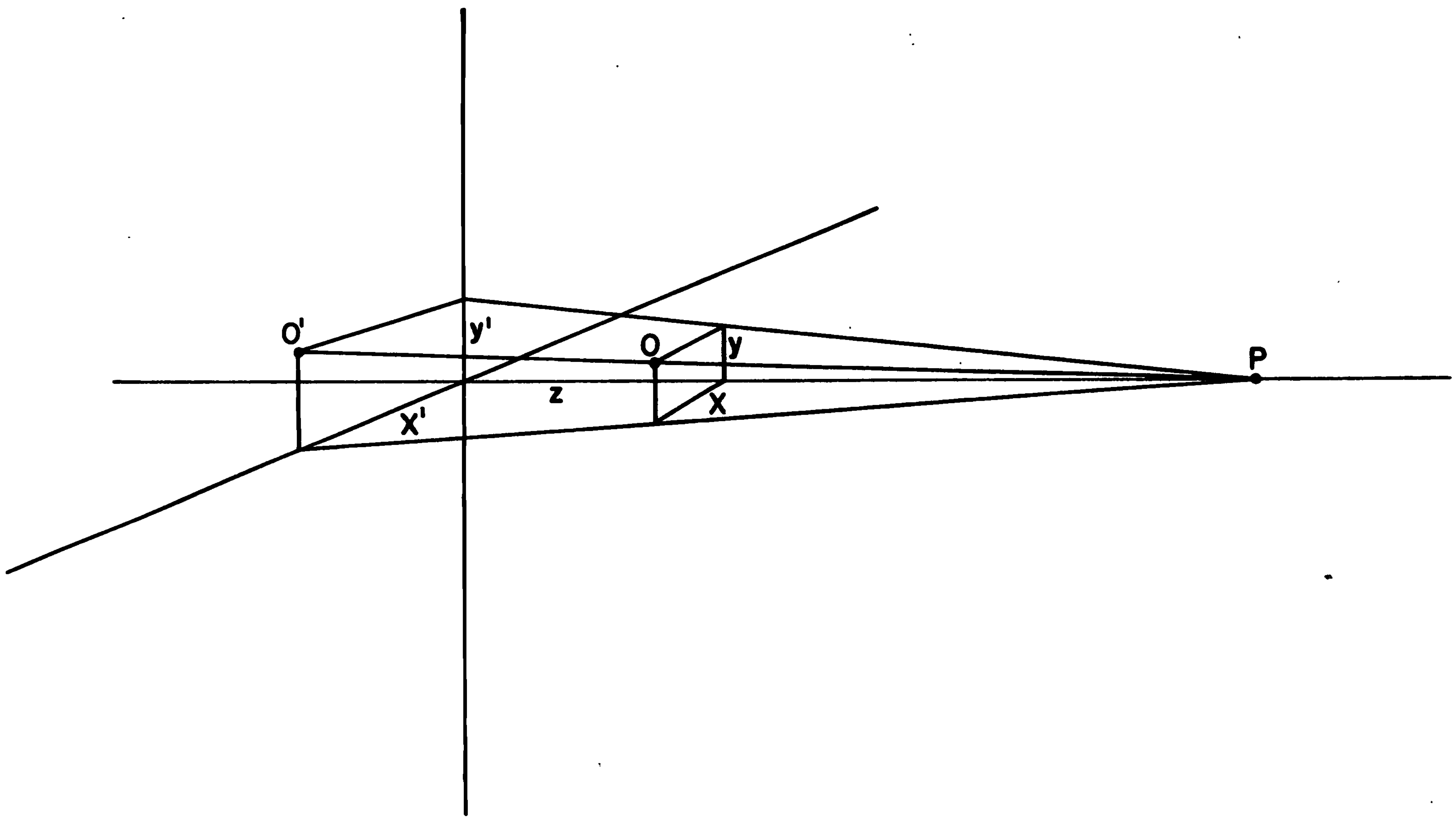

Fig. 8. The representation of point O as O' when viewed from P .

geometrically similar to the old one. That is, the relationships of the parts of the representation will change as the object is moved.

On the other hand, different objects may give rise to the same representation. In particular, a whole range of geometrically similar objects, increasing in overall size as they are placed further away from the viewing point, can be represented by a single figure on the picture plane.

Rectangles in Perspective

Having considered the basic projective equations, we should realize that they were not used directly to determine the stimuli for this experiment. Since the figures to be represented were rectangles, conveniently aligned with the vertical axis of the picture plane, it was possible to derive special projective equations applying to them.

Figure 9 provides a view of the situation from above (along the y -axis of Figure 8) as well as one from in front (along the z -axis). In the first of these views, BD is what can be seen of the rectangle and BC is what can be seen of its projection. The rectangle is inclined from the picture plane by the angle θ . Finally, A is the "vanishing point" of the rectangle, that is, the point toward which the top and bottom edges of its projection converge.

In the experiment, the projections were varied systematically. As a result, the primary interest was in specifying the rectangles in terms of their projections (and the viewing distance). Referring to Figure 9, we want to find expressions for $\overline{BD}$ and θ in terms of $\overline{AB}$, $\overline{BC}$, and $\overline{BP}$.

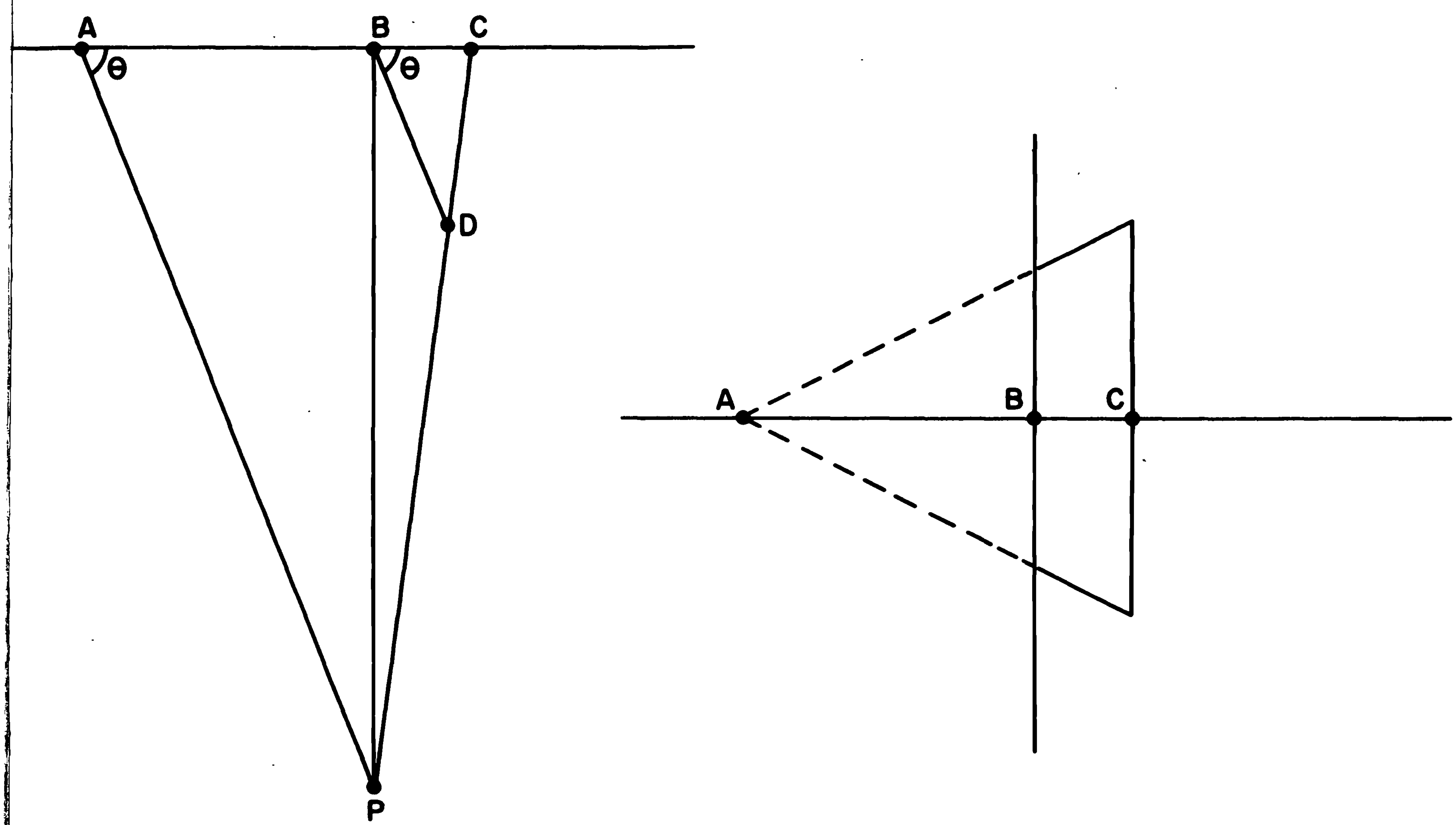

Fig. 9. The representation of a slanted rectangle when viewed from P .

Since triangles ACP and BCD are similar, $\overline{BD}/\overline{BC} = \overline{AP}/\overline{AC}$ or $\overline{BD} = \overline{BC}(\overline{AP}/\overline{AC})$. If we label the variables of interest as $x = \overline{BD}$, $a = \overline{AB}$, $b = \overline{BC}$, and $p = \overline{BP}$, the equation becomes $x = (b/(a + b))(a^2 + p^2)^{1/2}$. In addition, $\theta = \arctan (p/a)$, so the problem is solved.

Parameters of the Stimuli

Now it is possible to describe the stimuli more completely. As we remarked earlier, the central vertical line (the common inside edge of the rectangles) was fixed, and its ratio to p , the projecting distance, was set at .1. Taking the length of the edge as 2.5 makes $p = 25$.

We may remember that in the experiment each of two standard stimuli is paired with each member of two comparison series. In the units set above (which happen to be inches), standard one (S_1) has $\underline{a} = 50$, $\underline{b} = 1.5$, while for S_2 , $\underline{a} = 25$, $\underline{b} = 1.5$. In other words, the standards have the same projected width.

Let us use ϕ to denote the angle by which the top (or bottom) projected edge deviates from the horizontal. Then ϕ for S_1 is $\arctan (1.25/50) = 1.43210^\circ$ and, for S_2 , $\phi = 2.86241^\circ$.

Using the equations derived above, the rectangle represented by S_1 has a width of 1.62820 and comes out of the picture plane at an angle of 26.5651° . That represented by S_2 has $x = 2.00124$ and $\theta = 45^\circ$.

We can summarize the above results in terms of an inverse projective function $F : (b,\phi) \to (x,\theta)$.

For S_1 , $F(1.5, 1.43210^\circ) = (1.62820, 26.5651^\circ)$.

For S_2 , $F(1.5, 2.86421^\circ) = (2.00124, 45.0000^\circ)$.

The two comparison series (C_1 and C_2) may be similarly described. For each of them, $\underline{a}$ (hence ϕ) is fixed while $\underline{b}$ is varied systematically. In the case of C_1 , $\underline{a} = 2.5$ ($\phi = 26.5651^{\circ}$) and $\underline{b} = .50,.55,\ldots,1.45$. With C_2 , $\underline{a} = 5$ ($\phi = 14.0362^{\circ}$) and $\underline{b} = .75,.80,\ldots,1.70$. Thus each series has 20 components and the components in each case are generated by changing $\underline{b}$ in steps of .05.

We can easily describe the corresponding rectangles with the help of a modified version of F .

For C_1 , $F(.50,\ldots,1.45;\ 26.5651^{\circ}) = (4.18745,\ldots,9.22299;\ 84.2894^{\circ})$.

For C_2 , $F(.75,\ldots,1.70;\ 14.0362^{\circ}) = (3.32544,\ldots,6.46890;\ 78.6901^{\circ})$.

Finally, the stimulus in Figure 2 is S_1 paired with an element of C_1 which has $\underline{b} = 1.15$.

Data

The data for all 19 subjects[3] are found in Tables 12 and 14, immediately following this section. Table 12 gives the judgments in the order they were made, with an L if it was the left panel that appeared wider and an R if it was the right panel. Table 13 provides the code for interpreting these judgments. It may be read as follows: 1 S1;C1,14 R means that the first judgment of each subject was made between S_1 and the 14th element of C_1 (ordered so that the first element is the smallest) with the standard on the right.

Table 14 is the result of applying this code to Table 12. For ease (hopefully) of interpretation, the decoded data for each subject are

[3]Actually, there were only 17 subjects. Subject 7 wanted to try again, and, thus, become subject 8 as well. Subject 18 became subject 19 using a different interpretation of the instructions.

Table 12

Raw Data from Experiment with Each Subject's Judgment for Each of the 80 Slides, L and R Indicating Which Rectangle (Left or Right) Appeared Wider

SUBJECT NUMBER 1

1	L	21	R	41	R	61	L
2	R	22	L	42	L	62	R
3	R	23	R	43	L	63	R
4	L	24	R	44	R	64	L
5	L	25	L	45	L	65	R
6	R	26	L	46	L	66	R
7	L	27	L	47	L	67	R
8	R	28	R	48	L	68	L
9	R	29	L	49	R	69	R
10	L	30	L	50	L	70	R
11	R	31	L	51	L	71	R
12	L	32	R	52	L	72	L
13	L	33	R	53	R	73	L
14	R	34	L	54	R	74	R
15	R	35	L	55	L	75	R
16	L	36	L	56	L	76	R
17	L	37	R	57	L	77	R
18	L	38	L	58	L	78	L
19	R	39	L	59	R	79	R
20	L	40	R	60	R	80	R

SUBJECT NUMBER 2

1	L	21	L	41	R	61	L
2	R	22	R	42	L	62	R
3	R	23	R	43	R	63	R
4	L	24	R	44	R	64	R
5	L	25	L	45	R	65	R
6	R	26	L	46	L	66	R
7	L	27	L	47	L	67	R
8	R	28	R	48	L	68	L
9	R	29	L	49	R	69	R
10	L	30	L	50	L	70	R
11	L	31	R	51	L	71	R
12	L	32	L	52	L	72	R
13	L	33	L	53	L	73	L
14	R	34	R	54	R	74	R
15	R	35	L	55	R	75	R
16	L	36	L	56	L	76	R
17	R	37	R	57	L	77	R
18	L	38	L	58	R	78	L
19	R	39	L	59	R	79	R
20	L	40	R	60	R	80	R

SUBJECT NUMBER 3

1	R	21	L	41	L	61	L
2	R	22	R	42	L	62	R
3	L	23	L	43	R	63	L
4	L	24	R	44	L	64	R
5	L	25	R	45	R	65	L
6	R	26	L	46	L	66	R
7	L	27	R	47	R	67	L
8	R	28	R	48	L	68	L
9	R	29	L	49	L	69	L
10	L	30	L	50	R	70	R
11	L	31	R	51	L	71	L
12	L	32	L	52	L	72	R
13	R	33	L	53	L	73	R
14	R	34	R	54	L	74	R
15	L	35	L	55	R	75	R
16	L	36	L	56	L	76	R
17	R	37	R	57	R	77	R
18	L	38	L	58	R	78	L
19	L	39	L	59	R	79	L
20	L	40	R	60	L	80	R

SUBJECT NUMBER 4

1	L	21	L	41	R	61	L
2	R	22	L	42	L	62	R
3	R	23	L	43	R	63	R
4	L	24	R	44	L	64	R
5	L	25	L	45	L	65	R
6	R	26	L	46	L	66	R
7	R	27	L	47	L	67	R
8	R	28	R	48	L	68	L
9	R	29	L	49	R	69	R
10	L	30	L	50	L	70	R
11	L	31	R	51	L	71	R
12	L	32	L	52	L	72	R
13	L	33	L	53	L	73	L
14	R	34	R	54	R	74	R
15	R	35	L	55	R	75	R
16	L	36	L	56	L	76	R
17	R	37	R	57	L	77	R
18	L	38	L	58	R	78	L
19	R	39	L	59	R	79	R
20	L	40	R	60	R	80	R

Table 12 (Contd)

SUBJECT NUMBER 5

1 R	21 L	41 R	61 L
2 R	22 R	42 L	62 R
3 L	23 L	43 R	63 L
4 L	24 L	44 L	64 R
5 L	25 R	45 R	65 L
6 R	26 L	46 L	66 R
7 L	27 R	47 R	67 R
8 R	28 R	48 L	68 L
9 R	29 L	49 R	69 R
10 L	30 L	50 L	70 R
11 L	31 R	51 L	71 R
12 L	32 L	52 L	72 R
13 R	33 L	53 L	73 L
14 R	34 L	54 R	74 R
15 L	35 L	55 R	75 R
16 L	36 L	56 L	76 R
17 R	37 R	57 L	77 R
18 L	38 L	58 R	78 L
19 L	39 L	59 R	79 L
20 L	40 R	60 R	80 R

SUBJECT NUMBER 6

1 L	21 L	41 R	61 L
2 R	22 L	42 L	62 R
3 R	23 R	43 L	63 R
4 L	24 R	44 R	64 L
5 L	25 L	45 L	65 R
6 R	26 L	46 L	66 R
7 L	27 L	47 L	67 R
8 R	28 R	48 L	68 L
9 R	29 L	49 R	69 R
10 L	30 L	50 L	70 R
11 R	31 L	51 R	71 R
12 L	32 L	52 L	72 R
13 L	33 R	53 L	73 L
14 R	34 L	54 R	74 R
15 R	35 L	55 L	75 L
16 L	36 L	56 L	76 R
17 R	37 R	57 L	77 R
18 L	38 L	58 L	78 L
19 R	39 R	59 L	79 R
20 L	40 R	60 R	80 R

SUBJECT NUMBER 7

1 L	21 R	41 R	61 L
2 R	22 L	42 L	62 R
3 R	23 R	43 L	63 R
4 L	24 R	44 R	64 L
5 R	25 L	45 L	65 R
6 R	26 R	46 L	66 R
7 R	27 L	47 L	67 R
8 R	28 R	48 L	68 R
9 R	29 L	49 R	69 R
10 L	30 L	50 L	70 R
11 R	31 L	51 L	71 R
12 L	32 R	52 L	72 R
13 L	33 R	53 R	73 L
14 R	34 L	54 R	74 R
15 R	35 R	55 L	75 L
16 L	36 L	56 R	76 L
17 R	37 R	57 L	77 L
18 R	38 L	58 L	78 L
19 R	39 R	59 R	79 R
20 L	40 R	60 R	80 R

SUBJECT NUMBER 8

1 L	21 R	41 R	61 L
2 R	22 L	42 L	62 R
3 R	23 R	43 R	63 R
4 L	24 R	44 R	64 R
5 R	25 L	45 R	65 R
6 R	26 R	46 L	66 R
7 L	27 L	47 L	67 R
8 R	28 R	48 L	68 R
9 R	29 L	49 R	69 R
10 L	30 R	50 L	70 R
11 R	31 R	51 L	71 R
12 L	32 R	52 L	72 R
13 L	33 R	53 R	73 L
14 R	34 L	54 R	74 R
15 R	35 R	55 L	75 L
16 L	36 L	56 L	76 R
17 R	37 R	57 L	77 R
18 R	38 R	58 L	78 L
19 R	39 R	59 L	79 R
20 L	40 R	60 R	80 R

Table 12 (Contd)

SUBJECT NUMBER 9

1 R	21 L	41 R	61 L
2 R	22 R	42 L	62 R
3 R	23 L	43 R	63 L
4 L	24 R	44 L	64 R
5 L	25 R	45 R	65 L
6 R	26 L	46 L	66 R
7 L	27 R	47 R	67 R
8 R	28 R	48 L	68 L
9 R	29 L	49 R	69 R
10 L	30 L	50 L	70 R
11 L	31 R	51 L	71 R
12 L	32 L	52 L	72 R
13 R	33 L	53 L	73 L
14 R	34 R	54 R	74 R
15 R	35 L	55 R	75 R
16 L	36 L	56 L	76 R
17 R	37 R	57 L	77 R
18 L	38 L	58 R	78 L
19 L	39 L	59 R	79 L
20 L	40 R	60 L	80 R

SUBJECT NUMBER 10

1 L	21 R	41 R	61 L
2 R	22 L	42 L	62 R
3 R	23 R	43 L	63 R
4 L	24 R	44 R	64 L
5 R	25 L	45 L	65 R
6 R	26 R	46 L	66 R
7 R	27 L	47 L	67 R
8 R	28 R	48 L	68 L
9 R	29 L	49 R	69 R
10 L	30 R	50 L	70 R
11 R	31 L	51 R	71 R
12 L	32 R	52 L	72 L
13 L	33 R	53 R	73 L
14 R	34 L	54 R	74 R
15 R	35 R	55 L	75 L
16 L	36 L	56 R	76 L
17 R	37 R	57 L	77 R
18 R	38 R	58 L	78 R
19 R	39 R	59 L	79 R
20 L	40 L	60 R	80 L

SUBJECT NUMBER 11

1 L	21 L	41 R	61 L
2 R	22 R	42 L	62 R
3 R	23 R	43 R	63 R
4 L	24 R	44 L	64 R
5 L	25 L	45 R	65 R
6 R	26 L	46 L	66 R
7 L	27 L	47 R	67 R
8 R	28 R	48 L	68 L
9 R	29 L	49 R	69 R
10 L	30 L	50 L	70 R
11 L	31 R	51 L	71 R
12 L	32 L	52 L	72 R
13 L	33 L	53 L	73 L
14 R	34 L	54 R	74 R
15 R	35 L	55 R	75 R
16 L	36 L	56 L	76 R
17 R	37 R	57 L	77 R
18 L	38 L	58 L	78 L
19 R	39 L	59 R	79 R
20 L	40 R	60 R	80 R

SUBJECT NUMBER 12

1 R	21 L	41 R	61 L
2 R	22 R	42 L	62 R
3 R	23 L	43 R	63 R
4 L	24 R	44 R	64 R
5 L	25 R	45 R	65 L
6 R	26 L	46 L	66 R
7 L	27 R	47 R	67 R
8 R	28 R	48 L	68 L
9 R	29 L	49 R	69 R
10 L	30 L	50 L	70 R
11 L	31 R	51 L	71 R
12 L	32 L	52 L	72 R
13 R	33 L	53 L	73 L
14 R	34 R	54 R	74 R
15 R	35 L	55 R	75 R
16 L	36 L	56 L	76 R
17 R	37 R	57 L	77 R
18 L	38 L	58 R	78 L
19 L	39 L	59 R	79 R
20 L	40 R	60 L	80 R

Table 12 (Contd)

SUBJECT NUMBER 13

1 L	21 R	41 R	61 L
2 R	22 L	42 L	62 R
3 R	23 R	43 L	63 R
4 L	24 R	44 R	64 L
5 L	25 L	45 L	65 R
6 R	26 L	46 L	66 R
7 L	27 L	47 L	67 R
8 R	28 R	48 L	68 R
9 R	29 L	49 R	69 R
10 L	30 L	50 L	70 R
11 R	31 L	51 R	71 R
12 L	32 L	52 L	72 L
13 L	33 L	53 R	73 L
14 R	34 L	54 R	74 R
15 R	35 L	55 L	75 L
16 L	36 L	56 R	76 L
17 L	37 R	57 L	77 R
18 R	38 L	58 L	78 R
19 R	39 R	59 L	79 R
20 L	40 L	60 R	80 L

SUBJECT NUMBER 14

1 L	21 L	41 R	61 L
2 R	22 R	42 L	62 R
3 R	23 L	43 R	63 L
4 L	24 R	44 L	64 L
5 L	25 L	45 L	65 R
6 R	26 L	46 L	66 R
7 L	27 R	47 L	67 R
8 R	28 R	48 L	68 L
9 R	29 L	49 R	69 R
10 L	30 L	50 L	70 R
11 L	31 R	51 L	71 R
12 L	32 L	52 L	72 R
13 L	33 L	53 L	73 L
14 R	34 L	54 R	74 R
15 R	35 R	55 L	75 L
16 L	36 L	56 L	76 R
17 R	37 R	57 L	77 R
18 L	38 L	58 L	78 L
19 R	39 L	59 R	79 R
20 L	40 R	60 R	80 L

SUBJECT NUMBER 15

1 L	21 L	41 R	61 L
2 R	22 L	42 L	62 R
3 R	23 R	43 R	63 R
4 L	24 R	44 R	64 L
5 L	25 L	45 R	65 R
6 R	26 L	46 L	66 R
7 L	27 L	47 L	67 R
8 R	28 R	48 L	68 L
9 R	29 L	49 R	69 R
10 L	30 L	50 L	70 R
11 R	31 R	51 L	71 R
12 L	32 R	52 L	72 R
13 L	33 L	53 L	73 L
14 R	34 L	54 R	74 R
15 R	35 L	55 L	75 L
16 L	36 L	56 L	76 L
17 R	37 R	57 L	77 R
18 L	38 L	58 L	78 L
19 R	39 L	59 R	79 R
20 L	40 R	60 R	80 L

SUBJECT NUMBER 16

1 L	21 L	41 R	61 L
2 R	22 L	42 L	62 R
3 R	23 R	43 L	63 R
4 L	24 R	44 R	64 L
5 L	25 L	45 L	65 R
6 R	26 L	46 L	66 R
7 L	27 L	47 L	67 R
8 R	28 R	48 L	68 L
9 R	29 L	49 R	69 R
10 L	30 L	50 L	70 R
11 L	31 L	51 L	71 R
12 L	32 L	52 L	72 L
13 L	33 L	53 L	73 L
14 R	34 L	54 R	74 R
15 R	35 L	55 L	75 L
16 L	36 L	56 L	76 L
17 R	37 R	57 L	77 R
18 L	38 L	58 L	78 L
19 R	39 R	59 R	79 R
20 L	40 R	60 R	80 L

Table 12 (Contd)

SUBJECT NUMBER 17

1 R	21 L	41 L	61 R
2 L	22 R	42 L	62 R
3 L	23 L	43 R	63 L
4 L	24 R	44 L	64 R
5 L	25 R	45 R	65 L
6 R	26 L	46 L	66 R
7 L	27 R	47 R	67 R
8 R	28 R	48 L	68 L
9 R	29 L	49 R	69 R
10 L	30 L	50 R	70 R
11 L	31 R	51 L	71 L
12 L	32 L	52 L	72 R
13 R	33 L	53 L	73 R
14 R	34 R	54 L	74 R
15 L	35 L	55 R	75 R
16 L	36 L	56 L	76 R
17 R	37 R	57 R	77 R
18 L	38 L	58 R	78 L
19 L	39 L	59 R	79 L
20 L	40 R	60 L	80 R

SUBJECT NUMBER 18

1 R	21 L	41 L	61 L
2 L	22 R	42 L	62 R
3 L	23 L	43 R	63 L
4 L	24 L	44 L	64 R
5 L	25 R	45 R	65 L
6 R	26 L	46 L	66 R
7 L	27 R	47 R	67 L
8 R	28 R	48 L	68 L
9 L	29 R	49 L	69 L
10 R	30 L	50 R	70 R
11 L	31 R	51 L	71 L
12 L	32 L	52 L	72 R
13 R	33 L	53 L	73 R
14 R	34 R	54 L	74 L
15 L	35 L	55 R	75 R
16 L	36 L	56 L	76 R
17 R	37 R	57 R	77 R
18 L	38 L	58 R	78 L
19 L	39 L	.59 R	79 L
20 L	40 R	60 L	80 R

SUBJECT NUMBER 19

1 L	21 L	41 R	61 L
2 R	22 R	42 L	62 R
3 R	23 L	43 R	63 R
4 L	24 R	44 L	64 R
5 L	25 L	45 R	65 R
6 R	26 L	46 L	66 R
7 L	27 L	47 L	67 R
8 R	28 R	48 L	68 L
9 R	29 L	49 R	69 R
10 L	30 L	50 L	70 R
11 L	31 R	51 L	71 R
12 L	32 L	52 L	72 R
13 L	33 L	53 L	73 L
14 R	34 R	54 R	74 R
15 R	35 L	55 L	75 R
16 L	36 R	56 L	76 R
17 R	37 R	57 L	77 R
18 L	38 L	58 L	78 L
19 R	39 L	59 R	79 R
20 L	40 R	60 R	80 R

Table 13

Arrangement of Slides for Experiment, Showing Which Standard Rectangle Was Paired with Which Comparison, and on Which Side the Standard Appeared

1	S1;C1,14	R	21	S1;C1,5	L	41	S2;C1,14	L	61	S2;C1,19	R
2	S2;C2,13	L	22	S2;C2,11	R	42	S2;C2,19	R	62	S1;C2,19	L
3	S2;C1,16	L	23	S1;C1,9	L	43	S1;C1,4	R	63	S2;C1,11	L
4	S1;C2,17	R	24	S1;C2,12	L	44	S2;C2,10	L	64	S1;C2,8	R
5	S2;C1,1	L	25	S1;C1,13	R	45	S1;C1,6	R	65	S1;C1,10	L
6	S1;C2,3	R	26	S1;C2,2	L	46	S2;C2,18	R	66	S2;C2,3	R
7	S1;C1,3	L	27	S2;C1,13	R	47	S2;C1,10	R	67	S2;C1,20	L
8	S1;C2,18	L	28	S2;C2,1	R	48	S2;C2,2	L	68	S2;C2,8	L
9	S1;C1,18	L	29	S2;C1,18	R	49	S1;C1,17	L	69	S1;C1,19	L
10	S1;C2,14	R	30	S2;C2,6	L	50	S1;C2,13	R	70	S2;C2,5	R
11	S2;C1,6	L	31	S2;C1,9	R	51	S1;C1,1	L	71	S2;C1,12	L
12	S2;C2,16	R	32	S1;C2,7	L	52	S2;C2,20	R	72	S1;C2,6	R
13	S2;C1,15	R	33	S2;C1,5	L	53	S2;C1,4	L	73	S1;C1,15	R
14	S1;C2,15	L	34	S1;C2,11	R	54	S2;C2,14	L	74	S1;C2,16	L
15	S1;C1,16	L	35	S1;C1,2	L	55	S1;C1,7	R	75	S1;C1,8	R
16	S1;C2,20	R	36	S2;C2,17	R	56	S1;C2,5	L	76	S2;C2,9	R
17	S2;C1,3	R	37	S1;C1,20	L	57	S2;C1,17	R	77	S2;C1,2	R
18	S2;C2,7	L	38	S1;C2,1	L	58	S1;C2,10	R	78	S2;C2,4	L
19	S1;C1,12	L	39	S2;C1,7	L	59	S2;C1,8	R	79	S1;C1,11	L
20	S2;C2,15	R	40	S1;C2,4	R	60	S2;C2,12	L	80	S1;C2,9	R

Table 14

The Data of Table 12, Rearranged Using Table 13 So That for Each Comparison Width, It Is Noted Whether the Standard (-1) or the Comparison (1) Was Judged Wider

SUBJECT NUMBER 1

	C1		C2	
B	S1	S2	S1	S2
0.50	-1	-1		
0.55	-1	-1		
0.60	-1	1		
0.65	1	1		
0.70	1	1		
0.75	1	1	-1	-1
0.80	1	-1	-1	-1
0.85	-1	-1	-1	-1
0.90	1	1	-1	-1
0.95	1	1	-1	-1
1.00	1	1	1	-1
1.05	1	1	1	-1
1.10	1	1	1	-1
1.15	1	1	-1	-1
1.20	1	1	1	1
1.25	1	1	1	1
1.30	1	1	1	1
1.35	1	1	1	1
1.40	1	1	1	1
1.45	1	1	1	1
1.50			1	1
1.55			1	1
1.60			1	1
1.65			1	1
1.70			1	1

SUBJECT NUMBER 2

	C1		C2	
B	S1	S2	S1	S2
0.50	-1	-1		
0.55	-1	-1		
0.60	-1	-1		
0.65	-1	-1		
0.70	-1	-1		
0.75	-1	-1	-1	-1
0.80	-1	-1	-1	-1
0.85	-1	-1	-1	-1
0.90	1	-1	-1	-1
0.95	1	1	-1	-1
1.00	1	1	-1	-1
1.05	1	1	-1	-1
1.10	1	1	-1	-1
1.15	1	1	-1	-1
1.20	1	1	-1	1
1.25	1	1	-1	-1
1.30	1	1	1	1
1.35	1	1	1	1
1.40	1	1	1	1
1.45	1	1	1	1
1.50			1	1
1.55			1	1
1.60			1	1
1.65			1	1
1.70			1	1

SUBJECT NUMBER 3

	C1		C2	
B	S1	S2	S1	S2
0.50	-1	-1		
0.55	-1	-1		
0.60	-1	-1		
0.65	-1	-1		
0.70	-1	-1		
0.75	-1	-1	-1	-1
0.80	-1	-1	-1	-1
0.85	-1	-1	-1	-1
0.90	-1	-1	-1	-1
0.95	-1	-1	-1	-1
1.00	-1	-1	-1	-1
1.05	-1	-1	-1	-1
1.10	-1	-1	-1	-1
1.15	-1	-1	-1	-1
1.20	-1	-1	-1	-1
1.25	-1	-1	-1	-1
1.30	-1	-1	1	-1
1.35	1	1	-1	1
1.40	-1	1	1	-1
1.45	1	-1	1	1
1.50			1	1
1.55			1	1
1.60			1	1
1.65			1	1
1.70			1	1

SUBJECT NUMBER 4

	C1		C2	
B	S1	S2	S1	S2
0.50	-1	-1		
0.55	-1	-1		
0.60	1	-1		
0.65	-1	-1		
0.70	-1	-1		
0.75	1	-1	-1	-1
0.80	-1	-1	-1	-1
0.85	-1	-1	-1	-1
0.90	-1	-1	-1	-1
0.95	1	1	-1	-1
1.00	1	1	-1	-1
1.05	1	1	-1	-1
1.10	1	1	-1	-1
1.15	1	1	-1	-1
1.20	1	1	-1	-1
1.25	1	1	-1	1
1.30	1	1	1	1
1.35	1	1	1	1
1.40	1	1	1	1
1.45	1	1	1	1
1.50			1	1
1.55			1	1
1.60			1	1
1.65			1	1
1.70			1	1

Table 14 (Contd)

SUBJECT NUMBER 5

	C1		C2	
B	S1	S2	S1	S2
0.50	-1	-1		
0.55	-1	-1		
0.60	-1	-1		
0.65	-1	-1		
0.70	-1	-1		
0.75	-1	-1	-1	-1
0.80	-1	-1	-1	-1
0.85	-1	-1	-1	-1
0.90	-1	-1	-1	-1
0.95	-1	-1	-1	-1
1.00	-1	-1	-1	-1
1.05	-1	1	-1	-1
1.10	-1	-1	-1	-1
1.15	-1	1	-1	-1
1.20	1	-1	-1	-1
1.25	-1	-1	1	-1
1.30	1	1	-1	1
1.35	1	1	1	1
1.40	1	1	1	1
1.45	1	1	1	1
1.50			1	1
1.55			1	1
1.60			1	1
1.65			1	1
1.70			1	1

SUBJECT NUMBER 6

	C1		C2	
B	S1	S2	S1	S2
0.50	1	-1		
0.55	-1	-1		
0.60	-1	-1		
0.65	1	-1		
0.70	-1	1		
0.75	1	1	-1	-1
0.80	1	1	-1	-1
0.85	1	1	-1	-1
0.90	1	1	-1	-1
0.95	1	1	-1	-1
1.00	1	1	-1	-1
1.05	1	1	-1	-1
1.10	1	1	1	-1
1.15	1	1	-1	-1
1.20	1	1	1	1
1.25	1	1	1	1
1.30	1	1	1	1
1.35	1	1	1	1
1.40	1	1	1	1
1.45	1	1	1	1
1.50			1	1
1.55			1	1
1.60			1	1
1.65			1	1
1.70			1	1

SUBJECT NUMBER 7

	C1		C2	
B	S1	S2	S1	S2
0.50	-1	1		
0.55	1	1		
0.60	1	-1		
0.65	1	1		
0.70	1	1		
0.75	1	1	-1	-1
0.80	1	1	1	-1
0.85	1	-1	-1	-1
0.90	1	1	-1	-1
0.95	1	1	1	-1
1.00	1	1	-1	-1
1.05	1	1	1	1
1.10	1	1	1	1
1.15	1	1	-1	1
1.20	1	1	1	1
1.25	1	1	1	1
1.30	1	1	1	1
1.35	1	1	1	1
1.40	1	1	1	1
1.45	1	1	1	1
1.50			1	1
1.55			1	1
1.60			1	1
1.65			1	1
1.70			1	1

SUBJECT NUMBER 8

	C1		C2	
B	S1	S2	S1	S2
0.50	-1	1		
0.55	1	-1		
0.60	-1	-1		
0.65	-1	1		
0.70	1	1		
0.75	-1	1	1	-1
0.80	1	1	1	-1
0.85	1	1	-1	-1
0.90	1	-1	-1	-1
0.95	1	1	-1	-1
1.00	1	1	-1	1
1.05	1	1	1	1
1.10	1	1	-1	1
1.15	1	1	-1	-1
1.20	1	1	1	1
1.25	1	1	1	1
1.30	1	1	1	1
1.35	1	1	1	1
1.40	1	1	1	1
1.45	1	1	1	1
1.50			1	1
1.55			1	1
1.60			1	1
1.65			1	1
1.70			1	1

Table 14 (Contd)

SUBJECT NUMBER 9

B	C1 S1	C1 S2	C2 S1	C2 S2
0.50	-1	-1		
0.55	-1	-1		
0.60	-1	-1		
0.65	-1	-1		
0.70	-1	-1		
0.75	-1	-1	-1	-1
0.80	-1	-1	-1	-1
0.85	-1	-1	-1	-1
0.90	-1	-1	-1	-1
0.95	-1	-1	-1	-1
1.00	-1	-1	-1	-1
1.05	-1	1	-1	-1
1.10	-1	-1	-1	-1
1.15	-1	1	-1	-1
1.20	1	-1	-1	-1
1.25	1	1	-1	-1
1.30	1	1	1	-1
1.35	1	1	1	1
1.40	1	1	1	1
1.45	1	1	1	1
1.50			1	1
1.55			1	1
1.60			1	1
1.65			1	1
1.70			1	1

SUBJECT NUMBER 10

B	C1 S1	C1 S2	C2 S1	C2 S2
0.50	1	1		
0.55	1	-1		
0.60	1	-1		
0.65	1	1		
0.70	1	1		
0.75	1	1	1	-1
0.80	1	1	1	-1
0.85	1	1	-1	-1
0.90	1	1	1	1
0.95	1	1	1	-1
1.00	1	1	1	1
1.05	1	1	1	1
1.10	1	1	1	-1
1.15	1	1	1	1
1.20	1	1	1	1
1.25	1	1	1	1
1.30	1	1	1	1
1.35	1	1	1	1
1.40	1	1	1	1
1.45	1	1	1	1
1.50			1	1
1.55			1	1
1.60			1	1
1.65			1	1
1.70			1	1

SUBJECT NUMBER 11

B	C1 S1	C1 S2	C2 S1	C2 S2
0.50	-1	-1		
0.55	-1	-1		
0.60	-1	-1		
0.65	-1	-1		
0.70	-1	-1		
0.75	-1	-1	-1	-1
0.80	-1	-1	-1	-1
0.85	-1	-1	-1	-1
0.90	1	-1	-1	-1
0.95	1	-1	-1	-1
1.00	1	1	-1	-1
1.05	1	1	-1	-1
1.10	1	1	-1	-1
1.15	1	1	-1	-1
1.20	1	1	1	-1
1.25	1	1	1	-1
1.30	1	1	1	1
1.35	1	1	1	1
1.40	1	1	1	1
1.45	1	1	1	1
1.50			1	1
1.55			1	1
1.60			1	1
1.65			1	1
1.70			1	1

SUBJECT NUMBER 12

B	C1 S1	C1 S2	C2 S1	C2 S2
0.50	-1	-1		
0.55	-1	-1		
0.60	-1	-1		
0.65	-1	-1		
0.70	-1	-1		
0.75	-1	-1	-1	-1
0.80	-1	-1	-1	-1
0.85	-1	-1	-1	-1
0.90	-1	-1	-1	-1
0.95	-1	-1	-1	-1
1.00	1	1	-1	-1
1.05	-1	1	-1	-1
1.10	-1	-1	-1	-1
1.15	-1	1	-1	-1
1.20	1	-1	-1	1
1.25	1	1	-1	-1
1.30	1	1	1	-1
1.35	1	1	1	1
1.40	1	1	1	1
1.45	1	1	1	1
1.50			1	1
1.55			1	1
1.60			1	1
1.65			1	1
1.70			1	1

Table 14 (Contd)

SUBJECT NUMBER 13

B	C1 S1	C1 S2	C2 S1	C2 S2
0.50	1	-1		
0.55	-1	-1		
0.60	-1	1		
0.65	1	1		
0.70	1	-1		
0.75	1	1	-1	-1
0.80	1	1	-1	-1
0.85	1	1	-1	-1
0.90	1	1	1	1
0.95	1	1	1	-1
1.00	1	1	1	-1
1.05	1	1	-1	1
1.10	1	1	1	1
1.15	1	1	1	1
1.20	1	1	1	1
1.25	1	1	1	1
1.30	1	1	1	1
1.35	1	1	1	1
1.40	1	1	1	1
1.45	1	1	1	1
1.50			1	1
1.55			1	1
1.60			1	1
1.65			1	1
1.70			1	1

SUBJECT NUMBER 14

B	C1 S1	C1 S2	C2 S1	C2 S2
0.50	-1	-1		
0.55	1	-1		
0.60	-1	-1		
0.65	-1	-1		
0.70	-1	-1		
0.75	1	-1	-1	-1
0.80	1	-1	-1	-1
0.85	1	-1	-1	-1
0.90	-1	-1	-1	-1
0.95	1	1	-1	-1
1.00	1	-1	-1	-1
1.05	1	1	-1	-1
1.10	1	-1	1	-1
1.15	1	1	1	-1
1.20	1	1	1	-1
1.25	1	1	1	-1
1.30	1	1	1	1
1.35	1	1	1	1
1.40	1	1	1	1
1.45	1	1	1	1
1.50			1	1
1.55			1	1
1.60			1	1
1.65			1	1
1.70			1	1

SUBJECT NUMBER 15

B	C1 S1	C1 S2	C2 S1	C2 S2
0.50	-1	-1		
0.55	-1	-1		
0.60	-1	-1		
0.65	-1	-1		
0.70	-1	-1		
0.75	-1	1	-1	-1
0.80	1	-1	-1	-1
0.85	1	-1	-1	-1
0.90	1	-1	-1	-1
0.95	1	1	-1	-1
1.00	1	1	-1	-1
1.05	1	1	1	-1
1.10	1	1	1	-1
1.15	1	1	1	1
1.20	1	1	1	1
1.25	1	1	1	1
1.30	1	1	1	1
1.35	1	1	1	1
1.40	1	1	1	1
1.45	1	1	1	1
1.50			1	1
1.55			1	1
1.60			1	1
1.65			1	1
1.70			1	1

SUBJECT NUMBER 16

B	C1 S1	C1 S2	C2 S1	C2 S2
0.50	-1	-1		
0.55	-1	-1		
0.60	-1	-1		
0.65	1	-1		
0.70	-1	-1		
0.75	1	-1	-1	-1
0.80	1	1	-1	-1
0.85	1	-1	-1	-1
0.90	1	1	-1	-1
0.95	1	1	-1	-1
1.00	1	1	1	-1
1.05	1	1	-1	-1
1.10	1	1	1	-1
1.15	1	1	1	1
1.20	1	1	1	1
1.25	1	1	1	1
1.30	1	1	1	1
1.35	1	1	1	1
1.40	1	1	1	1
1.45	1	1	1	1
1.50			1	1
1.55			1	1
1.60			1	1
1.65			1	1
1.70			1	1

Table 14 (Contd)

SUBJECT NUMBER 17

B	C1 S1	C1 S2	C2 S1	C2 S2
0.50	-1	-1		
0.55	-1	-1		
0.60	-1	-1		
0.65	-1	-1		
0.70	-1	-1		
0.75	-1	-1	-1	-1
0.80	-1	-1	-1	-1
0.85	-1	-1	-1	-1
0.90	-1	-1	-1	-1
0.95	-1	-1	-1	-1
1.00	-1	-1	-1	-1
1.05	-1	-1	-1	-1
1.10	-1	-1	-1	-1
1.15	-1	-1	-1	-1
1.20	-1	-1	-1	-1
1.25	-1	-1	-1	-1
1.30	1	-1	1	-1
1.35	1	1	-1	-1
1.40	1	-1	1	-1
1.45	1	1	1	1
1.50			1	1
1.55			1	1
1.60			1	1
1.65			1	1
1.70			1	1

SUBJECT NUMBER 18

B	C1 S1	C1 S2	C2 S1	C2 S2
0.50	-1	-1		
0.55	-1	-1		
0.60	-1	-1		
0.65	-1	-1		
0.70	-1	-1		
0.75	-1	-1	-1	-1
0.80	-1	-1	-1	-1
0.85	-1	-1	-1	-1
0.90	-1	-1	-1	-1
0.95	-1	-1	-1	-1
1.00	-1	-1	-1	-1
1.05	-1	-1	-1	-1
1.10	-1	-1	-1	-1
1.15	-1	-1	-1	-1
1.20	-1	-1	-1	-1
1.25	-1	-1	-1	-1
1.30	-1	-1	-1	-1
1.35	-1	-1	-1	-1
1.40	-1	1	-1	-1
1.45	1	-1	1	1
1.50			-1	1
1.55			1	1
1.60			1	1
1.65			1	1
1.70			1	1

SUBJECT NUMBER 19

B	C1 S1	C1 S2	C2 S1	C2 S2
0.50	-1	-1		
0.55	-1	-1		
0.60	-1	-1		
0.65	-1	-1		
0.70	-1	-1		
0.75	-1	-1	-1	-1
0.80	1	-1	-1	-1
0.85	-1	-1	-1	-1
0.90	-1	-1	-1	-1
0.95	1	1	-1	-1
1.00	1	1	-1	-1
1.05	1	1	-1	-1
1.10	1	1	-1	-1
1.15	1	1	-1	-1
1.20	1	1	1	-1
1.25	1	1	-1	-1
1.30	1	1	1	1
1.35	1	1	1	1
1.40	1	1	1	1
1.45	1	1	1	1
1.50			1	1
1.55			1	-1
1.60			1	1
1.65			1	1
1.70			1	1

placed in a lopsided 3-way table. The entry for subject number 1 in row .50, column S1 and layer C1 is -1. This means that the subject judged S_1 wider than the first element of C_1 (which has $\underline{b} = .50$). An entry of 1 would mean that the opposite judgment occurred. To give one more example from the data of subject 1, consider the entry for row 1.55, column S2 and layer C2, which happens to be 1. This means that the element of C_2 with $\underline{b} = 1.55$ (the 17th) was judged wider than S_2 .

Analysis

To analyze these data we will apply the countback method for constructing confidence limits. Before deciding on a value for c , however, we should consider the way in which we would like to use the confidence limits.

In terms of an overall explanation of each subject's behavior, there are two natural hypotheses to apply. We may call them 2D and 3D. The 2D hypothesis is that subjects ignore the perspective cues and simply match the widths of the projected image. In the 3D hypothesis, it is assumed that the subjects treat the images as though they were the objects which generated them. We will interpret these hypotheses as giving locations for the 50% point in each of the four series.

The instructions attempted to direct the subjects toward 3D, but there were subjects who spontaneously reported their interpretation in terms of 2D.[4] On these grounds alone, we might expect to find some subjects of each type.

---

[4] Subject 18 was one of these.

An interesting question arises as to whether there are any subjects for whom neither 2D nor 3D is applicable. Preliminary observations suggested that this would be the most common situation. Consequently, the comparison series were chosen to be most appropriate to responses representing a compromise between 2D and 3D. In fact, perfect 3D behavior would require a subject to report every element in C_1 and C_2 as wider than either S_1 or S_2. On the other hand, a perfect 2D subject would find S_1 and S_2 wider than every element in C_1 and wider than all but the largest five in C_2. The effect of this choice of series is that confidence intervals for such subjects may be open-ended on one side or the other.

Returning to the discussion of confidence limits, there are many ways they could be used to test these hypotheses for each subject. The way used here is to reject a given hypothesis for a subject if at least three out of the four predicted 50% points do not fall between their appropriate confidence limits.

If the confidence coefficient for each confidence interval taken separately is $1 - P$, what is the probability that at least three of the four fail to contain their 50% point for at least one of N subjects? For any given subject, the probability of this occurrence is just the probability that none, or only one, contains its 50% point, or $P^4 + 4P^3(1 - P) = 4P^3 - 3P^4$. The probability of this happening for at least one of the N subjects must be at most $N(4P^3 - 3P^4)$. (This would be the exact value only if occurrences for individual subjects were mutually exclusive events.)

In the present case, with $N = 19$, confidence limits were found using $c = 3$. (These are reproduced in Table 15.) For this value of c , we know that $1 - P$ is greater than .875. Thus, the probability of three or four 50% points falling outside their confidence intervals for at least one of the 19 subjects is less than .134521. In view of the conservative nature of the lower bound on $1 - P$, this is quite a reasonable value. (It may be noted that, if $1 - P$ were in fact .95 for each subject, the probability would drop to .009144.)

Looking at Table 15, we can see that possible 3D subjects are 7, 8, 10, and 13. (Subject 8 is actually subject 7 trying again.) 3D must be rejected for all the others. Possible 2D subjects are 3, 17, and 18. Consequently, subjects 1, 2, 4, 5, 6, 9, 11, 12, 14, 15, 16, and 19 are neither 2D nor 3D, but somewhere in between.

Table 15

Countback Confidence Limits for the 50% Point of Each of the Four Series for Each Subject; Given in Terms of the Comparison Widths, with c = 3

Subject	S_1 vs C_1	S_2 vs C_1	S_1 vs C_2	S_2 vs C_2
1	(.50, 1.00)	(-∞, 1.00)	(.85, 1.30)	(1.00, 1.35)
2	(.70, 1.05)	(.75, 1.10)	(1.10, 1.45)	(1.05, 1.40)
3	(1.20, ∞)	(1.20, ∞)	(1.15, 1.50)	(1.20, 1.55)
4	(-∞, 1.05)	(.75, 1.10)	(1.10, 1.45)	(1.05, 1.40)
5	(1.05, 1.40)	(.90, 1.40)	(1.10, 1.45)	(1.10, 1.45)
6	(-∞, .85)	(.50, .85)	(.95, 1.30)	(1.00, 1.35)
7	(-∞, .70)	(-∞, 1.00)	(-∞, 1.30)	(.85, 1.20)
8	(-∞, .90)	(-∞, 1.05)	(-∞, 1.30)	(.85, 1.30)
9	(1.00, 1.35)	(.90, 1.35)	(1.10, 1.45)	(1.15, 1.50)
10	(-∞, .65)	(-∞, .75)	(-∞, 1.00)	(.75, 1.25)
11	(.70, 1.05)	(.80, 1.15)	(1.00, 1.35)	(1.10, 1.45)
12	(.85, 1.30)	(.85, 1.35)	(1.10, 1.45)	(1.05, 1.45)
13	(-∞, .75)	(-∞, .85)	(.75, 1.20)	(.75, 1.15)
14	(-∞, 1.05)	(.80, 1.25)	(.90, 1.25)	(1.10, 1.45)
15	(.60, .95)	(.60, 1.05)	(.85, 1.20)	(.95, 1.30)
16	(.50, .85)	(.65, 1.00)	(.85, 1.20)	(.95, 1.30)
17	(1.10, 1.45)	(1.20, ∞)	(1.15, 1.50)	(1.25, 1.60)
18	(1.25, ∞)	(1.25, ∞)	(1.30, 1.65)	(1.25, 1.60)
19	(.65, 1.05)	(.75, 1.10)	(1.05, 1.40)	(1.15, 1.70)

References

No attempt is made here to provide a comprehensive list of work on the sensitivity problem. For something approaching such a list, see Finney (1964).

Berkson, J. Approximation of chi-square by "probits" and by "logits." Journal of the American Statistical Association, 1946, 41, 70-74.

Berkson, J. Minimum X^2 and maximum likelihood solution in terms of a linear transformation, with particular reference to bio-assay. Journal of the American Statistical Association, 1949, 44, 273-278.

Berkson, J. A statistically precise and relatively simple way of estimating the bio-assay with quantal response, based on the logistic function. Journal of the American Statistical Association, 1953, 48, 565-599.

Berkson, J. Maximum likelihood and minimum chi-square estimates of the logistic function. Journal of the American Statistical Association, 1955, 50, 130-162. (a)

Berkson, J. Estimates of the integrated normal curve by minimum normit chi-square with particular reference to bio-assay. Journal of the American Statistical Association, 1955, 50, 529-549. (b)

Berkson, J. Estimation by least squares and maximum likelihood. In J. Neyman (Ed.), Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability. Vol. I: Contributions to the theory of statistics. Berkeley: University of California Press, 1956.

Bliss, C. I. The method of probits. Science, 1934, 79, 38-39.

Bliss, C. I. The determination of dosage-mortality curves from small numbers. Quarterly Journal of Pharmacy and Pharmacology, 1938, 11, 192-216.

Bock, R. D., & Jones, L. V. The measurement and prediction of judgment and choice. San Francisco: Holden-Day, 1968.

Brown, B. W. Some properties of the Spearman estimator in bioassay. Biometrika, 1961, 48, 293-302.

Cochran, W. G. Estimation of bacterial densities by means of the "most probable number." Biometrics, 1950, 6, 105-116.

Dixon, W. J., & Mood, A. M. A method for obtaining and analyzing sensitivity data. Journal of the American Statistical Association, 1948, 43, 109-126.

Finney, D. J. The estimation from individual records of the relationship between dose and quantal response. Biometrika, 1947, 34, 320-334.

Finney, D. J. Probit analysis: A statistical treatment of the sigmoid response curve. London: Cambridge University Press, 1952.

Finney, D. J. Statistical method in biological assay. (2nd ed.) New York: Hafner, 1964.

Fisher, R. A. On the mathematical foundations of theoretical statistics. Philosophical Transactions of the Royal Society, 1922, A 222, 309-368.

Fisher, R. A., & Yates, F. Statistical tables for biological, agricultural and medical research. (6th ed.) Edinburgh: Oliver and Boyd, 1963.

Gaddum, J. H. Reports on biological standards. III. Methods of biological assay depending on a quantal response. Medical Research Council Special Report Series, No. 183. London: His Majesty's Stationery Office, 1933.

Harris, E. K. Confidence limits for the LD_{50} using the moving average-angle method. Biometrics, 1959, 15, 424-432.

Kendall, M. G., & Stuart, A. The advanced theory of statistics. Vol. 2. (2nd ed.) London: Charles Griffin, 1967.

Little, R. E. A note on estimation for quantal response data. Biometrika, 1968, 55, 578-579.

Lord, F. M., & Novick, M. R. Statistical theories of mental test scores. Reading, Mass.: Addison-Wesley, 1968.

Miller, L. C. Biological assays involving quantal responses. Annals of the New York Academy of Sciences, 1950, 52, Art. 6: The place of statistical methods in biological and chemical experimentation, 903-919.

Wilks, S. S. Mathematical statistics. New York: Wiley, 1962.

Recent Applications of Maximum Likelihood
Techniques to Sensitivity Data

What follows is a sample of experiments published during 1968. This sample was chosen primarily to give an idea of the kinds of data presently being analyzed by some of the methods we have discussed. It also indicates journals where further examples may be found.

Probit Analysis

The anticonvulsant potency of inhibitors of carbonic anhydrase in young and adult rats and mice. Journal of Pharmacy and Experimental Therapeutics, 161, 329-334. C. E. Rauh and W. D. Gray.

Background light, temperature and visual noise in the turtle. Vision Research, 8, 787-800. W. R. A. Muntz and D. P. M. Northmore.

Effects of yeast and yeast fractions on the action of the so-called heat-stable exotoxin of Bacillus thuringiensis in Drosophila melanogaster. Journal of Invertebrate Pathology, 10, 379-386. J. M. Perron and G. Benz.

A five-year study of potential western corn rootworm resistance to diazinon and phorate in Nebraska. Journal of Economic Entomology, 61, 496-498. H. J. Ball.

Hypnotic and hypothermic effect of pregnant mare urine extracts in the rat. European Journal of Pharmacology, 4, 270-282. K. Voith and B. A. Kovacs.

The relationship of exposure rate and exposure time to radiation injury in sheep. Radiation Research, 33, 94-106. N. P. Page, E. J. Ainsworth and G. F. Leong.

Role of thymus in tolerance. Journal of Immunology, 101, 1322-1332. A. Horiuchi and B. H. Waksman.

Visual detection of signals in the presence of continuous and pulsed backgrounds. Perception and Psychophysics, 4, 207-213. B. Leshowitz, H. B. Taub, and D. H. Raab.

Dilution Series

Plate-dilution frequency technique for assay of microbial ecology. Applied Microbiology, 16, 330-334. R. F. Harris and L. E. Sommers.

Up-and-down Method

Induction of ovulation during pseudopregnancy in the rat. Neuroendocrinology, 3, 220-228. G. P. van Rees, J. A. M. J. van Dieten, E. Bijleveld, and E. R. A. Muller.