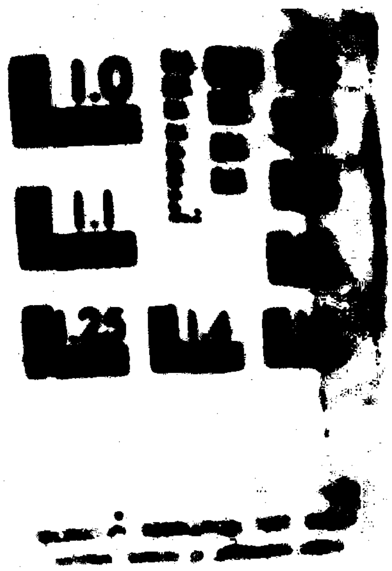
AD-A187 887 MODELING INCORRECT RESPONSES TO MULTIPLE-CHOICE ITEMS 1/2
WITH MULTILINEAR FO (U) ILLINOIS UNIV AT URBANA MODEL
BASED MEASUREMENT LAB F DRASGOW ET AL AUG 87
UNCLASSIFIED MEASUREMENT SER-87-1 N00014-83-K-0397 F/G 5/8 NL


1.0
1.1
1.25
1.4

MODELING INCORRECT RESPONSES TO MULTIPLE-CHOICE ITEMS WITH MULTILINEAR FORMULA SCORE THEORY

Fritz Drasgow, Michael V. Levine, Bruce Williams,
Mary E. McLaughlin, and Gregory Candell

Model-Based Measurement Laboratory
University of Illinois
210 Education Building
Champaign, IL 61820

August 1987

Prepared under contract No. N00014-83K-0397, NR 150-518
and No. N00014-86K-0482, NR 4421546
with the Personnel and Training Research Programs
Psychological Sciences Division
Office of Naval Research

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

A187 887

REPORT DOCUMENTATION PAGE

1a REPORT SECURITY CLASSIFICATION: Unclassified

1b RESTRICTIVE MARKINGS:

2a SECURITY CLASSIFICATION AUTHORITY:

2b DECLASSIFICATION / DOWNGRADING SCHEDULE:

3 DISTRIBUTION / AVAILABILITY OF REPORT: Approved for public release: distribution unlimited

4 PERFORMING ORGANIZATION REPORT NUMBER(S): Measurement Series 87-1

5 MONITORING ORGANIZATION REPORT NUMBER(S):

6a NAME OF PERFORMING ORGANIZATION: Michael V. Levine, Model-Based Measurement Lab.

6b OFFICE SYMBOL *(If applicable)*:

7a NAME OF MONITORING ORGANIZATION: Personnel and Training Research Programs, Office of Naval Research

6c ADDRESS *(City, State, and ZIP Code)*:
University of Illinois
210 Education Building, 1310 S. Sixth St.
Champaign, IL 61820

7b ADDRESS *(City, State, and ZIP Code)*:
Code 1142PT
800 North Quincy Street
Arlington, VA 22217

8a NAME OF FUNDING / SPONSORING ORGANIZATION:

8b OFFICE SYMBOL *(If applicable)*:

9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER:

8c ADDRESS *(City, State, and ZIP Code)*:

10 SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO	PROJECT NO	TASK NO	WORK UNIT ACCESSION NO
61153N	RR04204	RR04204-01	NR150-518 NR 4421546

11 TITLE *(Include Security Classification)*: Modeling Incorrect Responses to Multiple-Choice Items with Multilinear Formula Score Theory

12 PERSONAL AUTHOR(S): Drasgow, Fritz, Levine Michael V., Williams, Bruce, McLaughlin, Mary E., and Gregory Candell

13a TYPE OF REPORT: Technical Report

13b TIME COVERED: FROM ____ TO ____

14 DATE OF REPORT *(Year, Month, Day)*: 1987, August

15 PAGE COUNT: 117

16 SUPPLEMENTARY NOTATION:

17 COSATI CODES

FIELD	GROUP	SUB-GROUP
05	09	
	10	

18 SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*: Item response theory, polychotomous measurement, wrong answers, nonparametric modeling, formula score theory, parameter estimation, model fit. (cont.)

19 ABSTRACT *(Continue on reverse if necessary and identify by block number)*

Multilinear Formula Score theory provides powerful methods for solving psychological measurement problems of long standing. In this paper the question of information in incorrect option selection on multiple choice items is addressed. Multilinear formula scoring (MFS) is first used to estimate option characteristic curves for the Armed Services Vocational Aptitude Battery Arithmetic Reasoning test. Accurately estimated curves are obtained for real and simulated data. Then the statistical information about ability is computed for dichotomous and polychotomous scorings of the items. Moderate gains in information are obtained for low to slightly above average abilities. The dichotomous and polychotomous models are then compared for their relative performances in appropriateness measurement. The rates of detection of some types of aberrance responding were more than 100% higher for optimal polychotomous model index. Consequently the MFS polychotomous model provides opportunities for better testing by allowing more accurate ability estimates, improvements in the theory and practice of item writing, and more powerful appropriateness measurement.

20 DISTRIBUTION / AVAILABILITY OF ABSTRACT: ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS

21 ABSTRACT SECURITY CLASSIFICATION: unclassified

22a NAME OF RESPONSIBLE INDIVIDUAL: Michael V. Levine

22b TELEPHONE *(Include Area Code)*: 217/333-0092

22c OFFICE SYMBOL:

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

18. Suject terms (continued)

appropriateness measurement, optimal statistical test, cheating, coaching.

Abstract

Multilinear Formula Score theory provides powerful methods for solving psychological measurement problems of long standing. In this paper the question of information in incorrect option selection on multiple choice items is addressed. Multilinear formula scoring (MFS) is first used to estimate option characteristic curves for the Armed Services Vocational Aptitude Battery Arithmetic Reasoning test. Accurately estimated curves are obtained for real and simulated data. Then the statistical information about ability is computed for dichotomous and polychotomous scorings of the items. Moderate gains in information are obtained for low to slightly above average abilities. The dichotomous and polychotomous models are then compared for their relative performances in appropriateness measurement. The rates of detection of some types of aberrance responding were more than 100% higher for optimal polychotomous appropriateness indices than any dichotomous model index. Consequently the MFS polychotomous model provides opportunities for better testing by allowing more accurate ability estimates, improvements in the theory and practice of item writing, and more powerful appropriateness measurement.

Introduction

Multilinear formula score theory or multilinear formula scoring (MFS; Levine, 1985a, 1985b) is a nonparametric item response theory for which consistent and asymptotically efficient estimators of ability densities, item characteristic curves (ICCs), and option characteristic curves (OCCs) have been derived and programmed. MFS provides a powerful new approach to substantive questions of long standing. These questions include determining the shapes of ability distributions and the magnitudes of differences between ability distributions of various groups, determining the shapes of item characteristic curves for unidimensional and multidimensional tests, identifying biased and other faulty items, and assessing the extent to which two tests measure the same ability.

In this paper we focus on MFS's ability to estimate efficiently option response curves from small samples for responses that may fail to satisfy the local independence assumption of item response theory. The benefits of this endeavor shall be assessed in two ways. First, we determine the increase in information about ability due to polychotomous scoring of item responses. Here the term "information" is used in its statistical sense to mean the expected squared derivative of the logarithm of the likelihood function. Since the asymptotic standard error of the maximum likelihood estimate of an ability θ equals the square root of the reciprocal of the information function at θ, an increase in information due to polychotomous scoring is readily translated into percent test length reduction made possible by polychotomous scoring.

The second comparison is between the dichotomous and polychotomous item response model's potentials for supporting appropriateness measurement. Levine and Rubin (1979) introduced this term to refer to model-based methods for detecting response patterns that yield faulty measures of ability. For example, test scores are spuriously high when a low ability examinee copies some answers from a high ability neighbor or has been given answers to some questions by an informant. Spuriously low test scores result from alignment errors, atypical educations, unusual creativity, deliberate failure, and a variety of other sources.

Of course, the model-based detectability of a particular type of aberrance depends upon the item response model used to analyze the data; more specific (polychotomous) models are expected to be rejected more frequently when fitted to aberrant response patterns and thus provide superior appropriateness measurement. Recently Levine and Drasgow (1984, 1987) developed a technique for computing the power of the most powerful appropriateness measurement procedure supported by an item response model. By combining the new optimality results with MFS's ability to accurately recover the option characteristic curves needed for polychotomous modeling we intend to determine whether polychotomous modeling is negligibly or markedly superior to dichotomous modeling in detecting test anomalies.

This study also contributes to formula score theory in that it contains a verification of MFS theoretical results with simulation data.

Review of Multilinear Formula Score Theory

In this section we review MFS theory as it is used in this paper. The theory is more general than outlined here, but for the sake of clarity we describe only the special case required for the present research.

Let u_i denote the response to the ith item of an n item test scored $u_i = 1$ if correct and $u_i = 0$ if incorrect. The u_i generate the <u>elementary formula scores</u>, which can be enumerated as

$$
\begin{array}{l}
1 \\
u_1, u_2, \ldots, u_n \\
u_1u_2, u_1u_3, \ldots, u_{n-1}u_n \\
\ldots \\
u_1u_2 \ldots u_n .
\end{array}
$$

Traditional formula scoring (Lord and Novick, 1968, especially Chapter 14) generally uses only linear scores. When there is neither omitting nor polychotomous scoring, <u>linear formula scores</u> are formulas with a constant term plus a linear combination of the binary item scores, $u_1, u_2, \ldots, u_n$. (When there is omitting and polychotomous scoring, a linear score is a constant plus a <u>linear</u> combination of binary variables indicating omitting and option choice.)

Multilinear formula score theory generalizes traditional formula score theory by using quadratic scores (linear scores added to linear combinations of $u_1u_2, u_1u_3, \ldots, u_{n-1}u_n$), cubic scores (quadratic scores plus linear combinations of products of item scores for three different items), and higher order scores. Most of the results in this paper were obtained with fifth order scores. The new theory is called "multilinear" because frequent use is made of the fact that when all the scores except one are held constant, a "linear" score is obtained.

In this paper we shall assume that the regression of u_i on the latent trait θ is a three-parameter logistic ogive

$$
\begin{aligned}
E(u_i \mid \theta = t) &= c_i + \frac{1 - c_i}{1 + \exp[-Da_i(t - b_i)]} \\
&= P_i(t) ,
\end{aligned}
$$

where D is a scaling constant set equal to 1.702, a_i is the discrimination parameter, b_i is the difficulty parameter, and c_i is the lower asymptote of the ICC. By local independence, the regressions of the elementary formula scores on the latent trait can then be written

$$
\begin{array}{l}
1 \\
P_1(t),\ P_2(t),\ \ldots,\ P_n(t) \\
P_1(t)P_2(t),\ P_1(t)P_3(t),\ \ldots,\ P_{n-1}(t)P_n(t) \\
\ldots \\
P_1(t)P_2(t)\ \ldots\ P_n(t)\ ,
\end{array}
$$

where each $P_i(t)$ is a three-parameter logistic ICC.

There are 2^n regression functions listed above. More can be generated by taking linear combinations of the elementary formula scores and then computing their regressions on the latent trait. For example, the number-right score

$$X = u_1 + u_2 + \ldots + u_n\ ,$$

has the regression

$$E(X \mid t) = \sum_{i=1}^{n} P_i(t)\ .$$

The collection of regression functions of <u>all</u> linear combinations of elementary formula scores is called the <u>canonical space</u> of a test.

A major step in a MFS analysis of a test consists of finding a smaller number of functions to represent the large number (in fact, an infinite number) of functions in the canonical space. The smaller collection of functions is called an <u>orthonormal basis</u> for the canonical space.

Selecting an orthonormal basis for the canonical space is analogous to finding the principal components of a set of variables. In a principal components analysis, the basic idea is to create a new set of variables, the principal components, so that each of the original variables can be written as a linear combination of the principal components plus a small residual. A principal components analysis is valuable when there is a large number of original variables and the first few principal components explain almost all of their variance. In the same way functions in the canonical space are written as linear combinations of the orthonormal basis functions. For example, the ICC for the ith item can be written

$$P_i(t) = \sum_{k=1}^{K} \alpha_k h_k(t) \ ,$$

where K functions, denoted $h_1(t), \ldots, h_K(t)$ are used in the orthonormal basis and the α_k are the weights used in the linear combination. If K is sufficiently large, this representation is exact. If only the first J functions are used, instead of all K functions (where J is less than K), then there is some error. However, the residual

$$P_i(t) - \sum_{k=1}^{J} \alpha_k h_k(t) = \sum_{J+1}^{K} \alpha_k h_k(t)$$

will be small if the α_k are small for values of k larger than J . In fact, the area under the squared residual is exactly $\alpha_{J+1}^2 + \alpha_{J+2}^2 + \ldots + \alpha_K^2$.

In each MFS analysis a parsimonious representation of one or another collection of functions in the CS is important. MFS provides techniques that yield basis functions that give small α_k for large k , at least for the collection of functions being analyzed. Most MFS analyses require six to eight basis functions for an adequate representation of the functions being studied. Ten were used in this study.

To recapitulate, the analysis begins by estimating ICCs from the dichotomously scored item responses. Widely available programs such as LOGIST and BILOG can be used to this end. The estimated ICCs (and the assumption of local independence) are subsequently used to define the canonical space. Then a small number of orthonormal basis functions are selected so that the functions in the canonical space are well-approximated by linear combinations of the orthonormal basis functions.

The next step of the MFS analysis is to use the orthonormal basis functions to represent the option characteristic curves (OCCs). For technical reasons (see below), we first estimate orthonormal basis function weights for conditional option characteristic curves (COCCs). A COCC gives the probability of an option choice given that the person does not choose the correct option. A COCC equals its associated OCC divided by $(1-P_i(\theta))$. Hence the COCCs for an item sum to 1 for all θ values whereas the OCCs sum to $1-P_i(\theta)$, which becomes very small for large θ values. Each option characteristic curve is then represented as the product of two linear combinations of the h_j's, namely the representation of $1-P_i$ and a COCC. At this point the OCC can be represented by a single set of weights by calculating weights b_j's such that $\sum b_j h_j(\cdot)$ is approximately equal to $(1-P_i)$ times the COCC value. (An exact representation is not possible in general because a product of two functions in the canonical space is not necessarily in the canonical space.)

Since OCCs and COCCs were not included in the set of functions used to define the canonical space, there is both the mathematical question of how best to approximate the OCCs and COCCs by basis functions and the substantive question of whether or not the basis functions <u>can</u> adequately approximate OCCs and COCCs. The analysis proceeds item-by-item with the weights for all the options (including omit as an option) to each item simultaneously estimated by "marginal" maximum likelihood. The log likelihood that is maximized with respect to the weights is

$$(1) \qquad L = \sum_{j=1}^{N} \log P(u_j^*, v_{ij}^*) ,$$

where u_j^* is a vector containing the dichotomously scored item responses of the jth examinee and v_{ij}^* indicates the particular option on item i selected by examinee j. For a four option multiple-choice item, $v_{ij}^* = 1$ if option A is selected, ... $v_{ij}^* = 4$ if option D is selected, and $v^*_{ij} = 5$ if no response is made. Suppose all the items are recoded so that option A is always the correct response. Then Equation 1 can be rewritten as

$$(2) \qquad L = \sum_{\substack{j=1 \\ v_{ij}^*=1}}^{N} \log P(u_j^*) + \sum_{\substack{j=1 \\ v_{ij}^*\neq 1}}^{N} \log \int P(u_j^* \mid t)\, P(v_{ij}^* \mid t, u_{ij}=0) f(t)dt$$

where

$$(3) \qquad P(u_j^* \mid t) = \prod_{i=1}^{n} P_i(t)^{u_{ij}}[1 - P_i(t)]^{1-u_{ij}} ,$$

$$(4) \qquad P(v_{ij}^* \mid t, u_{ij} = 0) = \sum_{k=1}^{J} \alpha_k h_k(t) ,$$

and $f(t)$ is the ability density. Notice that Equation 3 is the likelihood function for the three-parameter logistic model [i.e., Lord's (1980) Equation (4-20) and Hulin, Drasgow, & Parson's (1983) Equation (2.6.2)]. It is the α_k's in Equation 4 that are to be estimated. Actually, each option has its own set of J α_k's, but to avoid notational complexity we have not added another subscript to the α_k's.

It is important to observe that local independence is not used to derive Equation 2 from Equation 1; only the definition of conditional probability is used. Thus, even when skipping items or not reaching items (response "5") fails to obey the assumption of local independence, an accurate estimate of the conditional probability of non-response for examinees at each ability level may be obtained.

Quadratic programming is used to obtain maximum likelihood estimates of orthonormal basis function weights for the COCCs in Equation 4. The weights α_k for the COCCs are easier to estimate than the weights for OCCs since the OCCs for easy items and OCCs for rarely chosen options are close to zero, which causes the α_k to become indeterminate; COCCs are not usually close to zero. Since the OCC at $\theta = t$ is equal to the COCC times $1 - P_i(t)$, the OCCs are available after the COCCs have been obtained. The COCCs are intrinsically interesting as well as mathematically tractable since their shapes can be used to study the properties of effective distractors.

The quadratic programming methods used by Levine and Williams (1987) are convenient because they allow plausible constraints to be placed on the COCCs. One constraint is <u>positivity</u>: COCCs are not allowed to become negative. In our analyses all COCCs were required to equal or exceed .001. A second constraint placed on COCCs is <u>smoothness</u>: The COCCs were not allowed to oscillate widely. The smoothness constraint can be implemented by restricting the third derivative of the COCCs to be less than .005. This condition can be thought of as requiring each small piece of the graph of the COCC to have a very accurate quadratic approximation. (A restriction on the second derivative would force the COCC to be locally linear and a first derivative constraint would force the COCC to be locally constant.)

In summary, orthogonal basis functions $h_j(t)$ are derived from ICCs, which are estimated by programs such as LOGIST or BILOG. COCCs are represented as linear combinations of the basis functions in Eq. 4, and marginal maximum likelihood estimates of the weights α_j in this equation are obtained. OCC values can then be obtained by multiplying COCC values times $(1-P_i)$.

Estimation and Information

Data set. The data set used in our analyses was a spaced sample of 2978 examinees; this data set is fully described in the Profile of American Youth (1982). These examinees answered the 30 item Armed Services Vocational Aptitude Battery (ASVAB) Arithmetic Reasoning (AR) test. Each item on this test has four options.

ICC estimation. The first step in the MFS analysis is to estimate ICCs from the dichotomously scored item responses. To this end, the item responses of the examinees described above were scored dichotomously. All unanswered items were scored as incorrect (since skipping and not reaching are treated as a separate--and incorrect--response option). Then the LOGIST (version 2B) computer program (Wood, Wingersky, & Lord, 1976) was used to estimate item and person parameters. Estimates of item discrimination parameters ranged from about 0.5 to 2.0 and estimates of item difficulties varied from about -3.0 to 1.4 (mean = .14, SD = .99).

[illegible] estimated [illegible]

abilities, suggesting a relative large number of examinees with very low abilities. The substantive interpretation of this fat left tail is that even among examinees who answered more than half of the items there may have been some who were poorly motivated and did not make a serious attempt to pass the exam. In fact examinees were paid to take the exam and consequently some of them may not have been adequately motivated. The test information function at [illegible] is very low and consequently bimodality [illegible] larger samples.

Insert Figure [illegible] about here

Estimation. The number of [illegible]s estimated for each item was four [illegible] response curves and an omit curve. Omits include both "skipped" responses and "not reached" responses. The number of orthonormal [illegible] used in the analysis was [illegible]. Thus, [illegible] weights [illegible] weights [illegible] were estimated for each item. Since the sum of the [illegible] the weights for the fourth [illegible] turn out to be a known [illegible] of the weights for the other three.

[illegible] of the [illegible]s estimated for all [illegible] AM items. [illegible] the estimated [illegible]s [illegible] page of Appendix [illegible] [illegible] The upper left panel contains the [illegible] [illegible] the upper right panel contains the [illegible] for [illegible] the bottom left panel contains the [illegible] for the [illegible] and the bottom right panel contains the omit [illegible]. [illegible] of the [illegible] of the estimated [illegible]s can be [illegible] by examining the vertical lines displayed in each panel. These lines were obtained by computing three parameter logistic ability estimates for all 11,914 examinees in the American Youth data set, forming 25 ability strata on the basis of estimated abilities by using the 4th, 8th, ..., 96th percentile points of the standard normal distribution as cutting scores, and then computing the proportion of examinees selecting each option among the subset of examinees who answered the item incorrectly. The centers of the vertical lines correspond to the observed proportions and they are plotted above the category medians (the 2nd, 6th, ..., 98th percentile points of the standard normal distribution). The vertical lines represent approximate 95% confidence intervals for the observed proportions (± two standard errors, where the observed proportion is used to compute the standard error). Observed proportions of 0 and 1 are plotted as plus signs and are offset slightly from their true locations so that they will be visible.

The AR items seem to be more-or-less ordered by difficulty. Consequently, the 95% confidence intervals for the first few items in Appendix 1 are very wide because these items are easy and so few examinees choose incorrect options. Confidence intervals for later items are much narrower and provide a severe test for COCC estimates. Item 27, for example, shows that the COCC estimates provide a very good description of option choice. Notice that the COCC for the omit category lies below most observed proportions. This occurs because examinees with high omitting rates were excluded from the sample used to estimate COCCs, but were included in the total sample used to compute the proportions displayed in Appendix 1.

COCC estimation verification. The figures presented in Appendix 1 show that MFS estimates of COCCs closely follow the actual patterns of item responses. It is difficult, however, to understand the accuracy of COCC estimates from these figures because the true COCCs are not known. To gain further insights into the properties of MFS estimates of COCCs, a simulation data set of 3000 response patterns was generated. Simulated abilities were sampled from the standard normal distribution, probabilities of correct and incorrect responses were determined from the ICCs obtained by the LOGIST run described previously, and probabilities of option selections (for responses simulated to be incorrect) were computed using the MFS estimated COCCs. Thus, the assumptions used to estimate COCCs correspond exactly to the way in which the data set was generated.

COCCs were re-estimated from the simulation data set. The true ability density (the standard normal) was used in Equation 2 and the true ICC values were used to compute probabilities of correct and incorrect responses. The true ability density and ICC values were used because we wanted to determine the errors of COCC estimates in a way that was not confounded with inaccuracies in density estimates and ICC estimates.

The results of the simulation study are shown in Appendix 2, which presents the re-estimated COCCs for all 30 items. Heavy lines indicate the re-estimated COCCs and thin lines indicate the true COCCs. Observed proportions and their approximate 95% confidence intervals are shown for the simulation sample of N = 3000. The observed proportions are not plotted if five or fewer incorrect responses were made in an ability stratum.

Item 2 shows estimated COCCs that are very close to the true COCCs for all ability levels. This is remarkable because there were almost no incorrect responses made by simulated examinees with above average ability. Item 3 shows that we cannot always expect to have well-estimated COCCs when there are no data available: Large differences between true and estimated COCCs occur at high ability levels. The COCCs are, however, accurately estimated in ability ranges for which there were very few incorrect responses.

From an inspection of the plots in Appendix 2 it seems evident that COCC values are accurately estimated when there are six or more incorrect responses in adjacent ability strata. Sometimes COCC values are well-estimated when fewer incorrect responses are available, but this seems to be a matter of chance. Notice, also, that COCCs for the omit option are not underestimated in this analysis as they were in the analysis of the real AR data. In the present analysis, all response vectors were used; there was no restriction on omitting as in the previous analysis. In this simulation study data were unidimensional in the sense that the probability of omitting depended only on ability, although it was permitted to vary from item to item. It would have been more realistic to use a two dimensional simulation model with examinees varying both in ability and tendency to omit.

Information function. Information functions for the dichotomous and polychotomous modelings of the AR test are shown in Figure 2. An expression for the information function of the three-parameter logistic model is

$$(5)\qquad \text{Information at } t = \sum_i \frac{[P_i'(t)]^2}{P_i(t)} + \sum_i \frac{[Q_i'(t)]^2}{Q_i(t)} ,$$

where $Q_i = 1 - P_i$ and P_i' and Q_i' are the first derivatives of P_i and Q_i. The information function of the polychotomous model is

$$(6)\qquad \text{Information at } t = \sum_i \frac{[P_i'(t)]^2}{P_i(t)} + \sum_i \sum_{j=2}^{J} \frac{[P_{ij}'(t)]^2}{P_{ij}(t)}$$

where P_{ij} is the OCC for option j on item i and P_{ij}' is its first derivative. The correct option makes the same contribution to information for both the dichotomous and polychotomous scorings, namely, the first term on the right sides of Equations 5 and 6. Thus, any differences in information are entirely due to the treatment of incorrect responses. Using Jensen's inequality (Halmos, 1950) it can be shown that

$$\sum_{j=2}^{J} \frac{[P'_{ij}(t)]^2}{P_{ij}(t)} \geq \frac{[Q'_i(t)]^2}{Q_i(t)}$$

(cf., Samejima, 1969; Park, 1983). Thus, any increase in information is entirely due to polychotomous scoring.

Insert Figure 2 about here

Figure 2 shows that there are moderate gains in information due to polychotomous scoring of the AR items for low to moderately high abilities. Little or no information is gained for high ability examinees; this latter finding is not surprising because high ability examinees are expected to answer nearly all the items correctly.

It should be noted that the AR items were not written with polychotomous scoring in mind and so the gains in information shown in Figure 2 are more-or-less accidental. Larger gains might be realized if item writers knew the attributes of incorrect options that typically lead to substantial increases in information.

Appropriateness Measurement

Purpose

In this section we compare the effectivenesses of dichotomous and polychotomous models for detecting aberrant responses patterns. By comparing detection rates of optimal indices it is possible to compare the maximum detection rates possible for a given form of aberrance. In this section, as in the previous section, the dichotomous model is a submodel of the polychotomous model; hence any increase in detection rates is due to modeling incorrect responses.

For an optimal index to be truly optimal, it must be computed from the true ICCs or OCCs and, therefore, the optimal indices for dichotomous and polychotomous scorings of the simulation data were computed using the simulation ICCs and OCCs. In any practical application, however, only estimated ICCs and OCCs will be available. Consequently, we decided to examine one aspect of the robustness of optimal indices by computing the optimal index for dichotomously scored responses using ICCs estimated by the LOGIST (Wood, Wingersky, & Lord, 1976) computer program. Further research designed to develop extensions of optimal indices for use in practical settings will be warranted if the optimal indices computed from estimated ICCs are found to be nearly as powerful as optimal indices computed from the true ICCs.

Several practical indices were also evaluated. Most of these indices were computed from the dichotomously scored item responses. One index, however, is the natural extension of a dichotomous model index to the polychotomous case. Detection rates for the practical indices indicated (1) which were relatively more powerful and less powerful; and (2) the extent to which the maximum detection rates were attained.

Overview

The ICCs and OCCs estimated for the AR test from the sample of N = 2,978 were used as the "true" item parameters in a simulation study. Initially, a sample of N = 3000 simulated response patterns was created and used as a test norming sample. This data set was used to determine the item and test statistics required to compute all but one (z_p) of the practical appropriateness indices listed in the next section. Then a normal sample of N = 4000 responses vectors was created. In addition, sixteen aberrant samples of N = 2000 were generated to simulate several forms of aberrance. Optimal indices and all the practical indices were then computed for the normal sample and aberrant samples. Rates of detection of aberrant responses vectors at various false alarm rates were determined for each appropriateness index and each form of aberrance.

Appropriateness Indices

In this section we list the appropriateness indices that are evaluated. For the sake of brevity we shall not provide extensive technical detail. This information is given by Levine and Drasgow (1984; 1987) for optimal indices and by Drasgow, Levine, and McLaughlin (1987) for practical indices. Additional references are given when appropriate.

Polychotomous model optimal indices (LR_p). Levine and Drasgow (1984) used the Neyman-Pearson lemma to derive a class of most powerful appropriateness indices. These indices require the probabilities of observing the polychotomously scored response vector v^* assuming that it was generated by a normal process $(P_{Normal}(v^*))$ and assuming that it was generated by a specified aberrant process $(P_{Aberrant}(v^*))$. The decision procedure that classifies response vectors as aberrant when

$$P_{Aberrant}(v^*) \geq \text{constant} \cdot P_{Normal}(v^*) ,$$

where the constant is chosen to control the false alarm rate or Type I error rate, is at least as powerful as any other test. Thus, the polychotomous model optimal indices studied here have the form

$$LR_p = P_{\text{Aberrant}}(v^*) \;/\; P_{\text{Normal}}(v^*) \;,$$

where the probabilities are computed using three-parameter logistic ICCs to determine conditional probabilities of correct responses and MFS OCCs to determine conditional probabilities of incorrect responses. Technical details about the form of LR_p for specific types of aberrance and an efficient computing algorithm are given by Levine and Drasgow (1984; 1987).

<u>Dichotomous model optimal indices</u> (LR_3). These indices are identical to the LR_p indices except that the only information used in their calculation is the pattern of correct and incorrect responses u^*, i.e., the dichotomously scored item responses. This class of indices, therefore, provides the highest rates of detection when the choice of incorrect option is ignored.

<u>Dichotomous model optimal indices computed using estimated item parameters</u> (LR'_3). For optimal indices to be truly optimal they must be computed using item parameters -- not item parameter estimates. In previous work (Levine & Drasgow, 1982), we found that the values of some appropriateness indices were almost unaffected when item parameter estimates were used in place of item parameters. In the present research, optimal indices for the three-parameter logistic model were also computed using estimated item parameters.

<u>Dichotomous and polychotomous model standardized ℓ_0</u> (z_3 and z_p). These indices, originally developed by Drasgow, Levine, and E. Williams (1985), are well-standardized (i.e., their conditional distributions given

ability are nearly invariant across ability levels) and are, therefore, well-suited to practical applications. In essence, they compare the likelihood of a vector of item responses to the expected likelihood given the examinee's ability estimate. In previous research (Levine & Rubin, 1979; Levine & Drasgow, 1982,; Drasgow, Levine, & E. Williams, 1985), it has been found that aberrant response vector tend to have likelihoods that are smaller than expected of normal response vectors, and thus, the standardized likelihoods z_3 and z_p serve as effective appropriateness indices.

Fit statistic (F1 and F2). Two fit statistics suggested by Rudner (1983) as generalizations of Rasch model fit statistics used by Wright and his colleagues are

$$F1 = \frac{1}{n} \sum_{i=1}^{n} [u_i - P_i(\hat{\theta})]^2 / [P_i(\hat{\theta})Q_i(\hat{\theta})]$$

and

$$F2 = \frac{1}{n} \sum_{i=1}^{n} [u_i - P_i(\hat{\theta})]^2 / \sum_{i=1}^{n} P_i(\hat{\theta})Q_i(\hat{\theta}) .$$

Notice that F1 and F2 tend to be large when an examinee misses items ($u_i = 0$) that should be answered correctly ($P_i(\hat{\theta})$ near 1) and correctly answers ($u_i = 1$) items that should be very difficult ($P_i(\hat{\theta})$ near 0) . Drasgow, Levine, and McLaughlin (1987) found F2 to be well-standardized. F1, however, was badly standardized because relatively many large values were observed for simulated normal, high ability examinees.

Caution indices (S, T2, and T4). Three "caution" indices were evaluated. The first is the original Sato caution index S described by Sato (1975) and Tatsuoka and Linn (1983). The other two caution indices are the second (T2) and fourth (T4) standardized extended caution indices developed by Tatsuoka (1984). Drasgow, Levine, and McLaughlin (1987) found T4 to be better standardized than T2, so T4 should be preferred when their detection rates are comparable.

Likelihood function curvature statistics (JK and O/E). It is expected that the likelihood function will be "flatter" for aberrant response vectors than normal response vectors at the maximum likelihood ability estimate $\hat{\theta}$. Two indices that provide measures of the flatness of the likelihood function were evaluated. The first (JK) is a normalized jackknife estimate of the variance of $\hat{\theta}$ and the second is the ratio of the observed and expected information about ability contained in the dichotomously scored item responses. Both of these indices are described by Drasgow, Levine, and McLaughlin (1987), who showed that JK and O/E are well standardized.

Method

Data Sets. A *test norming* sample of 3000 responses vectors was created by sampling 3000 numbers (θ's) from the normal (0,1) distribution truncated to the [-5.0, 3.5] interval. A normal sample of 4000 response vectors was also generated in this way. Two thousand aberrant response vectors were created in each of sixteen conditions. These conditions resulted from varying three factors: the type of aberrance (spuriously high; spuriously low), the severity of aberrance (mild; moderate), and the distribution from which simulated abilities were sampled.

Eight of the aberrant samples contained spuriously high response vectors and the remaining eight samples contained spuriously low responses vectors. Spuriously high responses patterns were created by first generating normal response vectors (using the AR three-parameter logistic ICCs to determine the probabilities of correct responses and the AR COCCs to determine probabilities of incorrect option selection) and then replacing a given percentage $\underline{k}$ of simulated responses (randomly sampled without replacement) with correct responses. Spuriously low response patterns were also created by first generating normal response vectors. Then a fixed percentage of items were randomly selected without replacement and the responses to these items replaced with random responses (i.e., a response was replaced by option A with probability .25, by option B with probability .25, ..., and by option D with probability .25). Mildly aberrant response patterns were generated by using k = 17% (i.e., 5 of 30 items). Moderately aberrant response patterns were created using k = 33% (i.e., 10 of 30 items).

The third variable manipulated was the ability level of the aberrant sample. Abilities for the spuriously high samples were sampled from four parts of the normal (0,1) distribution truncated to [-5.0, 3.5]: very low (0th through 9th percentiles), low (10th through 30th percentiles), low average (31st through 48th percentiles), and high average (49th to 64th percentiles). In all cases percentile points were determined after the truncation to [-5.0, 3.5]. These intervals were used because it is more important to detect spuriously high response patterns for low ability examinees than for high ability examinees. Similarly, it is more important to detect spuriously low responses by high ability examinees. Consequently, abilities were sampled from four average to high ability strata for the spuriously low samples: very high (93rd percentile and above), high (65th through 92nd percentiles), high average (49th through 64th percentiles), and low average (31st to 48th percentiles). The ability percentiles used here correspond to the percentiles forming United States Armed Service Vocational Aptitude Battery mental categories.

Analysis. All the item and test statistics required to compute the practical appropriateness indices were computed using the test norming sample. These quantities were computed as the first step in the analysis and then used in all subsequent analyses. LOGIST (Wood, Wingersky, & Lord, 1976) was used to estimate three-parameter logistic item parameters and a Fortran program was written to compute the other quantities required.

The practical appropriateness indices and LR_3' were then computed for the 4000 response vectors in the normal sample. The item and test statistics estimated from the test norming sample were used in these calculations. This procedure simulates the process by which practical appropriateness indices would be computed in many applications. Optimal indices were also computed for the normal sample for four aberrant conditions: 17% spuriously high, 33% spuriously high, 17% spuriously low, and 33% spuriously low. The ICCs and COCCs used to generated the data were used to compute LR_p and LR_3 .

The practical appropriateness indices were computed for each of the 16 aberrant samples. In addition, the 17% spuriously high optimal index was computed for the four samples with this form of aberrance, the 33% spuriously high optimal index was computed for the four samples with this form of aberrance, etc. The proper interpretation of the optimal indices computed in the present research is the following: They are optimal for the specified form of aberrance, say 17% spuriously high, in a population where the ability density is a truncated normal for both the normal and aberrant populations and a response vector is either normal or 17% spuriously high. The normal group does in fact have this ability distribution. By stratifying on a subinterval of [-5.0, 3.5] for the aberrant group, we determined the power of the index that is optimal for the population as a whole in a particular subpopulation.

Evaluation Criteria. The main criteria for evaluating the appropriateness indices were the proportions of aberrant response patterns correctly identified as aberrant when various proportions of normal response patterns were misclassified as aberrant. These proportions shall be presented for all 16 aberrance conditions. This allows us to determine what types of aberrant response patterns have acceptably high detection rates using optimal methods and using practical methods. The characteristics of response patterns that cannot be detected are evident as a consequence of examining the 16 aberrance conditions separately.

Results

The results for the spuriously high conditions are given in Tables 1 through 4. The results for the lowest ability group are shown in Table 1. In this table it is evident that cheating on five randomly selected items is not very detectable: At a 2% false alarm rate only 28% of the simulated cheaters are detected by the optimal LR_p index. The best of the practical indices, z_3 and F2 , detect 18% and 20%, respectively. Cheating on 10 items (the 33% condition) is reasonably detectable. For example, LR_p detects 61% and LR_3 detects 54% at a 2% false alarm rate. At this false alarm rate, z_3 , F2 , and T4 detect 44%, 41%, and 50%, respectively. Finally, detection rates for optimal indices computed from true and estimated ICCs are very similar for almost all false alarm rates in Table 1.

Insert Tables 1 through 4 about here

The dete[illegible]

small[illegible] [illegible] % [illegible]

[illegible] % [illegible] % [illegible] % [illegible]

[illegible] % [illegible] % [illegible] % [illegible]

[illegible] dete[illegible] rates [illegible]

[illegible]

[illegible]

[illegible]

[illegible] significant difference [illegible]

[illegible] the frequency and relative frequency [illegible]

[illegible]criminating [illegible] items with low lower asymptotes [illegible]

Seventeen of the [illegible] AT items satisfy these three conditions. [illegible] contrast, none of the [illegible] AR items meet these conditions and only three items have [illegible]. In sum, high average examinees have reasonably good chances [illegible] and correctly [illegible] each AR item and so correct responses obtained by cheating are not clearly aberrant.

The results for the spuriously low samples are given [illegible] through [illegible] Table [illegible] it is evident that [illegible]% spuriously low responding [illegible] simulated low average examinees is moderately detectable [illegible]% detection rate with [illegible]% false alarms [illegible] other [illegible] higher detection rates [illegible] high average examinees [illegible] Table [illegible] Again [illegible] better than [illegible] high rates of detection [illegible] high examinees are shown in Tables [illegible] index, with detection rates of [illegible]% and [illegible]% for [illegible]% false alarm rate [illegible] [illegible]% spuriously low treatment.

[illegible], has detection rates that are close to [illegible], in all of the tables. Thus, little power for detecting inappropriate response patterns is lost when the true three parameter logistic [illegible]s are replaced by estimated [illegible]s.

Insert Tables [illegible] through [illegible] about here

<u>Discussion</u>

In this paper we have described and implemented Levine's (1985a, 1985b) approach to polychotomous measurement. It was used to estimate [illegible]s for a sample of [illegible] examinees who responded to the ASVAP AR test. Good to excellent fits were obtained when the estimated [illegible]s were compared to empirical proportions computed form the responses of a larger sample of [illegible] examinees. A simulation data set was also used to investigate [illegible] estimates. Very accurate estimates were obtained for ability ranges with sufficient numbers of examinees who responded incorrectly.

The test information function of the polychotomous model was found to be moderately larger than the three-parameter logistic information function for low to moderately high ability levels. Since there <u>is</u> information in incorrect options it seems prudent to use it if items are expensive to write, the number of items that can be administered is severely limited, or very accurate ability estimates are required. Furthermore, we can now study systematically the differences in items with informative incorrect options and items with essentially noninformative incorrect options. It may be possible to identify different characteristics of these two types of items and thereby help item writers increase the information about ability provided by tests by writing items with highly informative incorrect options.

An appropriateness measurement simulation study was also conducted to compare the polychotomous model with a dichotomous submodel, namely the three-parameter logistic. Several important results were obtained. First, for the spuriously low treatment that simulates atypical educations, misgridding answers to a portion of the test, unusual creativity, etc., we found that optimal three-parameter logistic appropriateness indices fell far short of their optimal polychotomous model counterparts. At some false alarm rates, the rates of detection of aberrant response vectors were more than 100% higher for the polychotomous optimal indices. Thus appropriateness measurement constitutes one important practical testing problem where substantial gains are made by the use of a polychotomous item response model.

The results of the appropriateness measurement simulation study also showed that the practical polychotomous model index z_p was not a particularly good index: Its detection rates were not close to optimal for either spuriously high or spuriously low treatments. This result, in conjunction with the results described, previously point to the need to devise better polychotomous appropriateness indices that can be used in practical situations.

A third result obtained in the appropriateness measurement research is that the z_3 , F2 , and F4 indices effectively detect aberrance in relation to three-parameter logistic optimal indices (but not polychotomous model optimal indices). Therefore, if one is satisfied with dichotomous scoring of item responses for some particular application, then z_3 , F2 , F4 can be used with confidence to detect inappropriate test scores.

Means for implementing appropriateness measurement in practical settings are discussed by Drasgow and Guertler (1987).

Finally, the LR_3' indices provided detection rates that were nearly as high as the rates provided by the optimal LR_3 indices. Thus, the three-parameter logistic optimal indices seem to be robust to item parameter estimation error. This result is surprising because extensive computations are required to evaluate LR_3'; small errors (in ICC values) would be expected to grow progressively larger as the computations progressed. Nonetheless, only small differences between values of LR_3 and LR_3' were observed for individual response patterns. Thus, we are encouraged to continue research on "almost-optimal" indices that are based on likelihood ratios and could be used in practical settings.

Conclusion

COCC estimation provides opportunities to improve testing in a variety of ways: ability estimation, the theory and practice of item writing, appropriateness measurement. Applications in areas such as item and test bias and adaptive testing may also be fruitful. Consequently, we conclude that there is information in incorrect responses and that polychotomous item response models can make important contributions to psychological testing.

References

Drasgow, F. & Guertler, E. (1987). Detecting inappropriate test and scale scores in practical setting. Journal of Applied Psychology, 72, 10-18.

Drasgow, F., Levine, M.V., & McLaughlin, M.E. (1987). Detecting inappropriate test scores with optimal and practical appropriateness indices. Applied Psychological Measurement, 11, in press.

Drasgow, F., Levine, M.V., & Williams, E. (1985). Appropriateness measurement with polychotomous item response models and standardized indices. British Journal of Mathematical and Statistical Psychology, 38, 67-86.

Halmos, P.R. (1950). Measure theory. Princeton, NJ: Van Nostrand.

Hulin, C.L., Drasgow, F., & Parsons, C.K. (1983). Item response theory: Application to psychological measurement. Homewood, IL: Dow Jones-Irwin.

Levine, M.V. (1985). The trait in latent trait theory. In D.J. Weiss (Ed.), Proceedings of the 1982 Item Response Theory/Computerized Adaptive Testing Conference. Minneapolis: University of Minnesota, Department of Psychology, Computerized Adaptive Testing Laboratory. (a)

Levine, M.V. (1985). Classifying and representing ability distributions. Measurement Series 85-1, Model-Based Measurement Laboratory, 210 Education Building, Department of Educational Psychology, University of Illinois, 1310 S. Sixth Street, Champaign, IL 61820. (b)

Levine, M.V. (1987). Constrained maximum likelihood estimation of ability distributions. Manuscript in preparation.

Levine, M.V., & Drasgow, F. (1982). Appropriateness measurement: Review, critique and validating studies. British Journal of Mathematical and Statistical Psychology, 35, 42-56.

Levine, M.V., & Drasgow, F. (1984). Performance envelopes and optimal appropriateness measurement. Measurement Series 84-5, Model-Based Measurement Laboratory, 210 Education Building, Department of Educational Psychology, University of Illinois, 1310 S. Sixth Street, Champaign, IL 61820.

Levine, M.V. & Drasgow, F. (1987). Optimal appropriateness measurement. Psychometrika, in press.

Levine, M.V., & Rubin, D.F. (1979). Measuring the appropriateness of multiple choice test scores. Journal of Educational Statistics, 4, 269-290.

Levine, M.V., & Williams, B. (1987). Methods for estimating ability densities. Manuscript in preparation.

Lord, F.M. (1965). A strong true-score theory, with applications. Psychometrika, 30, 239-270.

Lord, F.M. (1969). Estimating true-score distributions in psychological testing (An empirical Bayes estimation problem). Psychometrika, 34, 259-299.

Lord, F.M. (1970). Item characteristic curves estimated without knowledge of their mathematical form: A confrontation of Birnbaum's logistic model. Psychometrika, 35, 43-50.

Lord, F.M. (1975). Formula scoring and number-right scoring. Journal of Educational Measurement, 12, 7-11.

Lord, F.M. (1980). Applications of item response theory to practical testing problems. Hillsdale, NJ: Erlbaum.

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

Mislevy, R.J., & Bock, R.D. (1983). Implementation of the EM algorithm in the estimation of item parameters: The BILOG computer program. In D.J. Weiss (Ed.), Proceedings of the 1982 Item Response Theory/Computerized Adaptive Testing Conference. Minneapolis: University of Minnesota, Department of Psychology, Computerized Adaptive Testing Laboratory.

Park, R.K. (1983). Application of a graded response model to the assessment of job satisfaction. Unpublished doctoral dissertation, University of Illinois.

Profile of American Youth: 1980 nationwide administration of the Armed Services Vocational Aptitude Battery (1982). Washington, DC: Office of the Assistant Secretary of Defense (Manpower, Reserve Affairs, and Logistics).

Rudner, L.M. (1983). Individual assessment accuracy. Journal of Educational Measurement, 20, 207-219.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. Psychometrika Monograph, 34.

Sato, T. (1975). The construction and interpretation of S-P tables. Tokyo: Meiji Tosho (in Japanese).

Tatsuoka, K.K. (1984). Caution indices based on item response theory. Psychometrika, 49, 95-110.

Tatsuoka, K.K., & Linn, R.L. (1983). Indices for detecting unusual response patterns: Links between two general approaches and potential applications. Applied Psychological Measurement, 7, 81-96.

Wingersky, M.S., Barton, M.A., & Lord, F.M. (1982). LOGIST user's guide. Princeton, NJ: Educational Testing Service.

Wood, R.L., Wingersky, M.S., & Lord, F.M. (1976). LOGIST - A computer program for estimating examinee ability and item characteristic curve parameters. Research Memorandum 76-6. Princeton, NJ: Educational Testing Service.

Acknowledgments

This research was supported by Contract No. N00014-83K-0397, NR 150 518, from the Office of Naval Research, Michael V. Levine, Principal Investigator and by Contract No. F41689-84-D-0002 from the U.S. Air Force Human Resources Laboratory to the Human Factors and Logistics Division of Universal Energy Systems, Inc., 4401 Dayton-Xenia Road, Dayton, OH 45432. We wish to thank Malcolm Ree, Randy Park, and James Earles for their help.

Table 1

Selected ROC Points for Spuriously High Response Patterns Generated from the 0-9% Ability Range

False Alarm Rate	Proportion Detected by											
	LR_p	LR_3	LR'_3	z_p	z_3	F1	F2	S	T2	T4	JK	O/E
17% Spuriously High Treatment												
.001	04	04	01	00	03	00	01	00	00	01	00	00
.005	11	12	11	03	06	00	08	00	04	04	02	02
.01	16	19	17	05	12	02	13	03	07	06	03	04
.02	28	29	26	08	18	04	20	12	13	11	06	07
.03	34	33	30	11	25	07	24	18	18	14	09	09
.04	38	37	34	13	29	10	28	24	22	18	13	12
.05	43	40	38	15	33	15	32	27	26	22	15	14
.07	48	45	44	19	41	24	40	37	32	26	22	19
.10	52	50	49	26	51	36	50	49	42	33	29	25
33% Spuriously High Treatment												
.001	23	24	17	02	10	00	04	00	06	12	00	00
.005	40	33	27	07	25	00	15	00	28	27	00	04
.01	45	45	43	12	30	01	27	06	37	34	00	09
.02	61	54	52	17	44	05	41	17	50	46	01	17
.03	67	59	58	22	50	16	47	24	60	52	02	24
.04	71	64	63	25	56	23	55	32	65	57	03	37
.05	72	67	66	31	62	30	59	37	69	61	03	37
.07	77	71	70	37	66	42	68	47	74	67	07	47
.10	81	75	75	46	75	57	76	60	81	73	19	57

Table 2

Selected ROC Points for Spuriously High

Response Patterns Generated from the 10-30% Ability Range

False Alarm Rate	Proportion Detected by											
	LR_p	LR_3	LR'_3	z_p	z_3	F1	F2	S	T2	T4	JK	O/E
17% Spuriously High Treatment												
.001	02	01	00	00	02	00	00	00	00	01	00	00
.005	09	07	07	01	05	00	03	00	05	04	00	01
.01	14	14	14	04	09	00	06	00	07	07	00	03
.02	26	25	22	06	14	01	11	04	14	12	01	05
.03	31	29	29	08	19	03	14	06	20	16	02	07
.04	34	33	33	10	23	06	18	10	24	20	02	10
.05	40	36	37	12	27	09	21	12	27	23	03	14
.07	46	43	43	16	34	14	27	18	33	28	06	20
.10	52	50	51	23	43	24	37	28	42	34	14	27
33% Spuriously High Treatment												
.001	16	16	13	00	04	00	00	00	03	09	00	01
.005	31	27	23	03	14	00	07	00	20	23	00	06
.01	37	40	39	05	20	00	15	01	28	29	00	10
.02	53	50	50	08	30	03	27	06	42	41	00	20
.03	61	56	57	12	37	08	34	10	51	47	00	27
.04	65	63	62	14	42	12	42	16	58	53	00	34
.05	68	66	65	19	49	17	46	20	62	58	00	40
.07	73	70	70	25	54	28	56	29	67	63	05	51
.10	78	74	75	33	64	44	67	41	74	70	18	60

Table 3

Selected ROC Points for Spuriously High

Response Patterns Generated from the 31-48% Ability Range

False Alarm Rate	LR_p	LR_3	LR_3'	z_p	z_3	F1	F2	S	T2	T4	JK	O/E
17% Spuriously High Treatment												
.001	00	00	00	00	01	00	00	00	00	01	00	00
.005	03	03	04	00	03	00	01	00	04	04	00	00
.01	06	07	08	02	06	00	02	00	06	06	00	01
.02	15	15	14	03	09	00	05	00	12	12	00	05
.03	20	19	19	05	14	03	07	02	17	15	00	08
.04	24	23	24	06	17	06	10	03	21	18	00	10
.05	29	26	28	07	20	07	13	04	23	22	00	13
.07	36	34	35	10	25	12	18	07	29	26	01	20
.10	43	42	43	15	33	18	26	12	36	32	07	29
33% Spuriously High Treatment												
.001	06	10	07	00	02	00	00	00	02	06	00	01
.005	17	16	14	01	07	00	03	00	12	16	00	05
.01	22	27	26	02	10	00	08	00	18	22	00	08
.02	39	36	37	04	17	04	16	02	27	32	00	17
.03	48	43	45	05	22	08	21	05	36	38	00	23
.04	53	51	49	07	27	12	27	07	41	43	00	29
.05	56	55	54	09	33	16	31	09	45	47	00	34
.07	63	61	61	13	37	23	40	14	50	53	07	44
.10	71	67	68	20	46	36	51	22	59	60	19	53

The column headings "Proportion Detected by" span the LR through O/E columns.

Table 4

Selected ROC Points for Spuriously High Response Patterns Generated from the 49-64% Ability Range

False Alarm Rate	LR_p	LR_3	LR_3'	z_p	z_3	F1	F2	S	T2	T4	JK	O/E
17% Spuriously High Treatment												
.001	00	00	00	00	00	00	00	00	00	00	00	00
.005	00	00	00	00	01	00	00	00	02	03	00	00
.01	02	01	01	00	03	01	01	00	04	04	00	00
.02	07	06	03	01	05	01	03	00	07	08	00	00
.03	11	09	05	01	08	04	04	00	11	11	00	06
.04	14	13	07	02	10	06	07	01	14	14	00	09
.05	18	16	09	03	13	08	08	01	16	17	00	12
.07	25	23	14	06	17	11	13	03	20	21	01	17
.10	33	30	23	09	23	16	19	05	26	27	07	24
33% Spuriously High Treatment												
.001	01	02	00	00	00	00	00	00	00	02	00	00
.005	05	04	02	03	03	00	01	00	05	07	00	01
.01	08	10	05	00	04	02	04	00	07	10	00	02
.02	19	16	09	01	07	07	08	01	12	17	00	06
.03	28	23	14	03	10	11	11	02	16	20	00	08
.04	34	32	18	03	12	14	15	03	20	25	00	11
.05	37	37	21	05	16	17	17	04	23	29	00	14
.07	48	45	29	08	19	23	23	07	28	35	03	20
.10	60	55	41	13	25	31	31	12	35	40	11	28

Table 5

Selected ROC Points for Spuriously Low

Response Patterns Generated from the 31-48% Ability Range

False Alarm Rate	Proportion Detected by											
	LR_p	LR_3	LR_3'	z_p	z_3	F1	F2	S	T2	T4	JK	O/E
17% Spuriously Low Treatment												
.001	01	00	00	00	00	00	00	00	00	00	00	00
.005	05	01	01	03	02	00	01	00	02	02	00	00
.01	09	03	03	05	04	01	02	00	03	03	01	01
.02	15	06	07	08	07	02	04	00	06	07	01	02
.03	18	10	12	12	10	04	05	01	09	09	02	03
.04	21	14	15	14	13	07	07	03	12	12	03	05
.05	24	17	18	15	15	10	09	04	14	14	05	07
.07	29	22	23	21	19	17	12	07	18	17	07	10
.10	35	28	28	27	26	25	17	11	23	22	12	14
33% Spuriously Low Treatment												
.001	07	01	01	01	02	00	00	00	00	01	00	00
.005	14	03	04	07	05	00	04	00	03	04	01	01
.01	22	08	09	12	10	02	07	00	05	07	02	01
.02	30	14	16	18	15	05	11	03	09	11	04	03
.03	36	20	22	23	20	09	13	06	14	15	07	04
.04	41	24	26	27	23	13	17	10	16	19	10	06
.05	45	29	30	31	26	17	19	11	19	22	13	07
.07	51	36	37	36	32	27	24	17	22	27	17	11
.10	59	44	44	44	38	36	31	25	29	33	24	16

Table 6

Selected ROC Points for Spuriously Low

Response Patterns Generated from the 49-64% Ability Range

False Alarm Rate	Proportion Detected by											
	LR_p	LR_3	LR_3^*	z_p	z_3	F1	F2	S	T2	T4	JK	O/E
17% Spuriously Low Treatment												
.001	13	02	02	00	02	00	00	00	00	01	00	00
.005	22	09	08	04	04	00	01	00	05	05	00	00
.01	26	14	13	07	09	03	03	00	08	07	00	01
.02	32	21	20	11	14	09	07	01	13	13	00	04
.03	34	26	25	16	19	17	10	03	19	16	00	06
.04	38	30	28	19	22	21	13	04	22	20	00	08
.05	41	33	31	23	25	24	15	05	26	22	00	11
.07	46	37	35	29	31	31	20	08	29	27	03	16
.10	51	42	40	37	38	34	28	13	36	32	09	21
33% Spuriously Low Treatment												
.001	24	09	08	02	08	00	00	00	03	05	00	00
.005	34	16	15	15	14	00	06	00	14	12	00	02
.01	43	25	24	23	22	03	11	01	18	17	00	03
.02	51	33	32	31	29	12	17	04	26	25	00	08
.03	55	39	38	37	36	22	22	07	33	29	01	12
.04	58	43	41	41	39	29	26	10	38	34	01	16
.05	61	46	45	45	43	35	29	13	41	38	02	19
.07	66	51	50	52	49	44	36	19	45	44	06	25
.10	71	58	56	60	57	52	45	27	53	51	14	33

Table 7

Selected ROC Points for Spuriously Low

Response Patterns Generated from the 65-92% Ability Range

False Alarm Rate	Proportion Detected by											
	LR_p	LR_3	LR_3'	z_p	z_3	F1	F2	S	T2	T4	JK	J/E
17% Spuriously Low Treatment												
.001	35	15	14	01	05	01	00	00	01	03	00	00
.005	45	33	31	09	11	11	05	01	11	09	00	00
.01	51	41	38	15	19	24	11	05	15	14	00	00
.02	57	46	45	20	27	40	20	13	25	21	00	02
.03	60	50	49	25	34	45	24	13	32	26	00	03
.04	63	53	51	31	38	53	30	24	36	30	00	04
.05	65	55	53	35	42	57	34	27	39	35	00	06
.07	68	59	57	42	50	61	41	33	43	40	04	10
.10	71	63	61	51	57	65	50	41	53	48	12	16
33% Spuriously Low Treatment												
.001	53	31	29	04	26	00	03	00	15	22	00	02
.005	61	44	41	25	39	05	21	02	40	39	00	09
.01	67	53	51	34	52	20	34	09	47	46	00	15
.02	72	59	57	45	61	42	47	21	59	57	00	25
.03	75	63	61	52	67	54	54	29	67	61	00	30
.04	77	66	63	57	70	60	61	36	71	67	00	38
.05	78	68	66	61	74	64	64	40	74	70	00	43
.07	81	72	70	69	79	71	71	49	79	75	15	52
.10	84	75	74	77	84	79	79	58	84	80	33	61

Table 8

Selected ROC Points for Spuriously Low

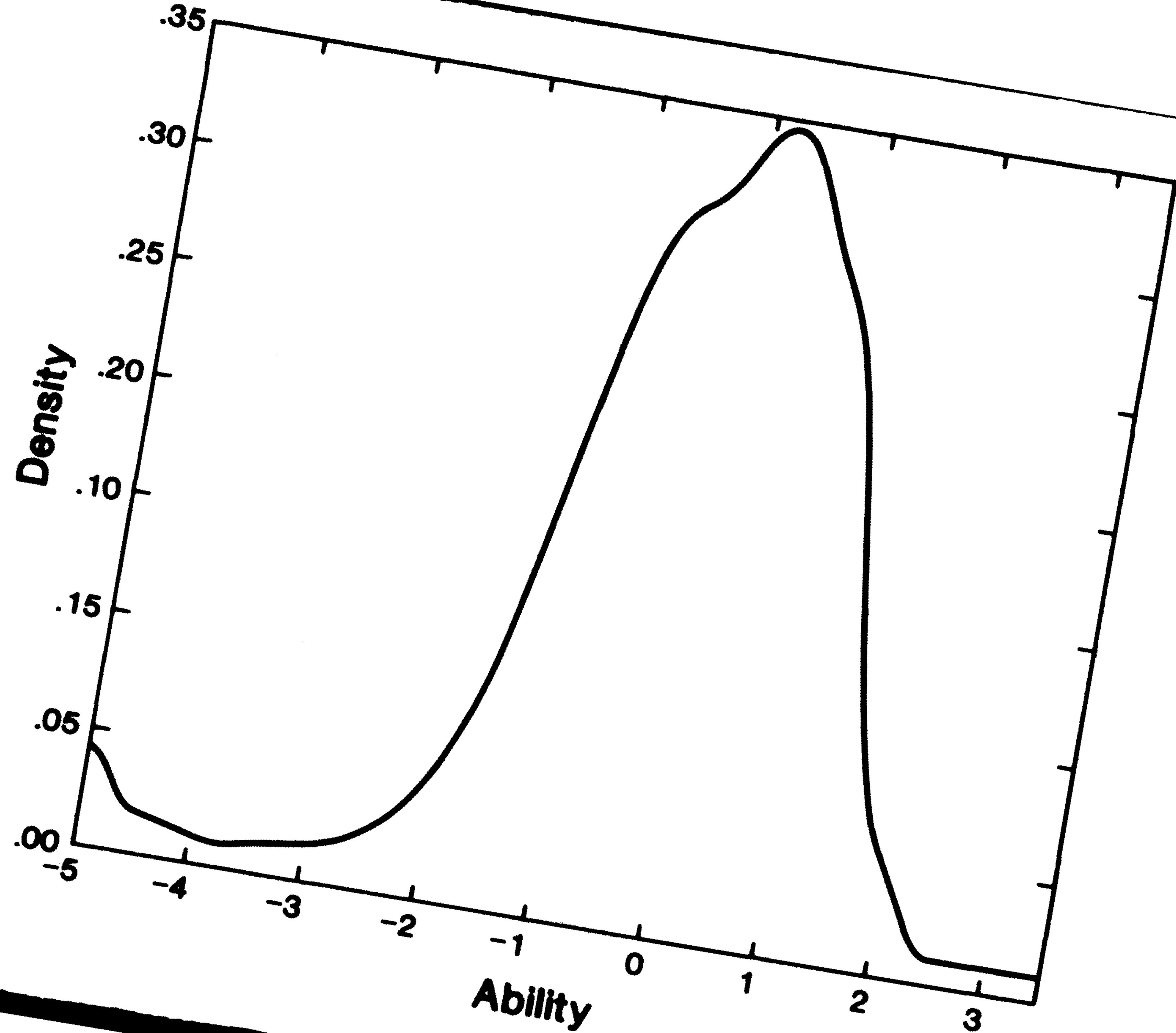
Response Patterns Generated from the 43-100% Ability Range

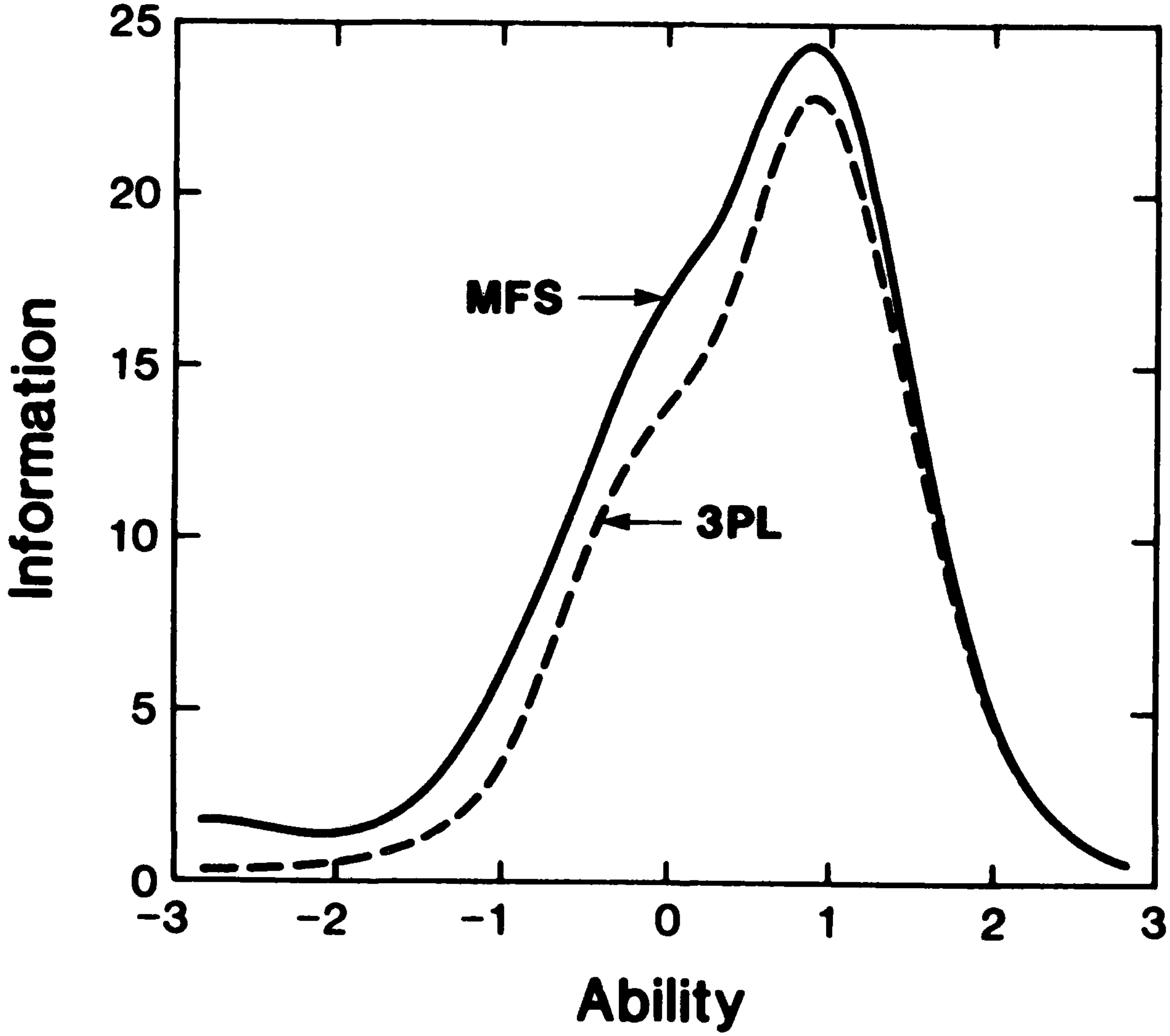
False Alarm Rate	Proportion Detected by											
	LR_p	LR_1	LR_1'	z_p	z_1	F1	F2	C	T2	T4	JK	E
17% Spuriously Low Treatment												
.001	45	22	22	04	04	11	01	00	02	03	00	00
.005	55	42	40	13	11	27	09	09	15	11	00	00
.01	60	49	46	18	20	43	18	22	21	18	00	00
.02	67	54	53	26	29	55	30	35	33	29	00	00
.03	69	58	56	32	37	60	35	41	41	35	00	00
.04	71	60	58	37	41	63	41	48	47	41	00	01
.05	72	62	60	40	46	66	45	51	51	47	00	01
.07	74	65	62	48	54	71	53	58	56	53	02	03
.10	77	68	66	58	63	75	63	65	64	62	11	06
33% Spuriously Low Treatment												
.001	64	42	40	04	32	02	06	00	20	33	00	02
.005	72	53	51	27	49	17	32	08	51	52	00	08
.01	76	61	59	39	62	36	46	21	59	60	00	13
.02	81	67	64	51	71	59	61	39	69	70	00	22
.03	83	70	68	59	77	69	67	48	74	74	00	27
.04	85	72	70	64	79	74	72	55	80	78	00	33
.05	86	74	73	68	82	78	75	59	83	81	00	38
.07	87	77	75	75	86	84	80	68	86	84	21	48
.10	90	79	77	82	90	87	86	76	90	87	41	57

Figure Captions

1. Ability Density for National Opinion Research Center sample and Arithmetic Reasoning test.

2. Information functions for dichotomous and polychotomous scorings of the Arithmetic Reasoning test.


Density
.35
.30
.25
.20
.10
.15
.05
.00
-5
-4
-3
-2
-1
0
1
2
3
Ability

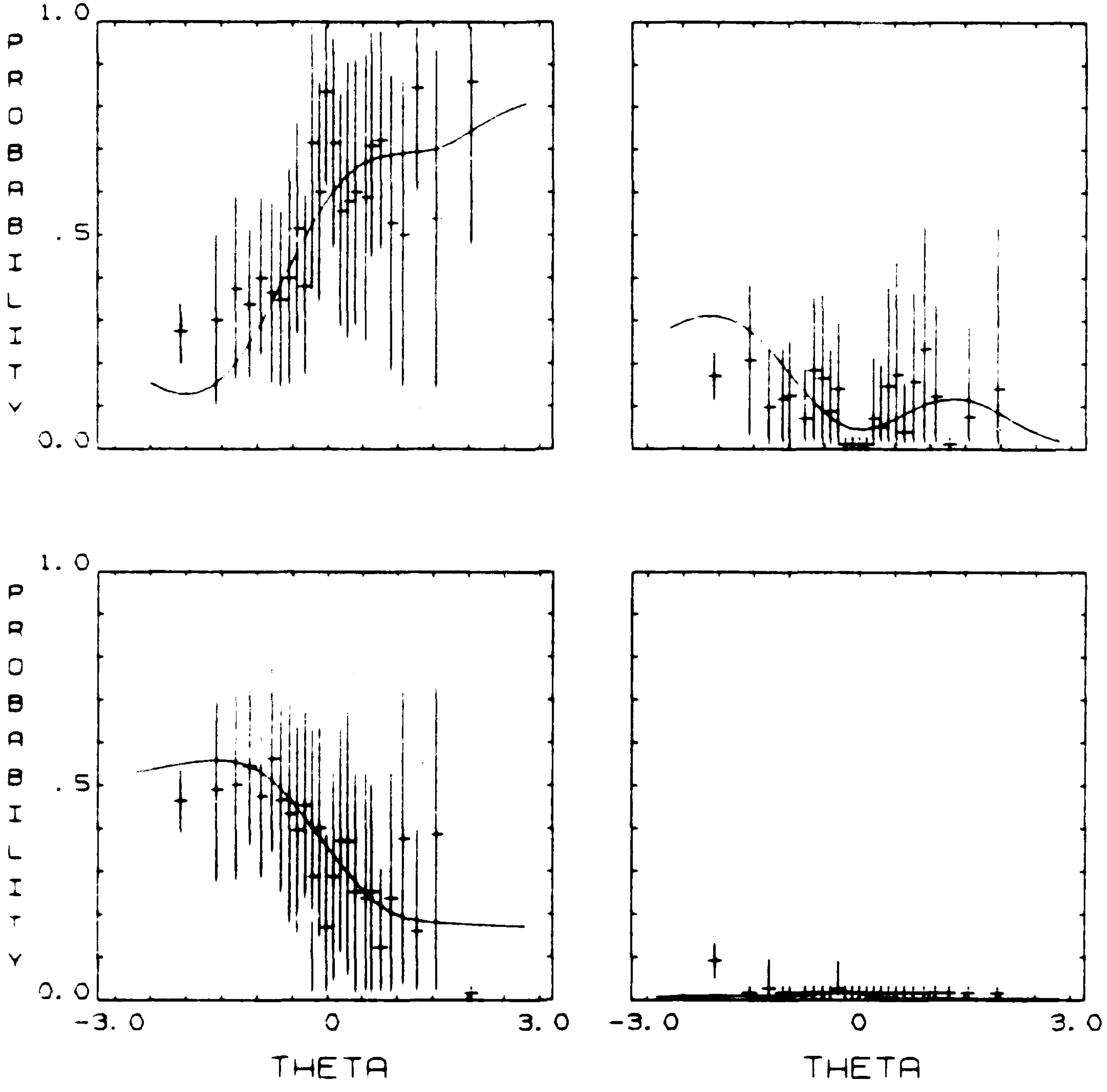


Information
Ability
MFS
3PL
0
5
10
15
20
25
−3
−2
−1
0
1
2
3

Appendix 1

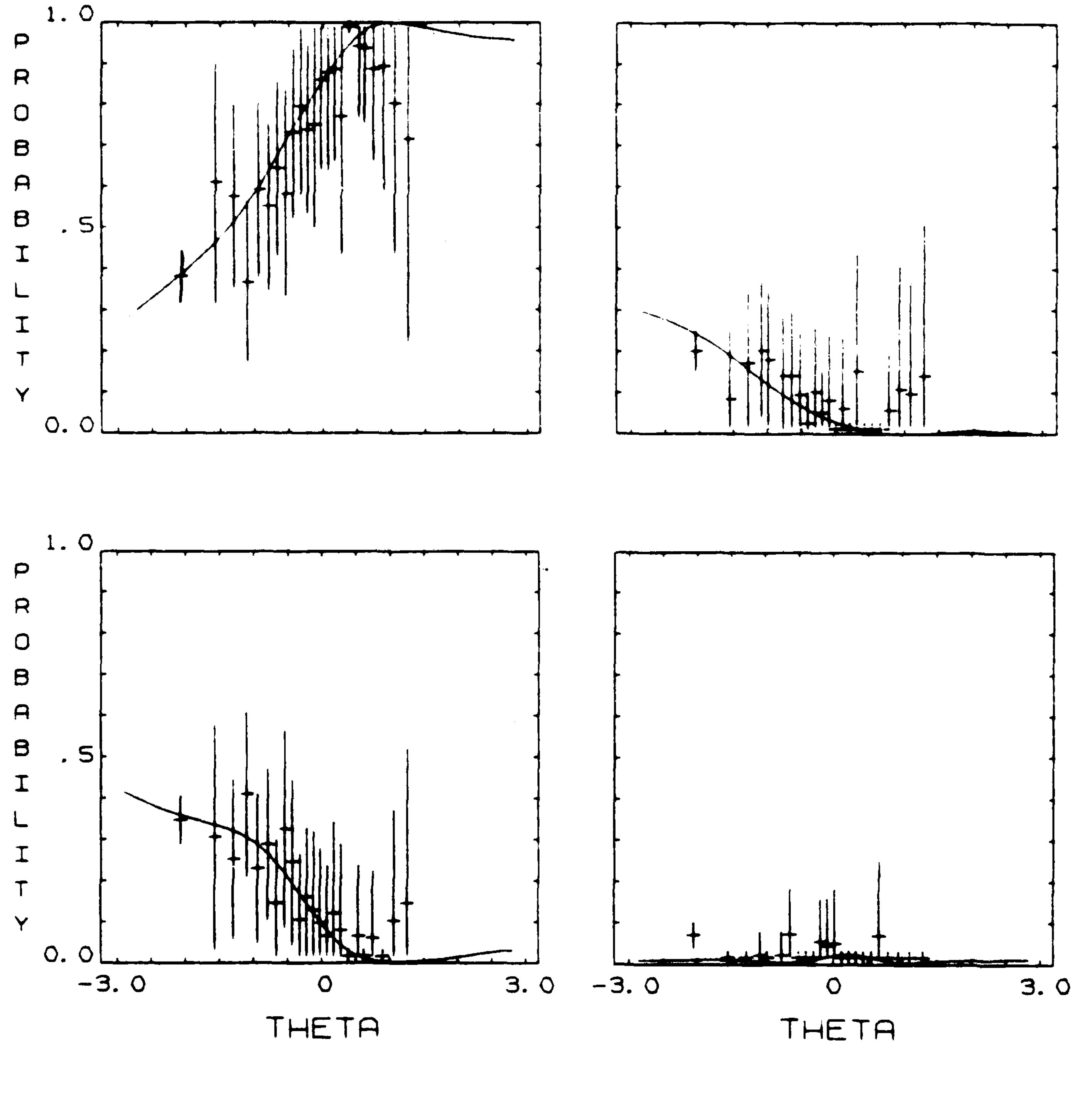
Goodness of Fit of Arithmetic Reasoning COCCs Estimated from a Sample of N=2,978 and Evaluated Using the Entire Sample of N=11,914.

ITEM 1

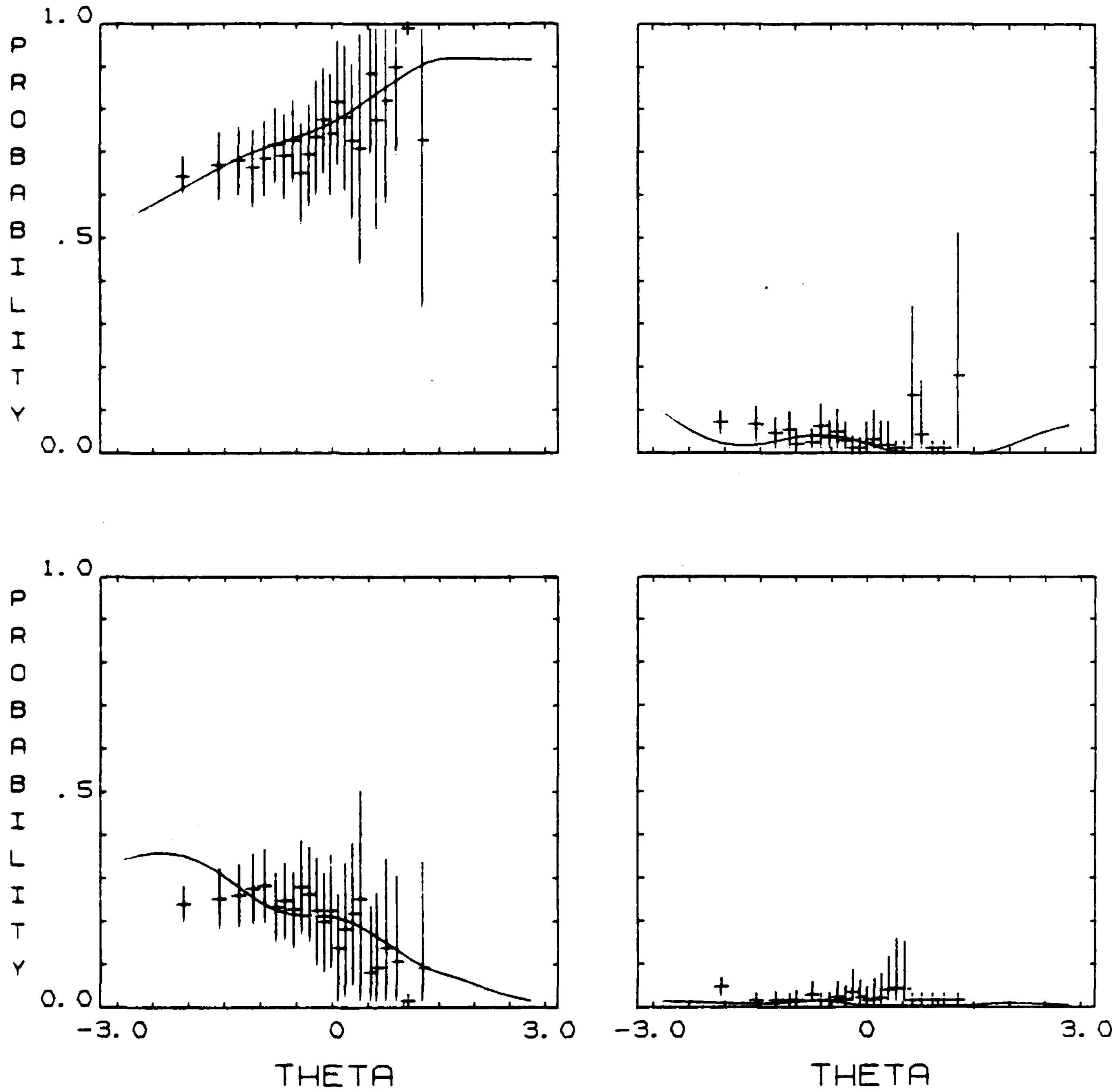


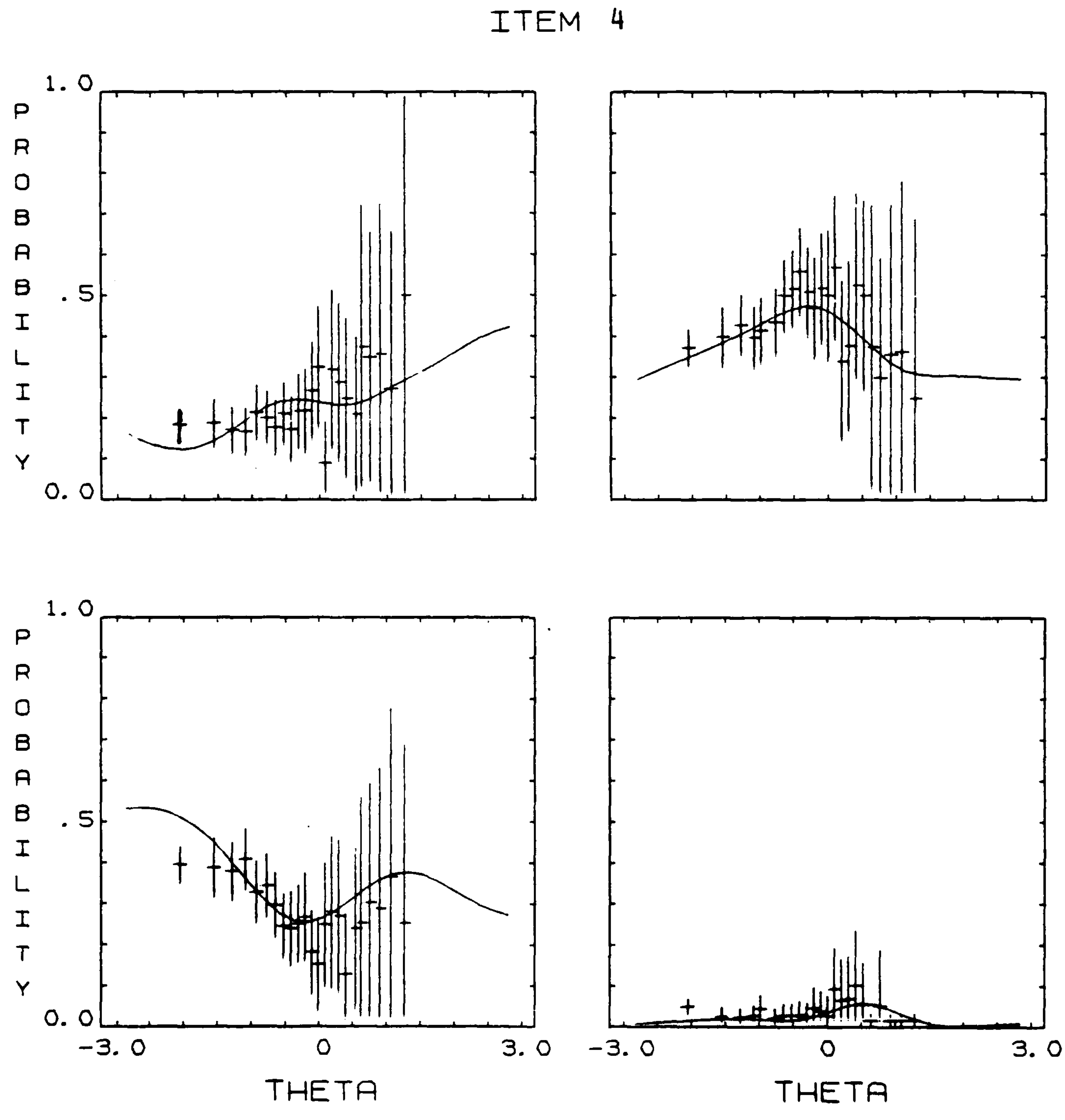
1.0
PROBABILITY
.5
0.0
1.0
PROBABILITY
.5
0.0
-3.0
0
3.0
-3.0
0
3.0
THETA
THETA

ITEM 2



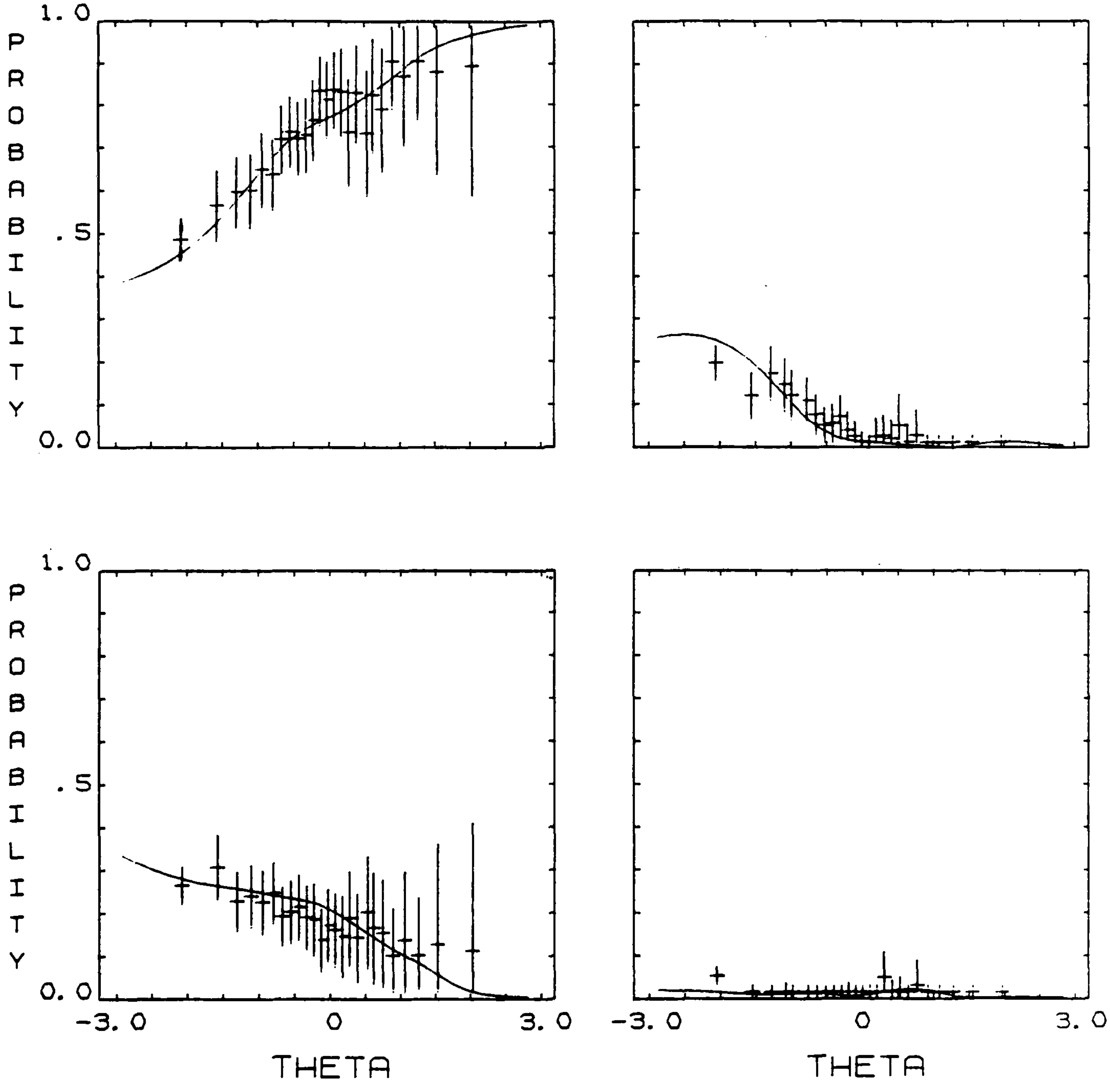
ITEM 3



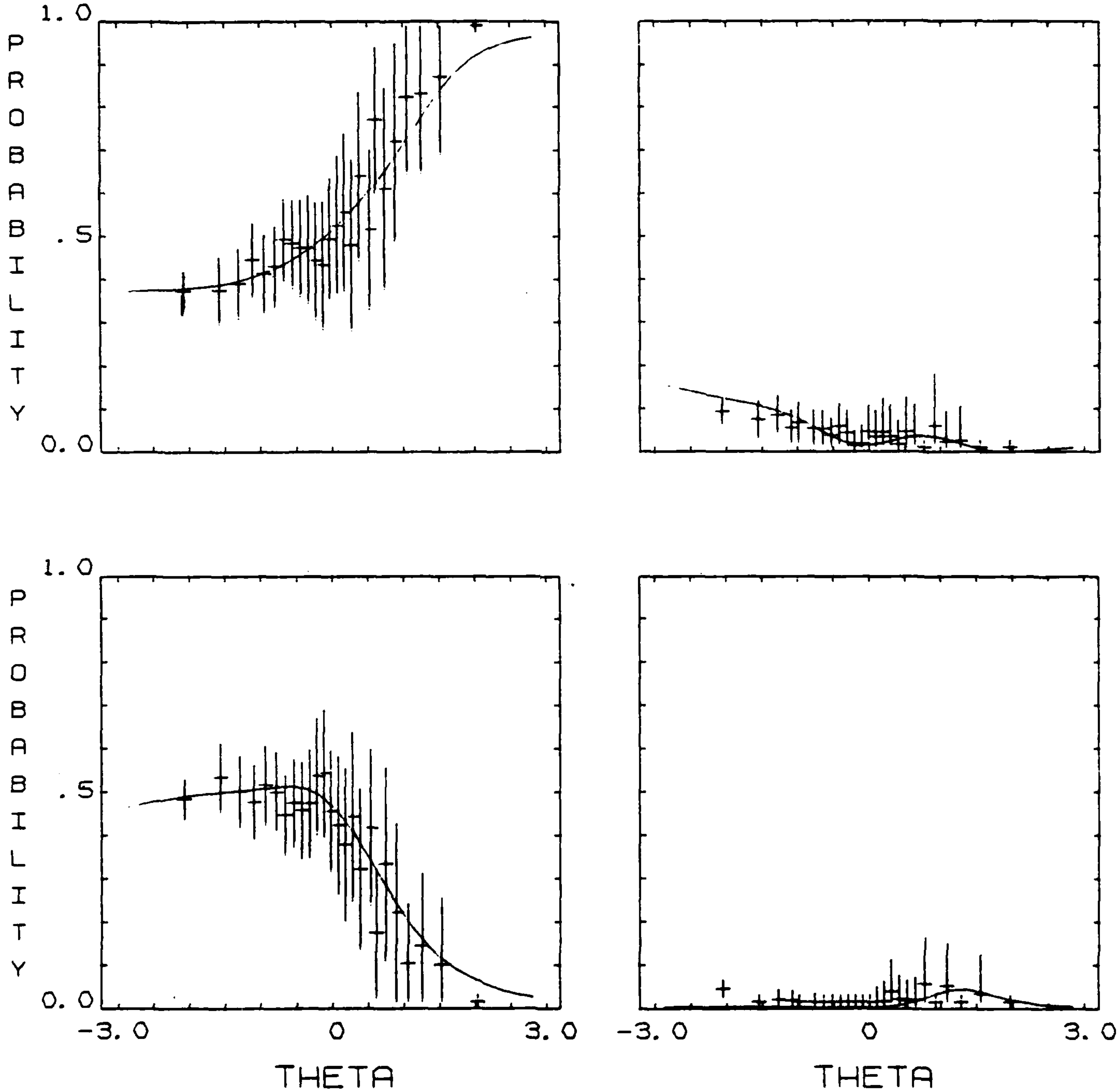


ITEM 4
1.0
.5
0.0
PROBABILITY
-3.0
0
3.0
THETA

ITEM 5

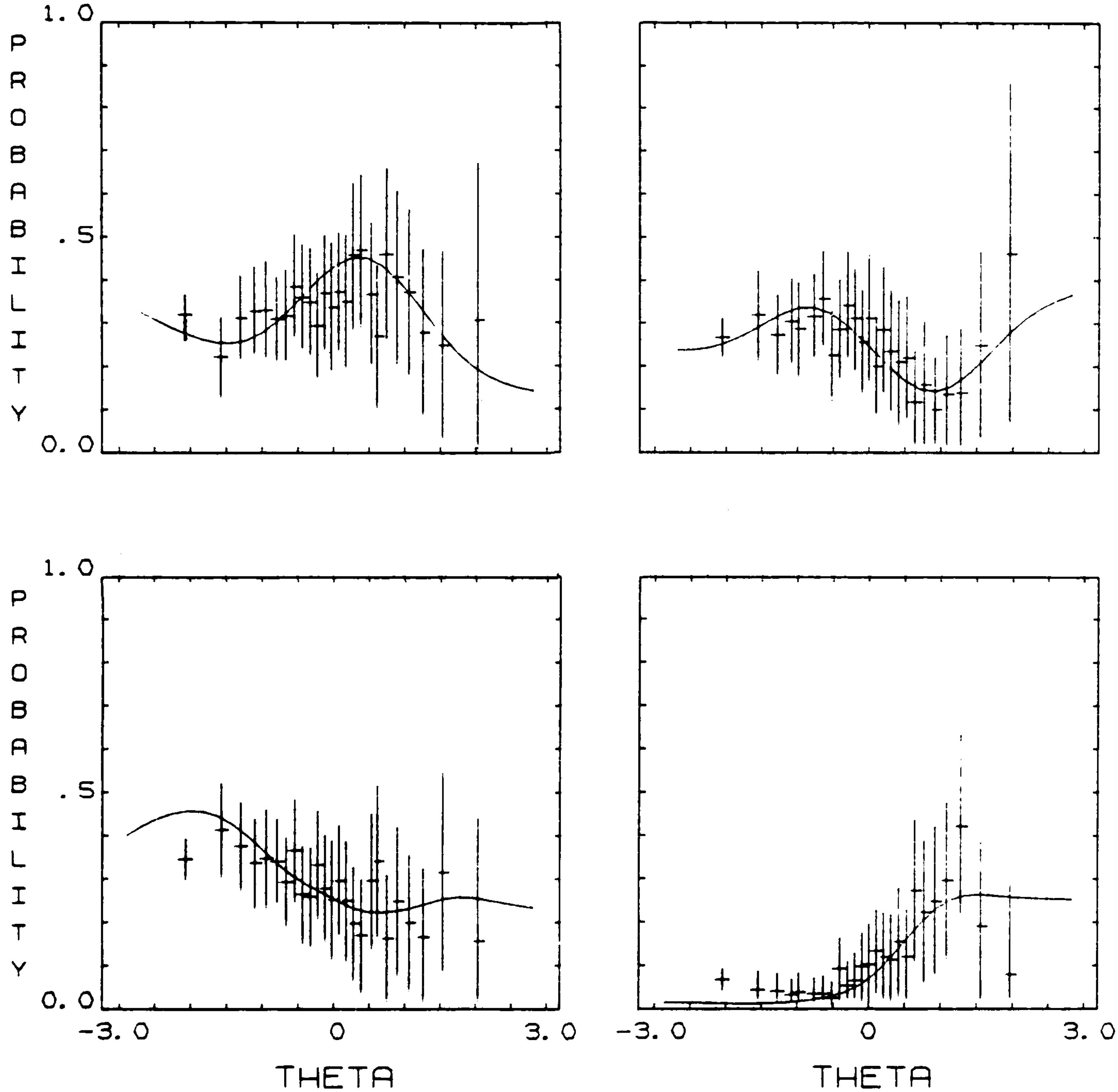


ITEM 6

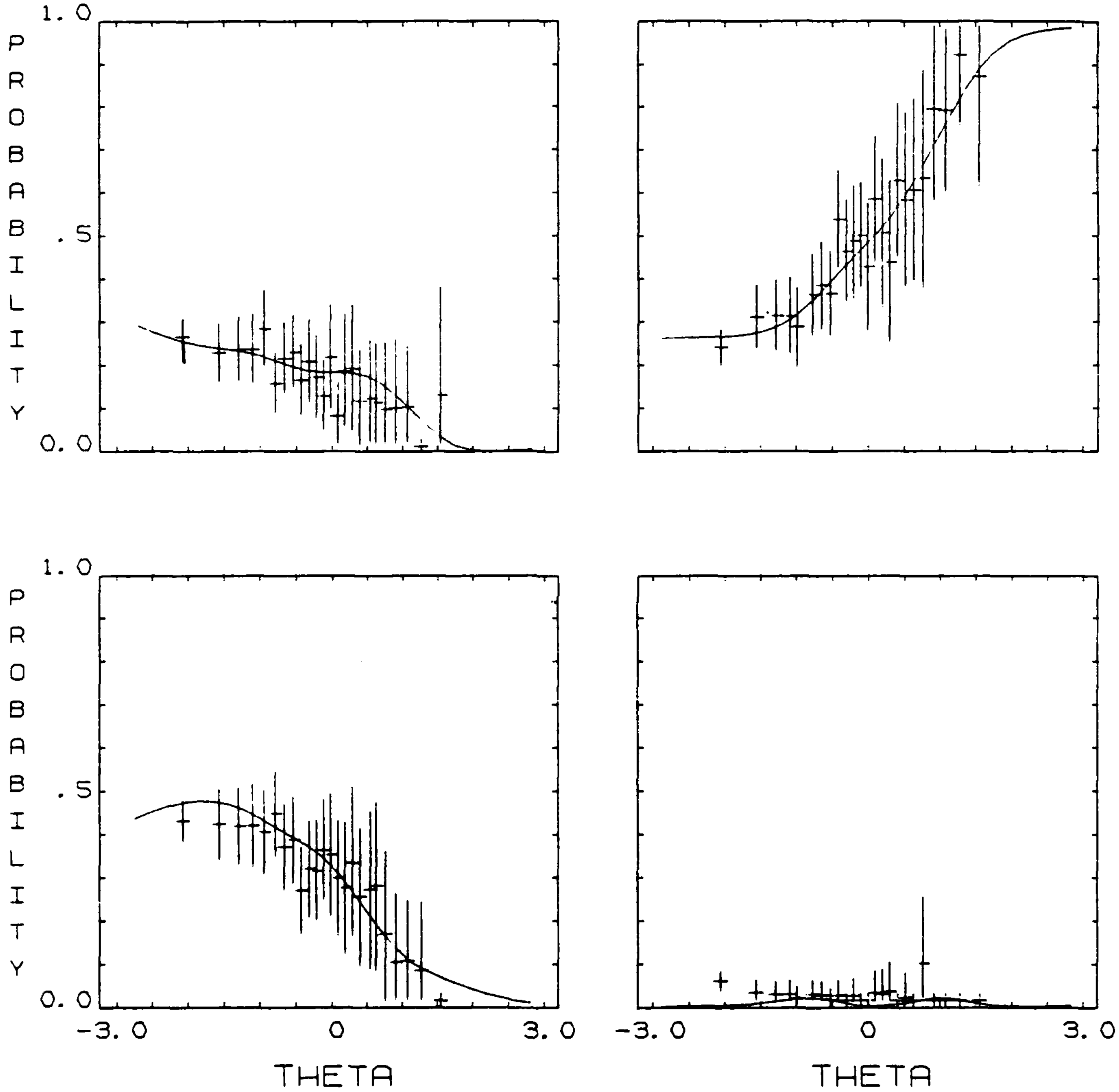


1.0
.5
0.0
PROBABILITY
1.0
.5
0.0
PROBABILITY
-3.0
0
3.0
-3.0
0
3.0
THETA
THETA

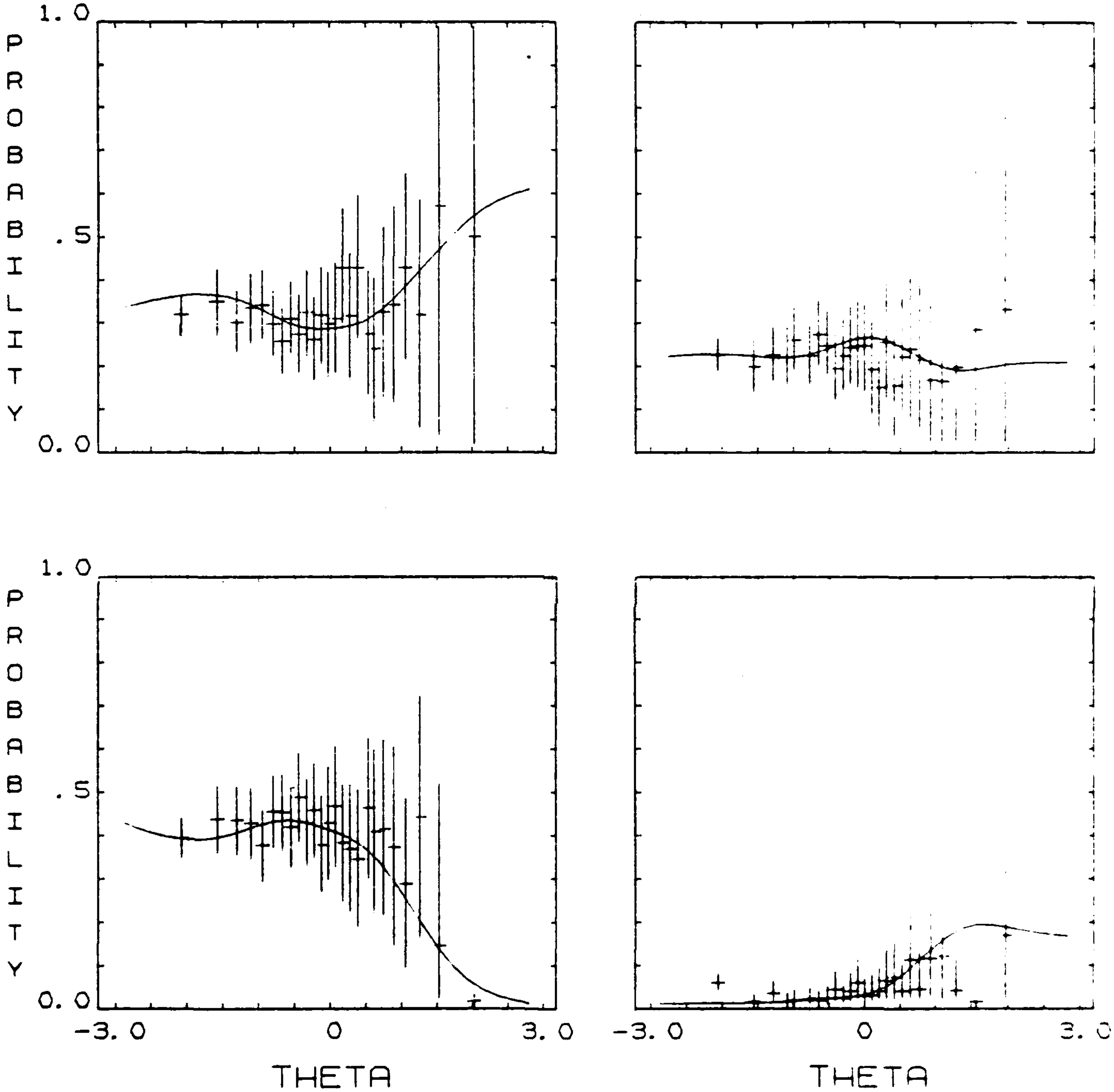
ITEM 7



ITEM 8

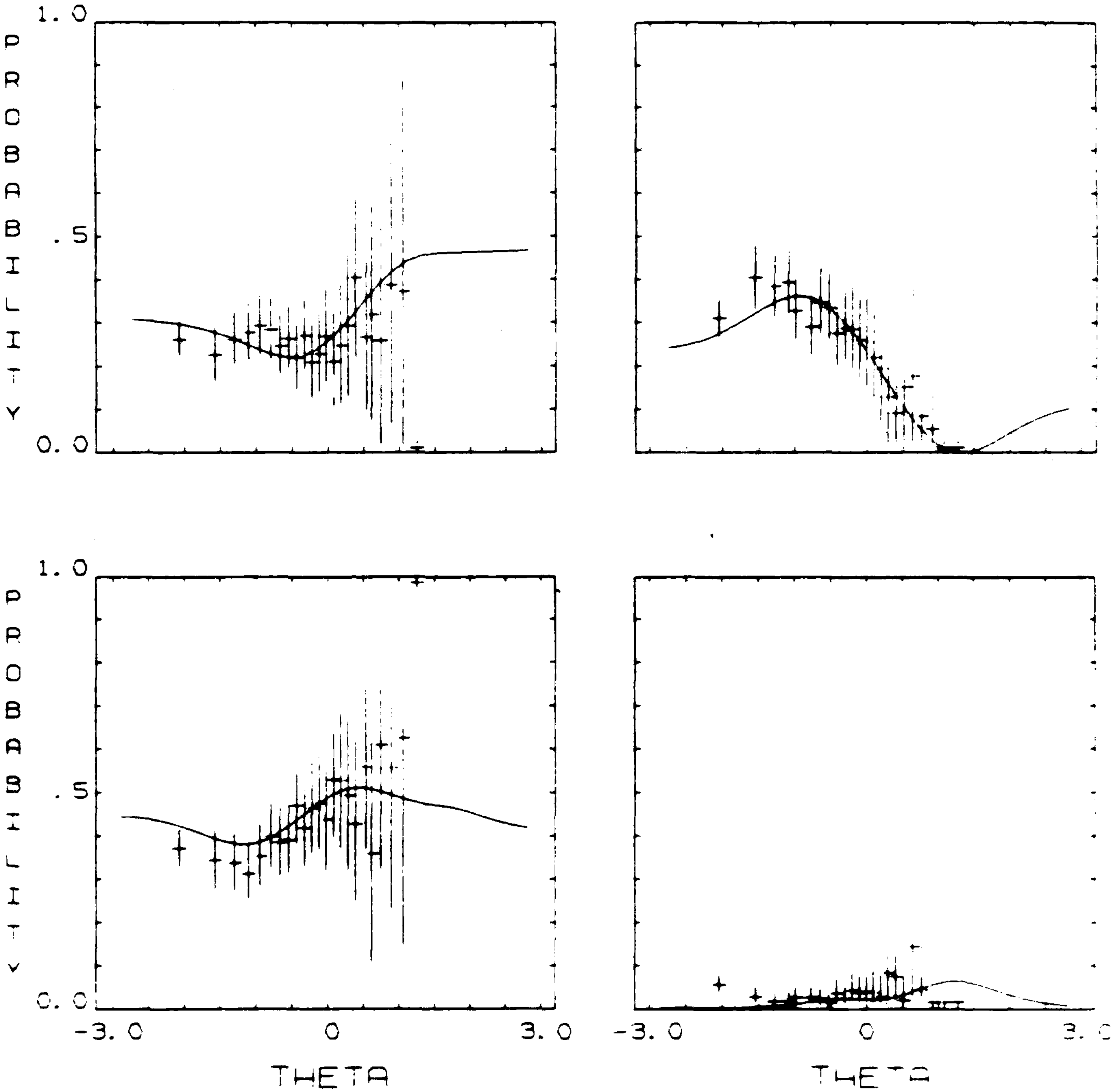


ITEM 9

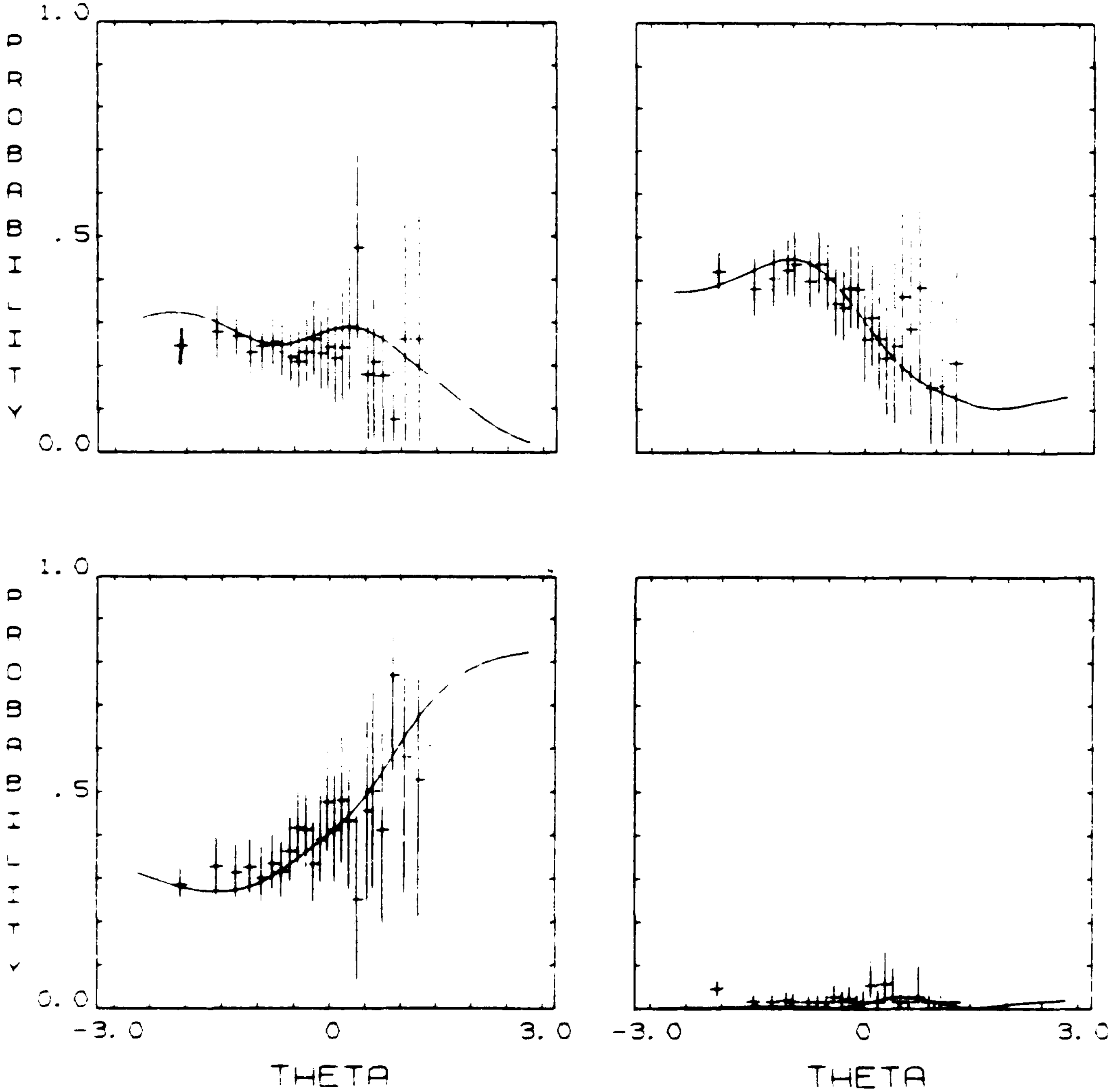


1.0
PROBABILITY
.5
0.0
1.0
PROBABILITY
.5
0.0
-3.0
0
3.0
THETA
-3.0
0
3.0
THETA

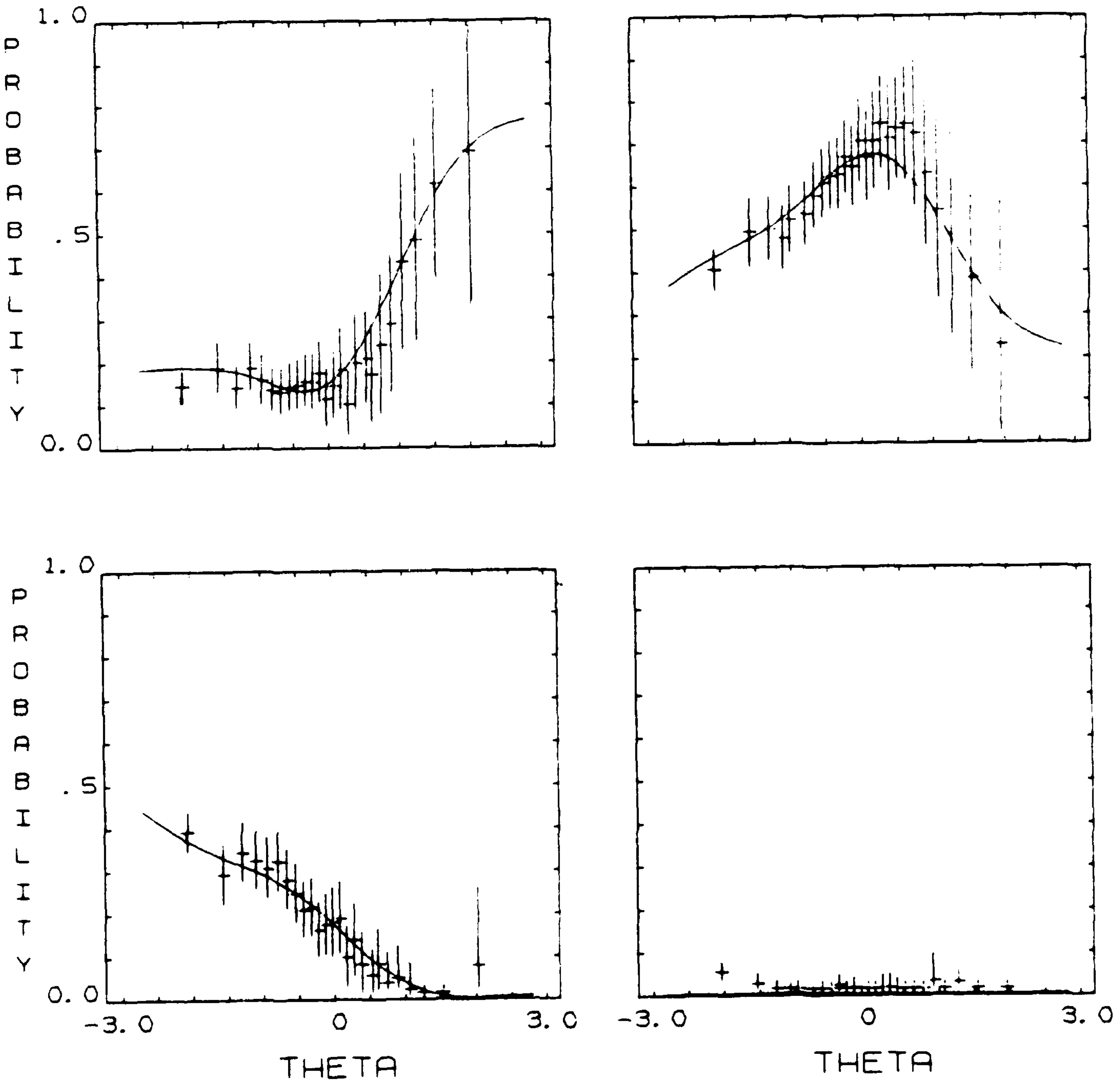
ITEM 10



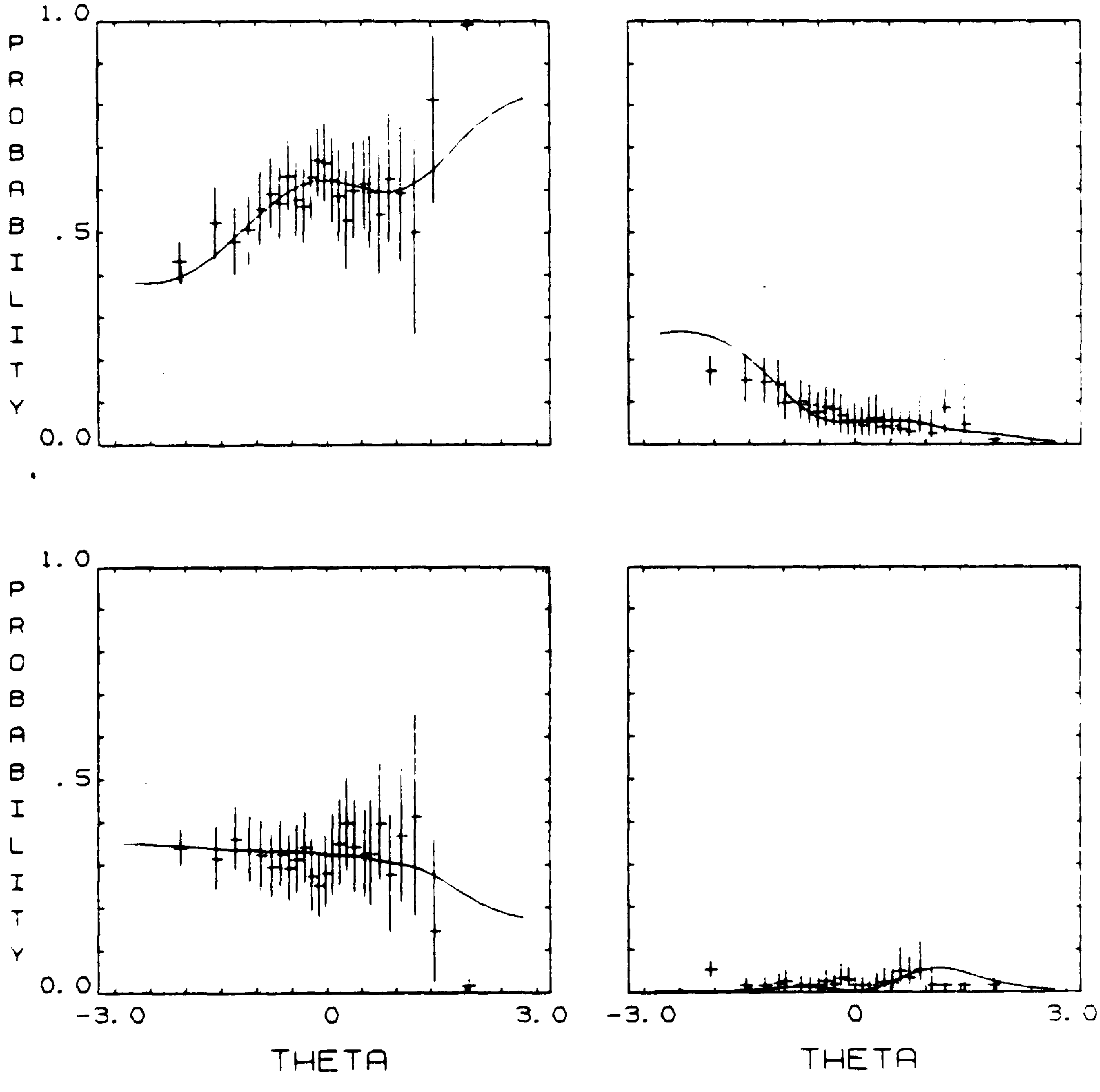
ITEM 11



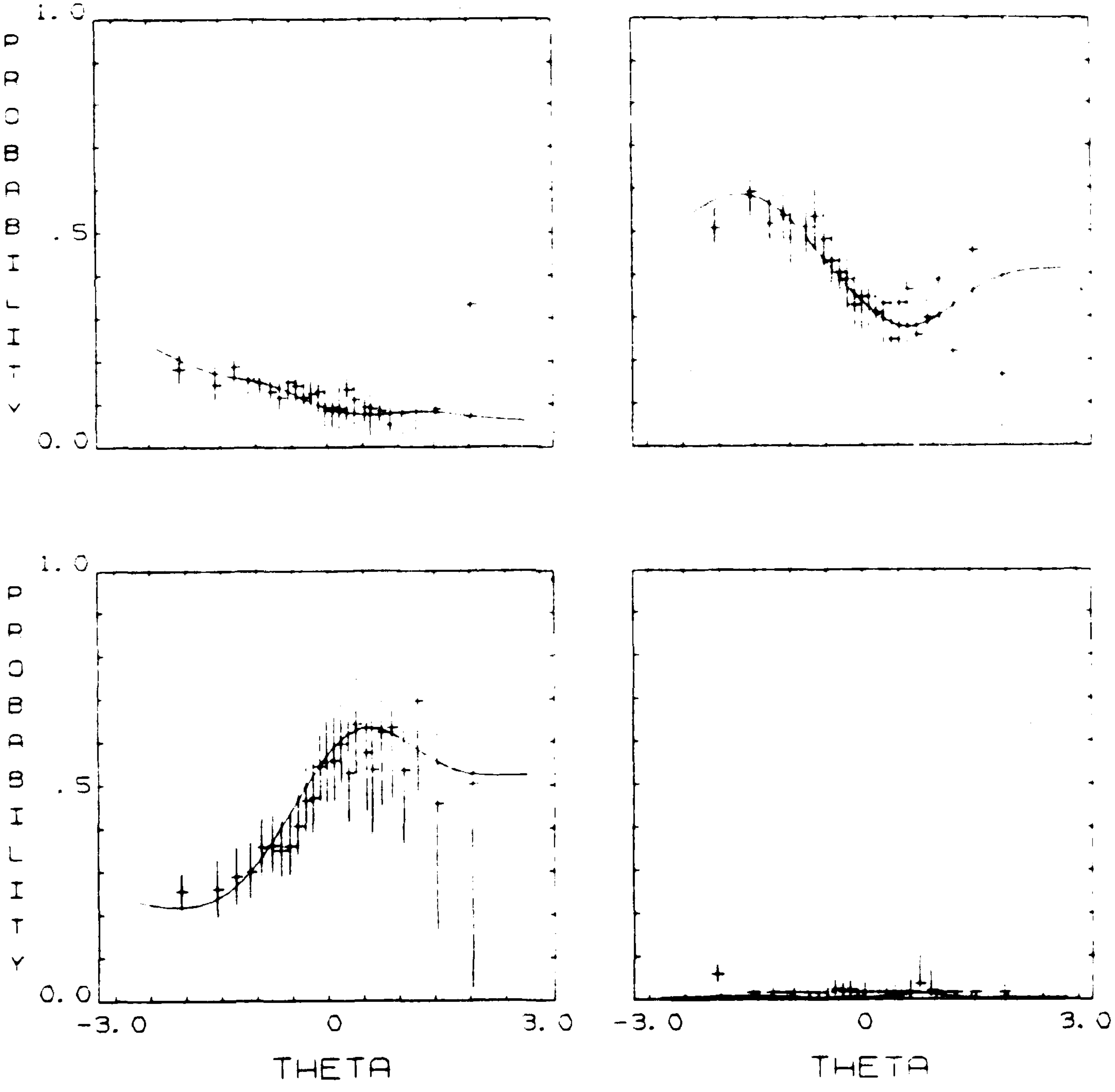
ITEM 12



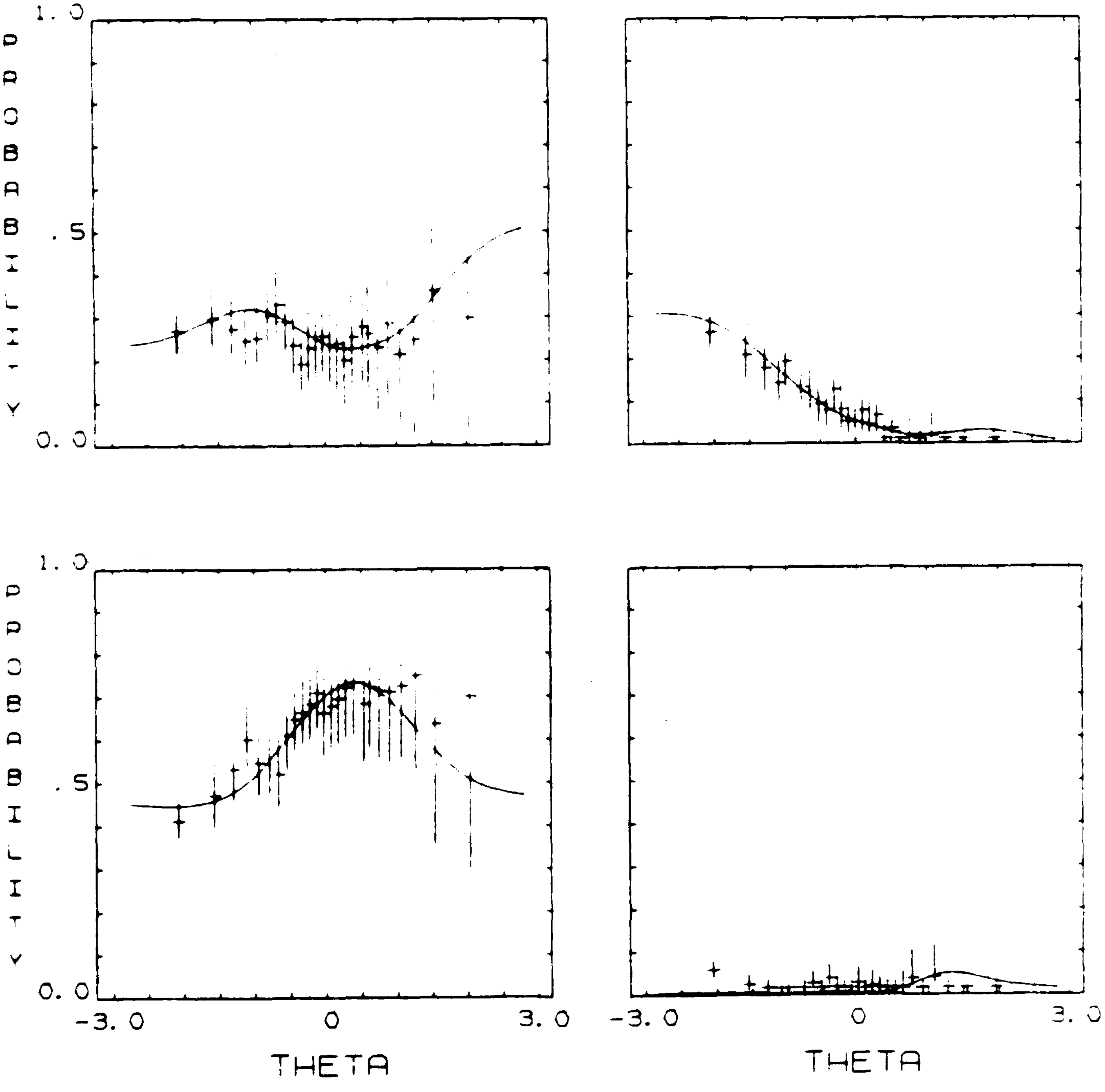
ITEM 13



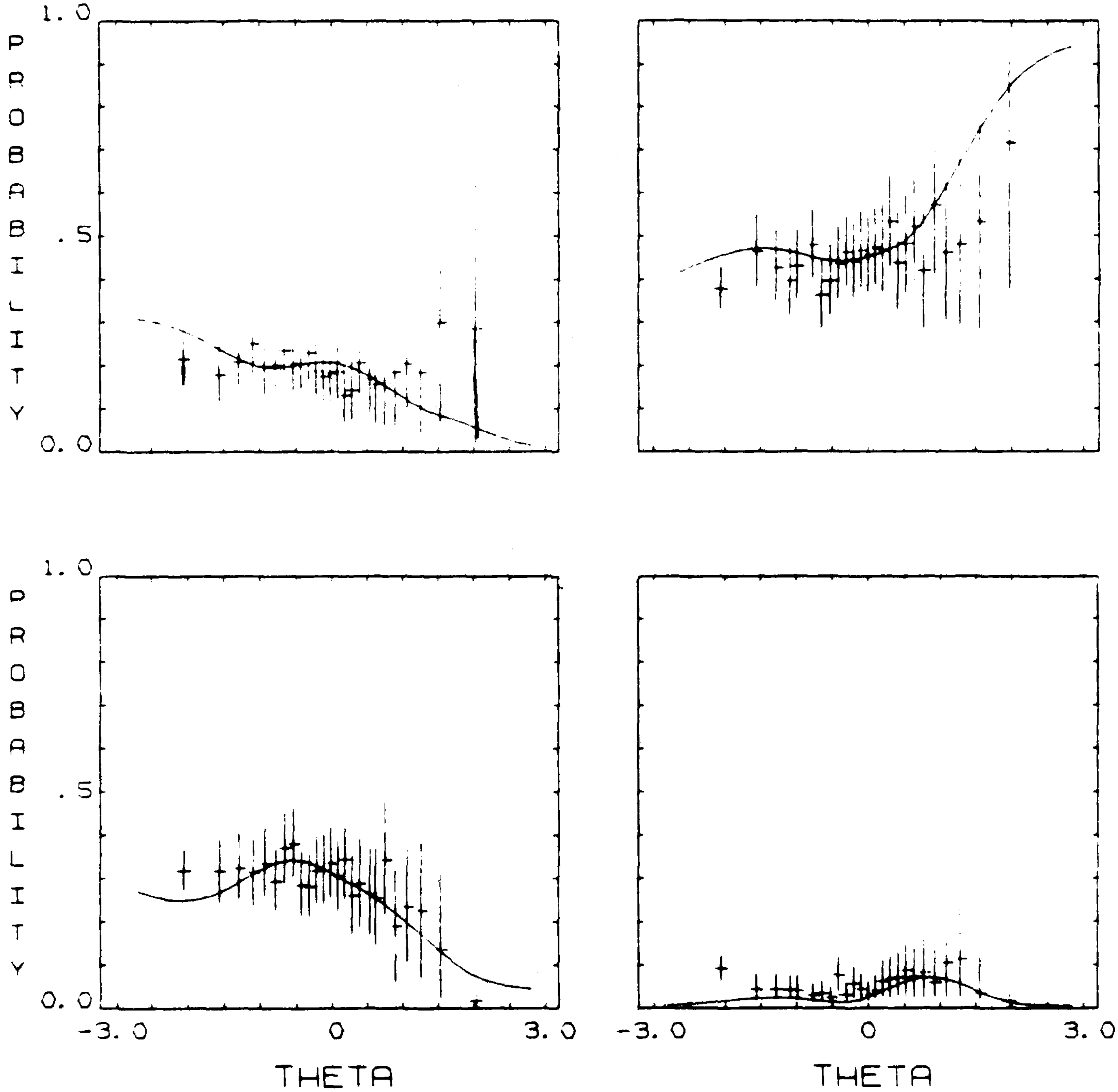
ITEM 14



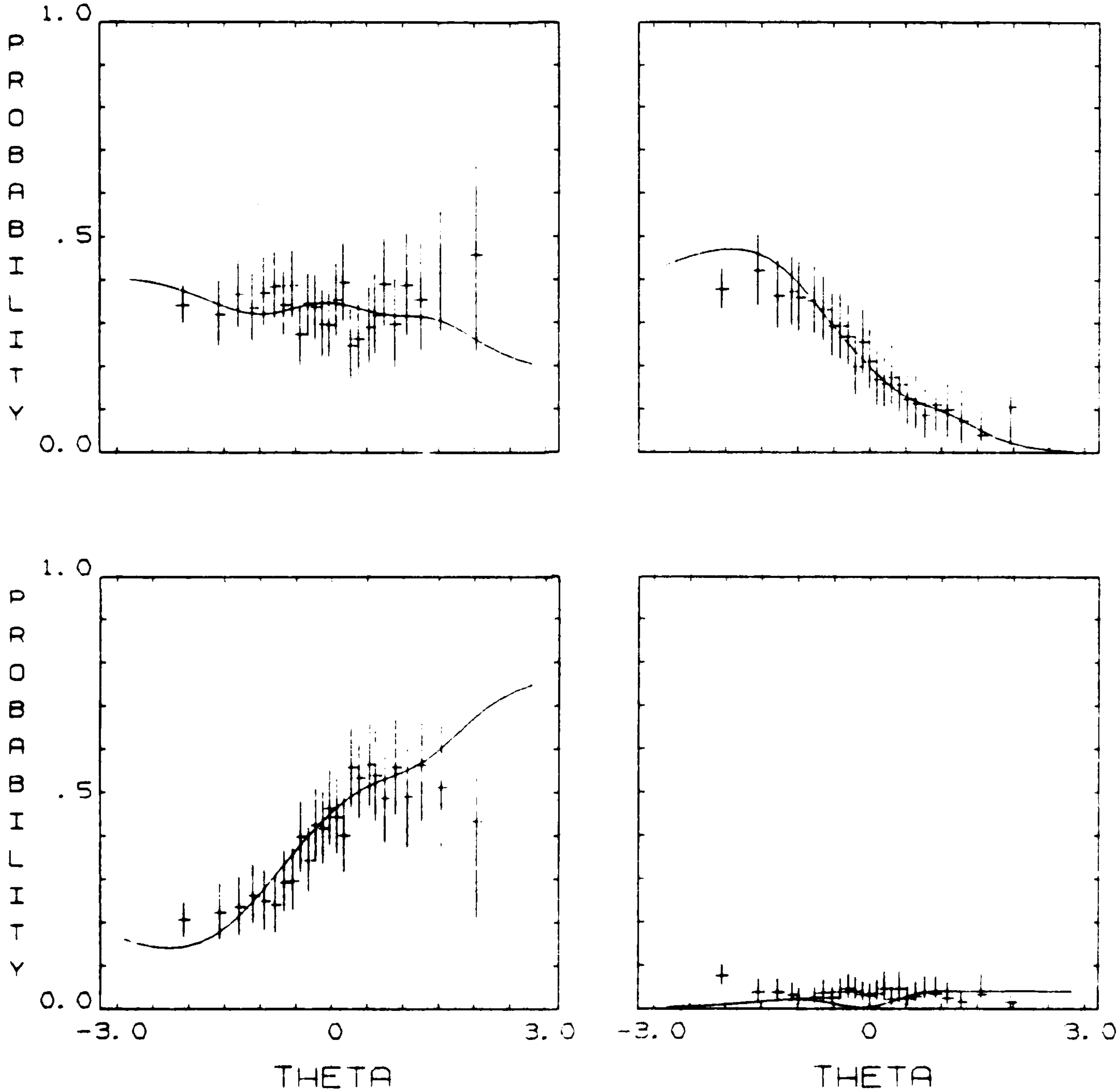
ITEM 15



ITEM 16

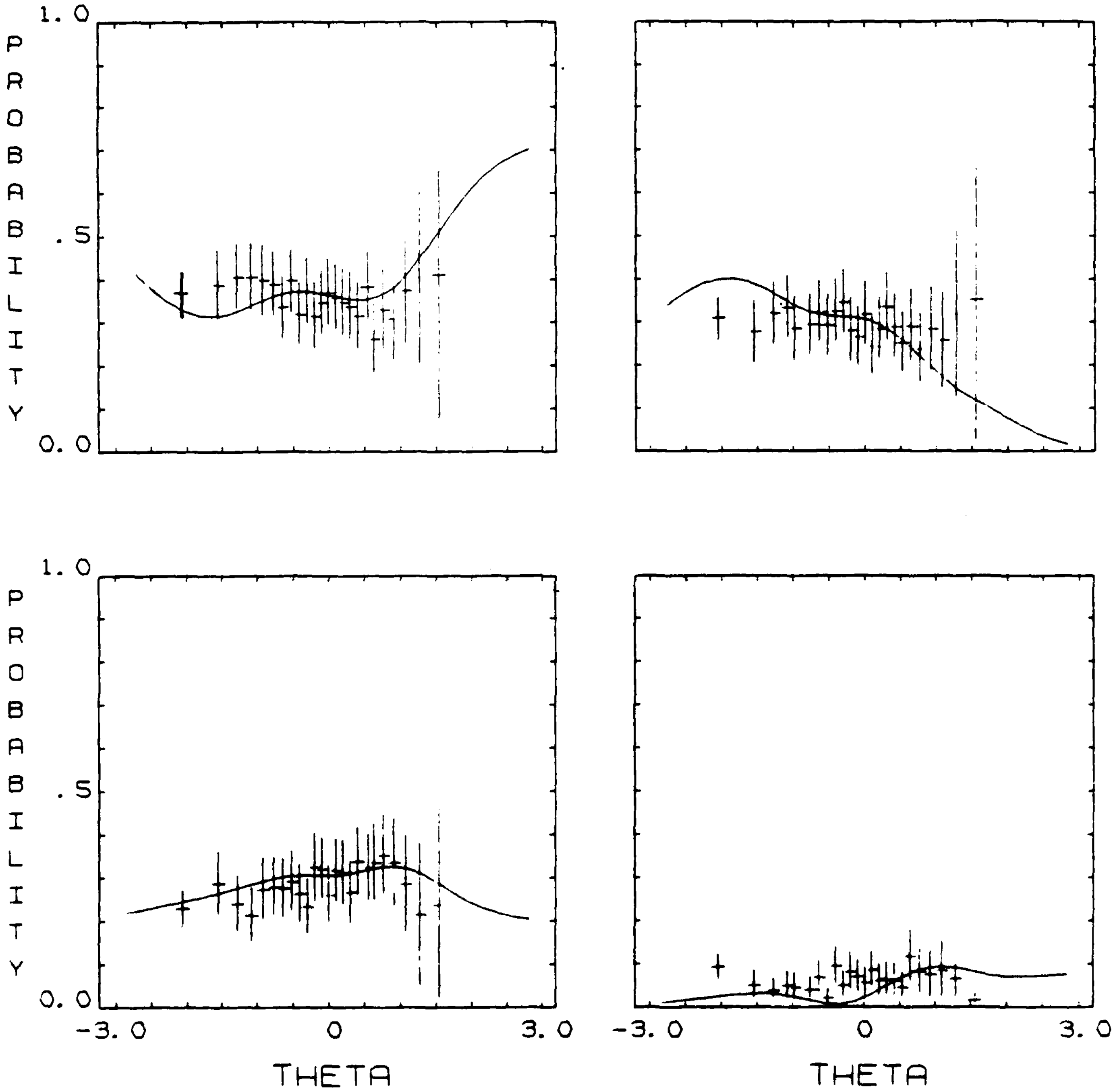


ITEM 17

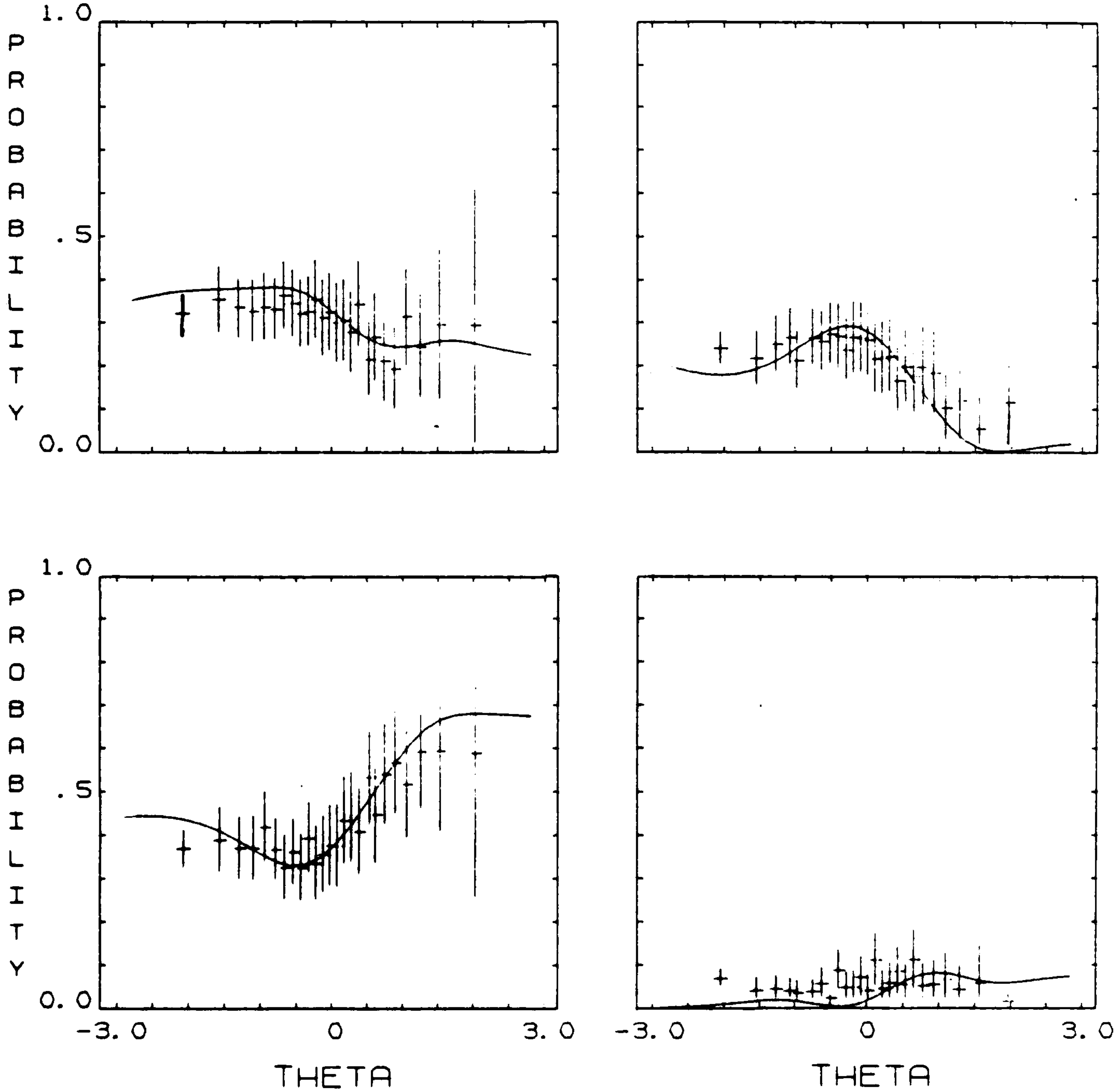


1.0
PROBABILITY
.5
0.0
1.0
PROBABILITY
.5
0.0
-3.0
0
3.0
THETA
-3.0
0
3.0
THETA

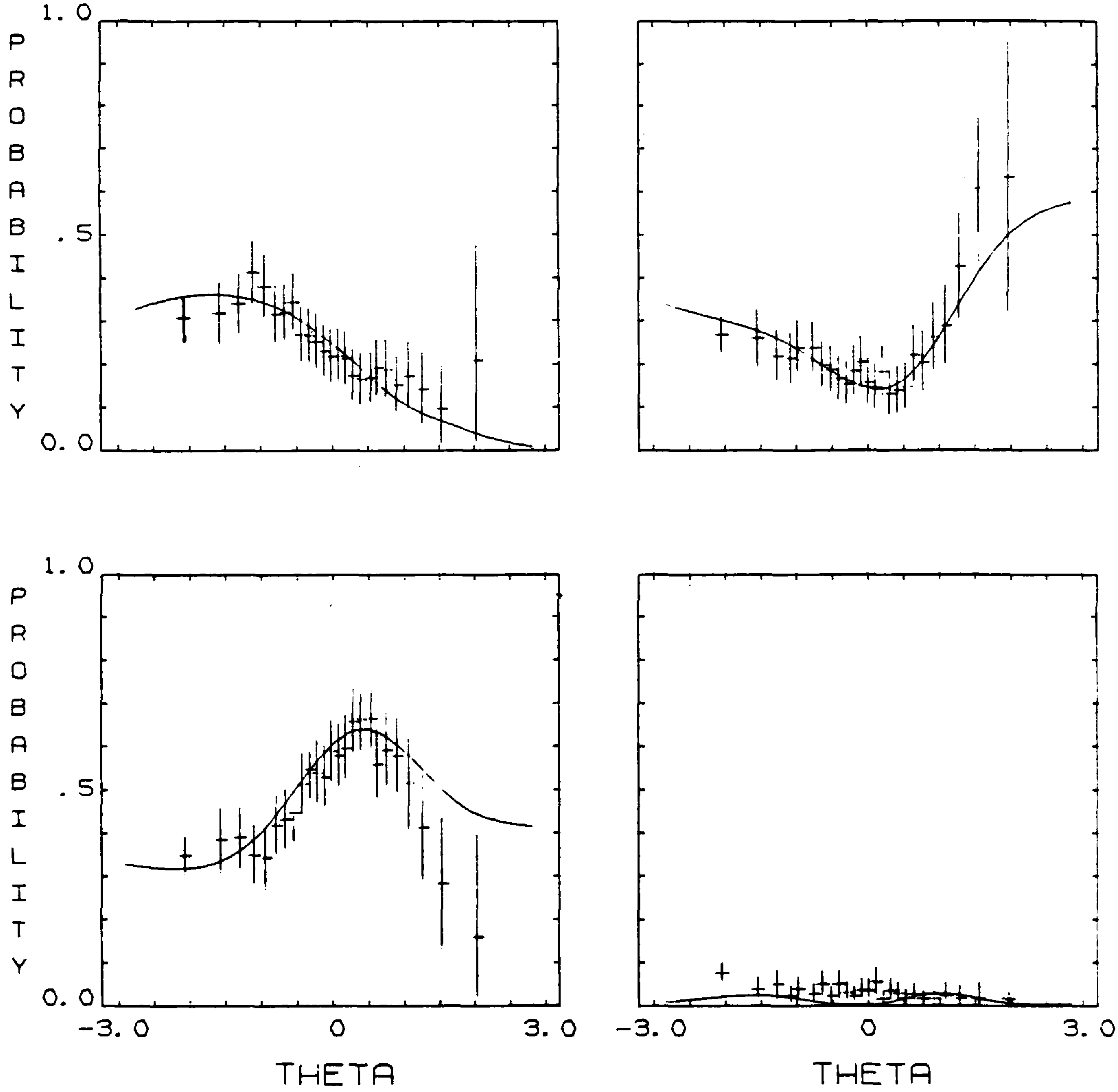
ITEM 18


1.0
PROBABILITY
.5
0.0
1.0
PROBABILITY
.5
0.0
-3.0
0
3.0
-3.0
0
3.0
THETA
THETA

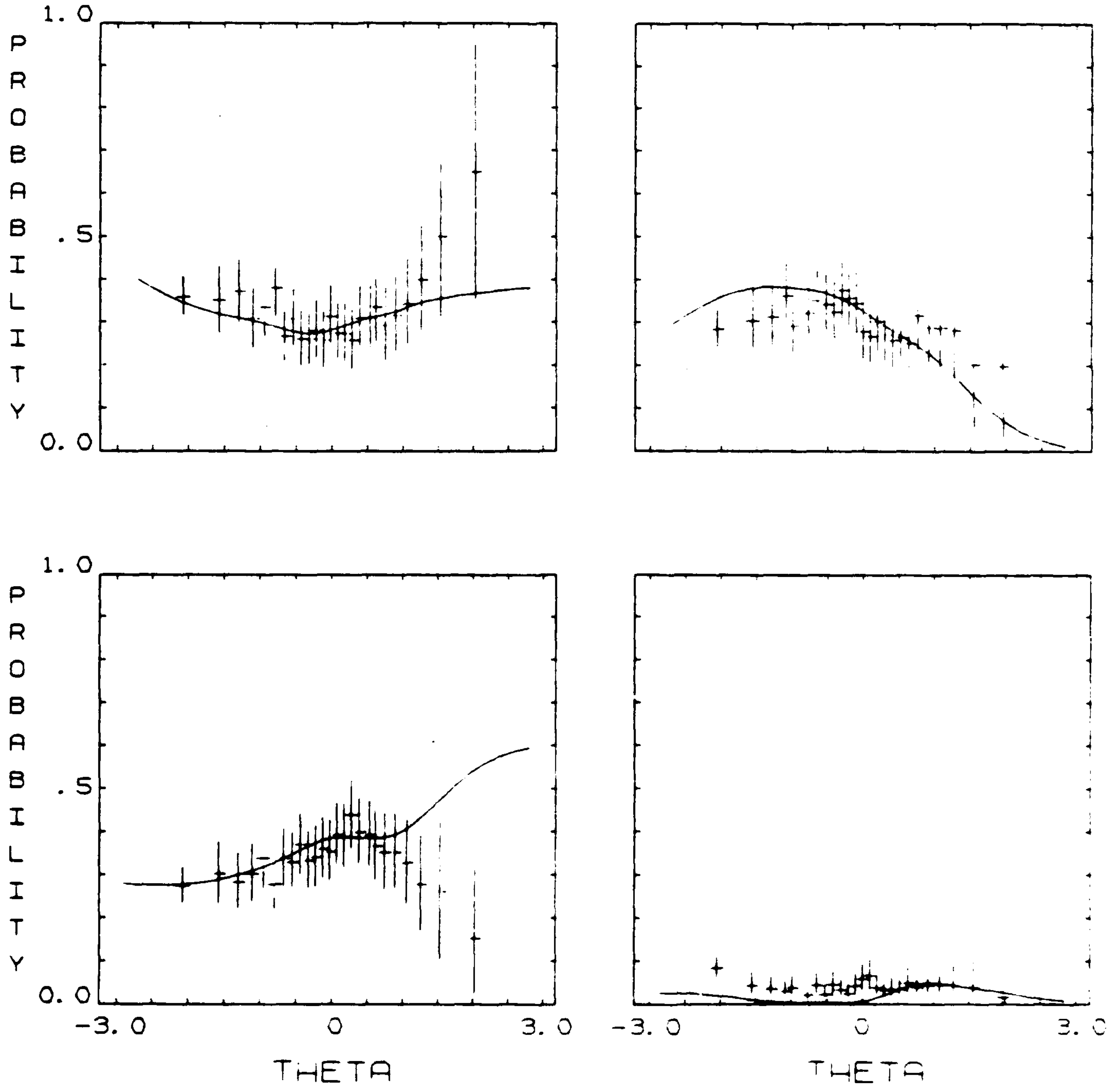
ITEM 19


1.0
.5
0.0
PROBABILITY
1.0
.5
0.0
PROBABILITY
-3.0
0
3.0
THETA
-3.0
0
3.0
THETA

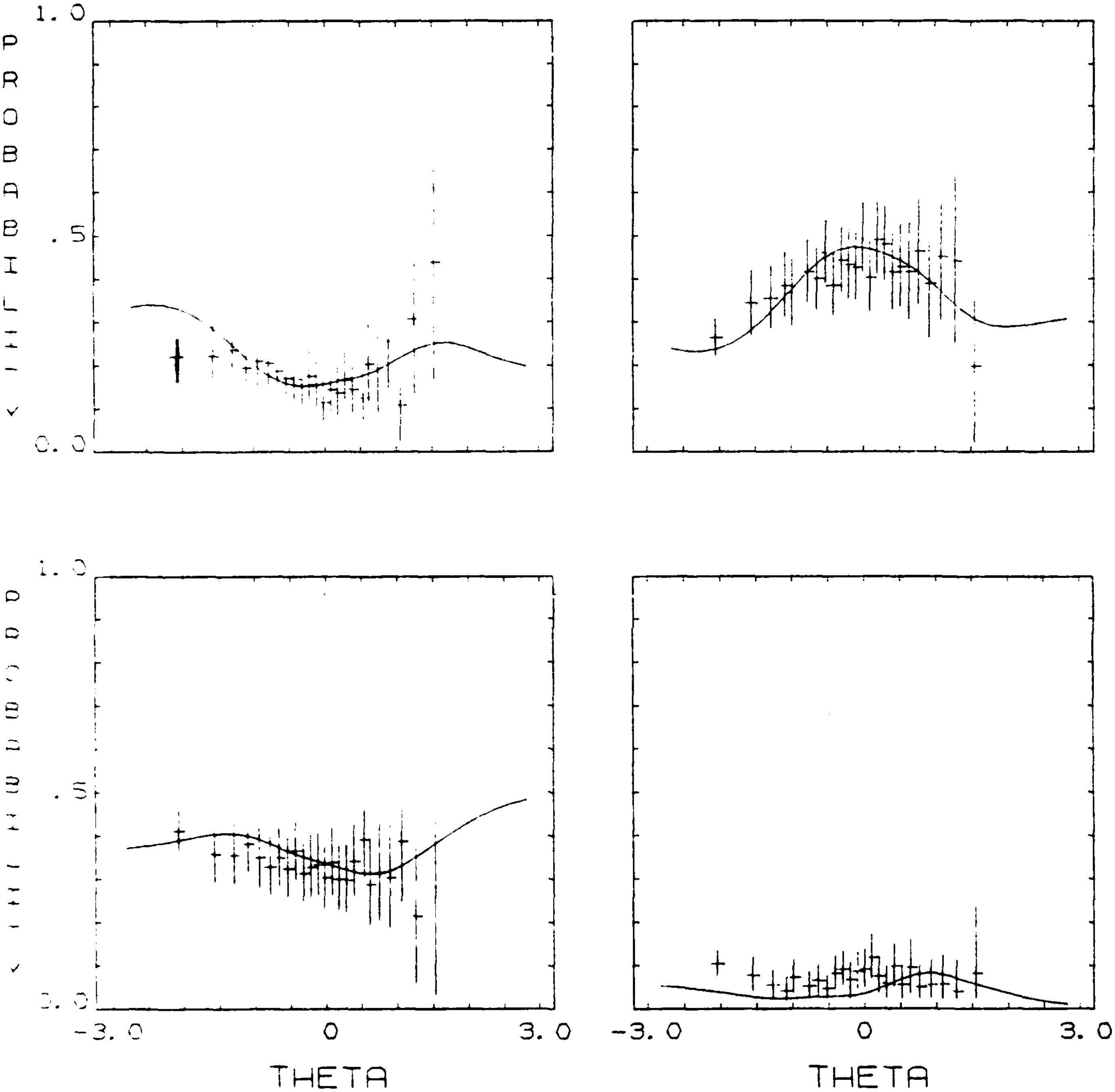
ITEM 20



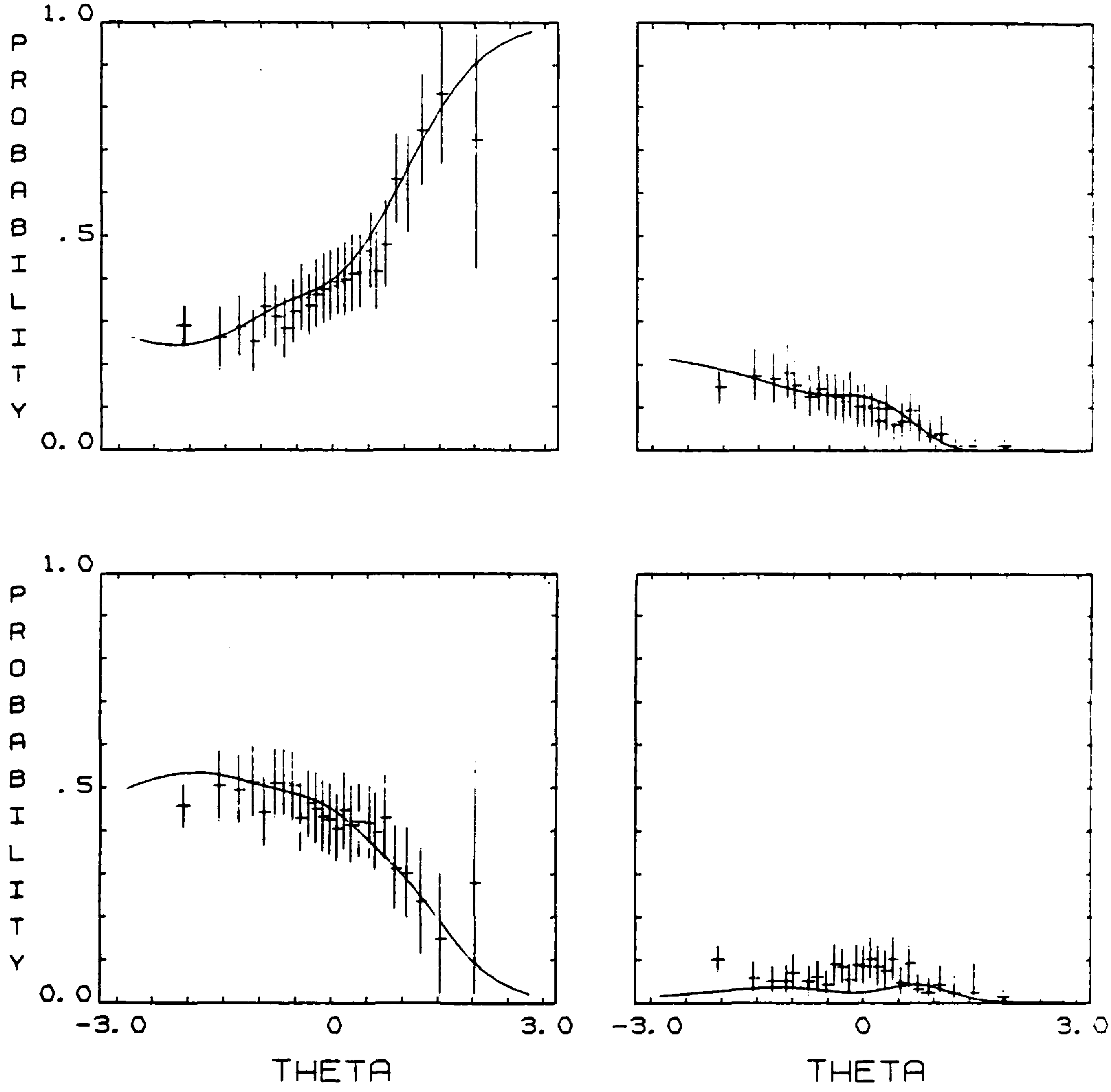
ITEM 21



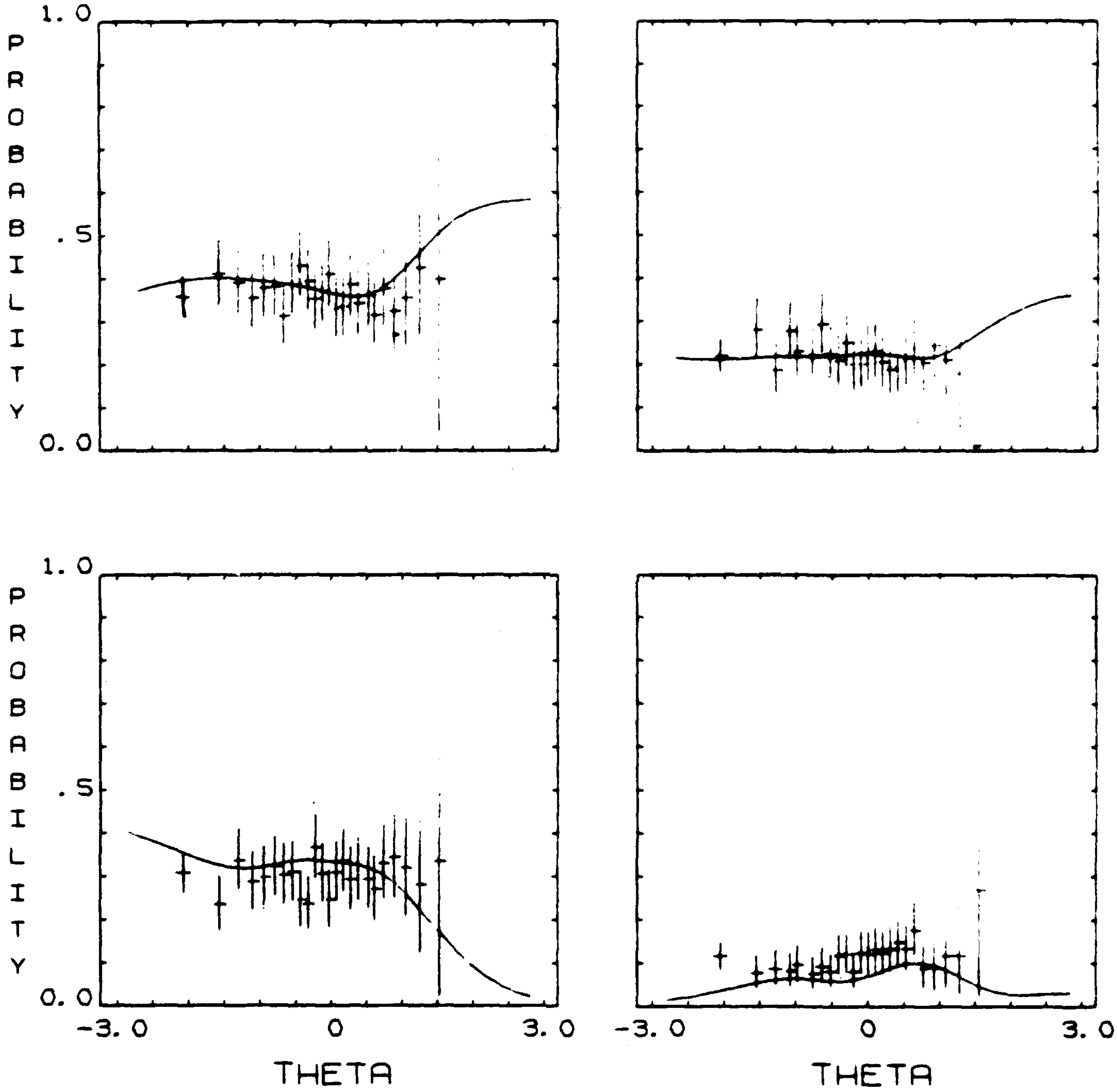
ITEM 22


1.0
PROBABILITY
.5
0.0
1.0
PROBABILITY
.5
0.0
-3.0
0
3.0
THETA
-3.0
0
3.0
THETA

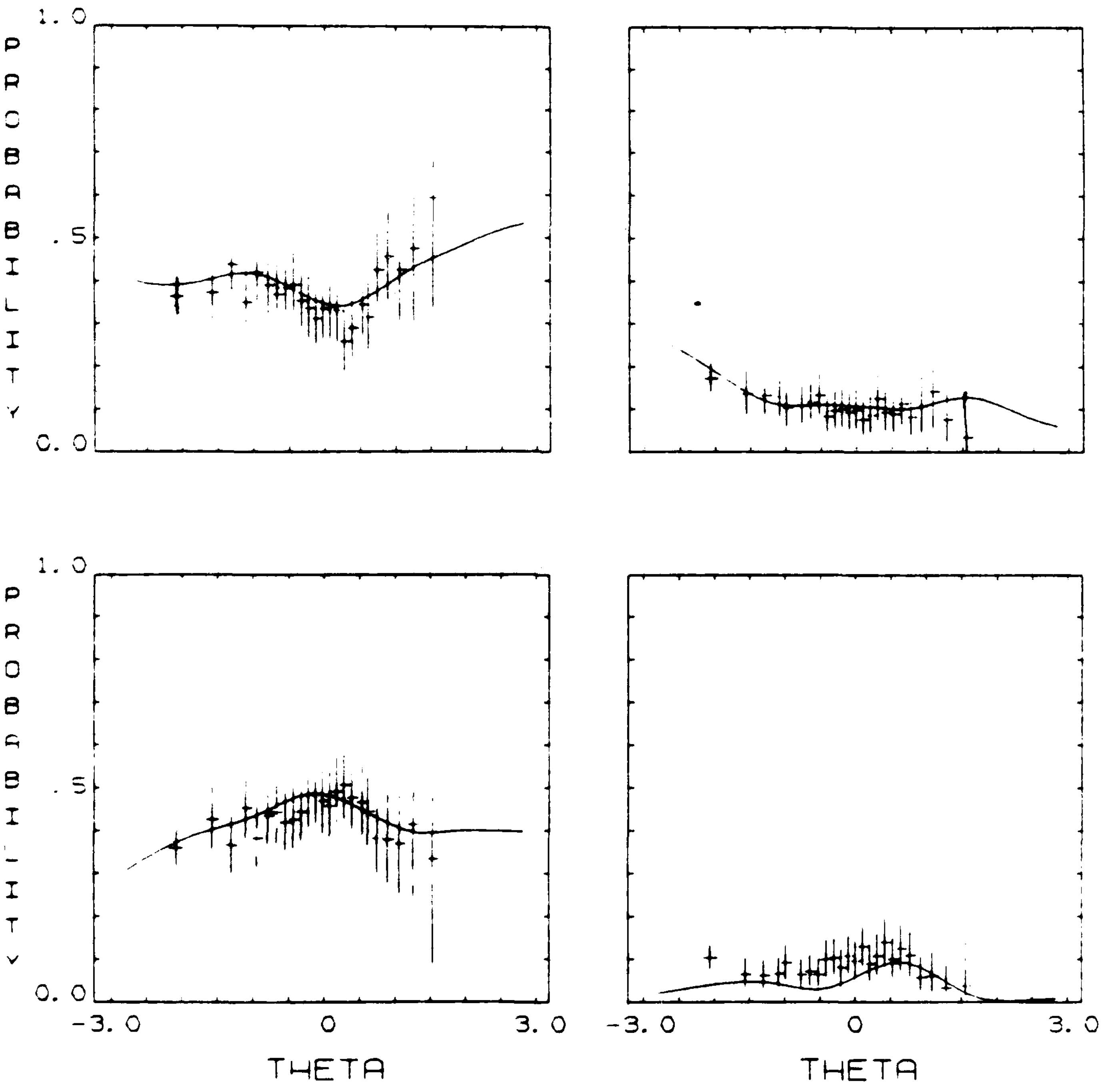
ITEM 23



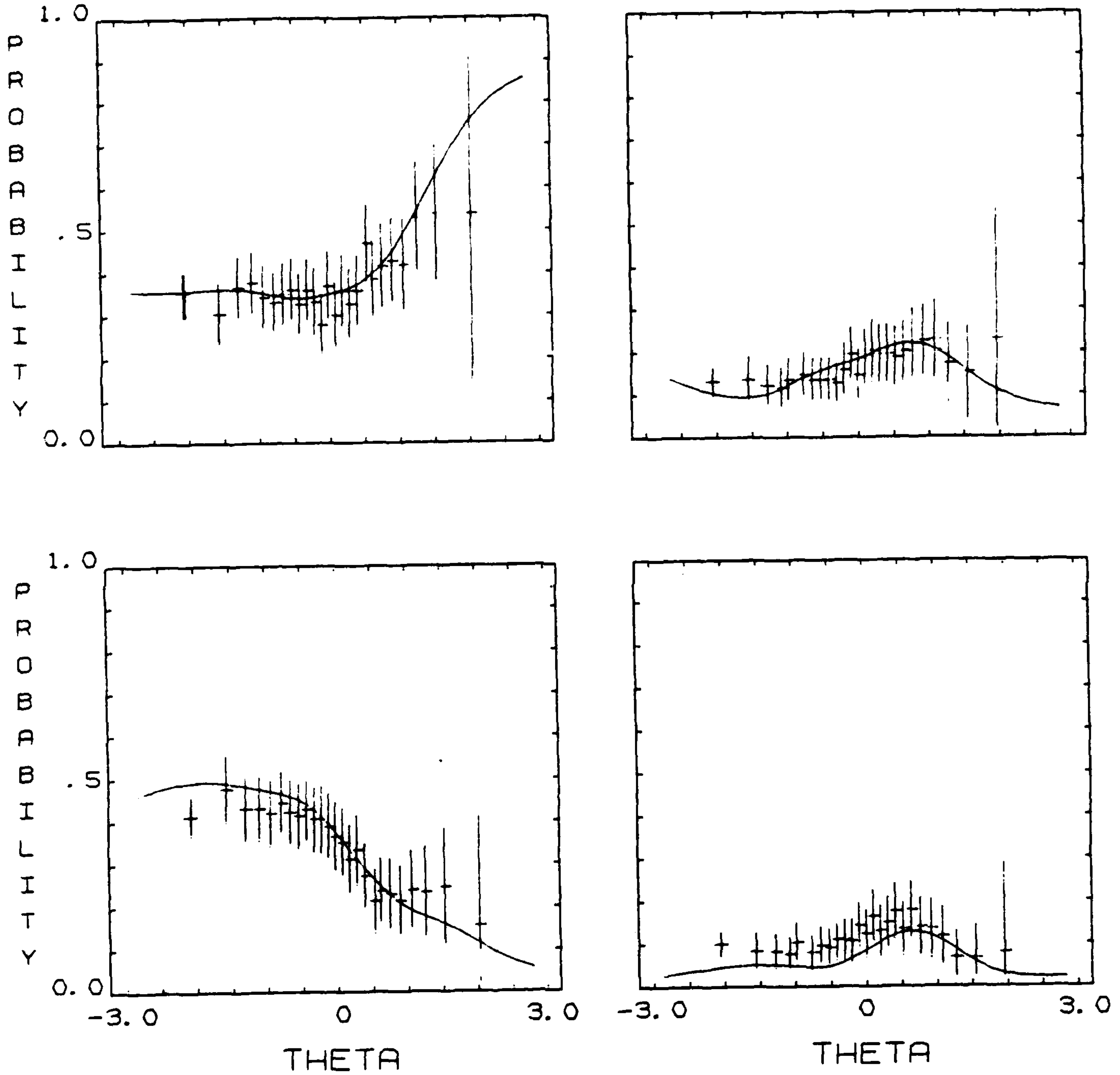
ITEM 24



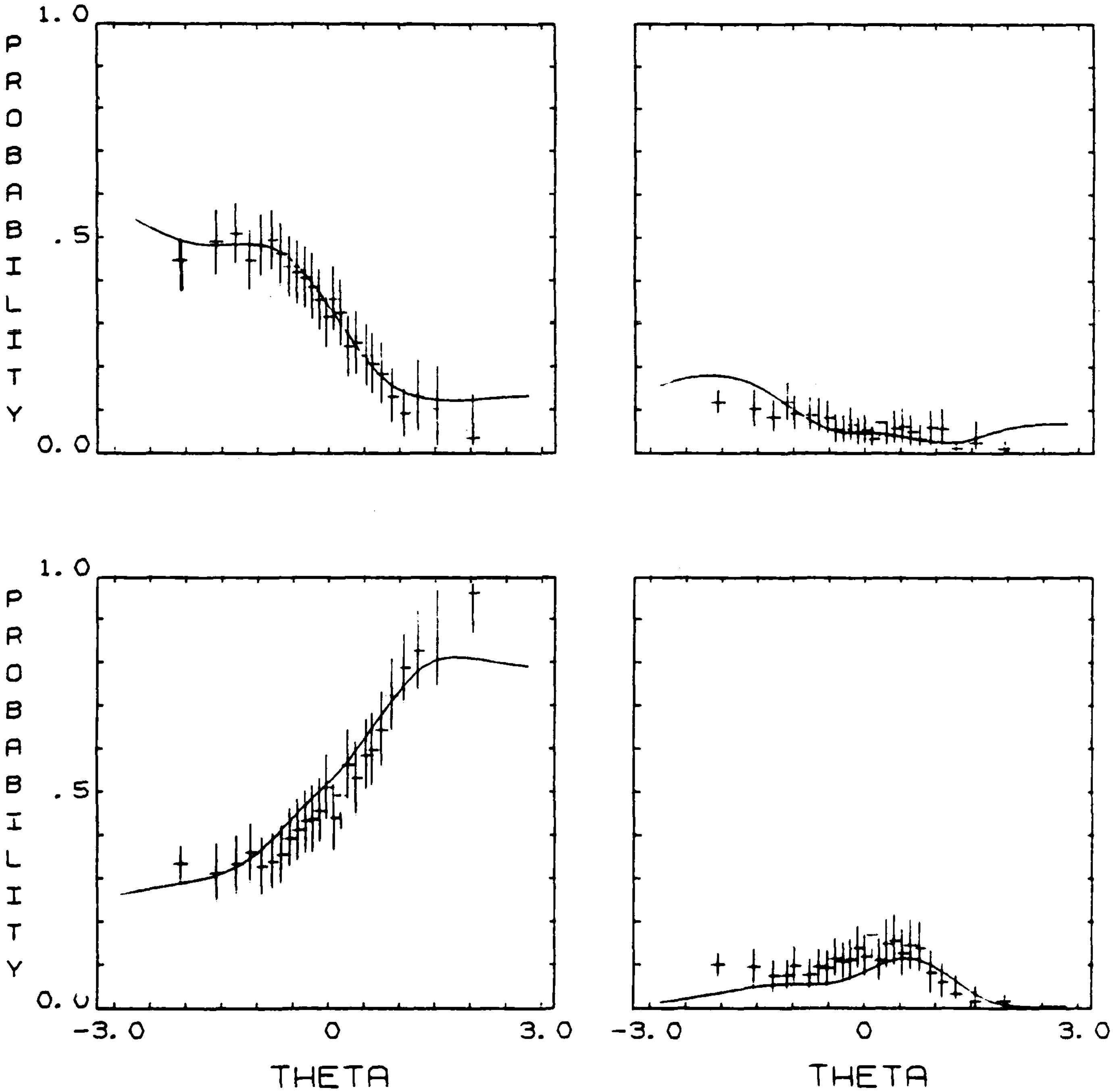
ITEM 25



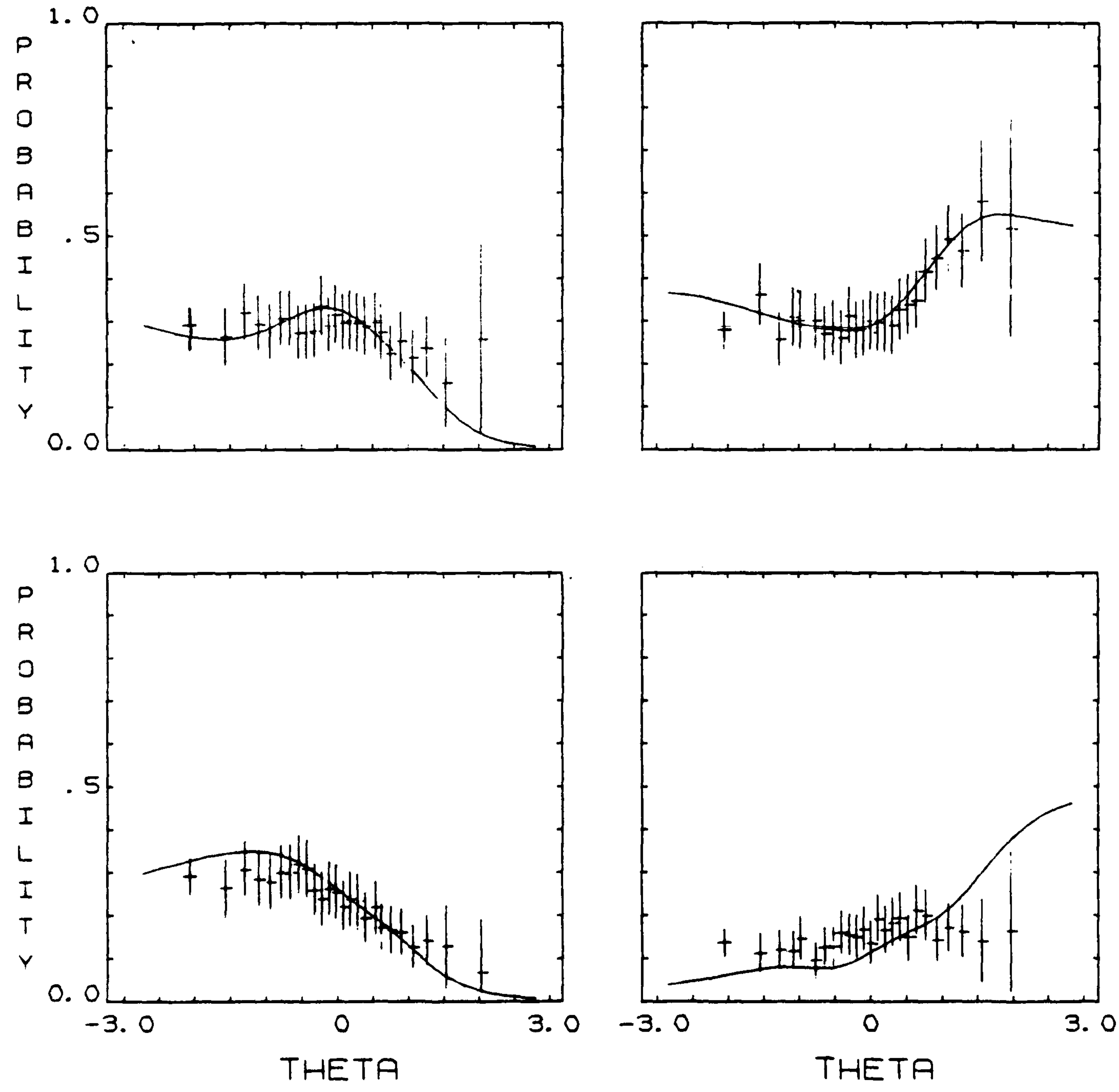
ITEM 26

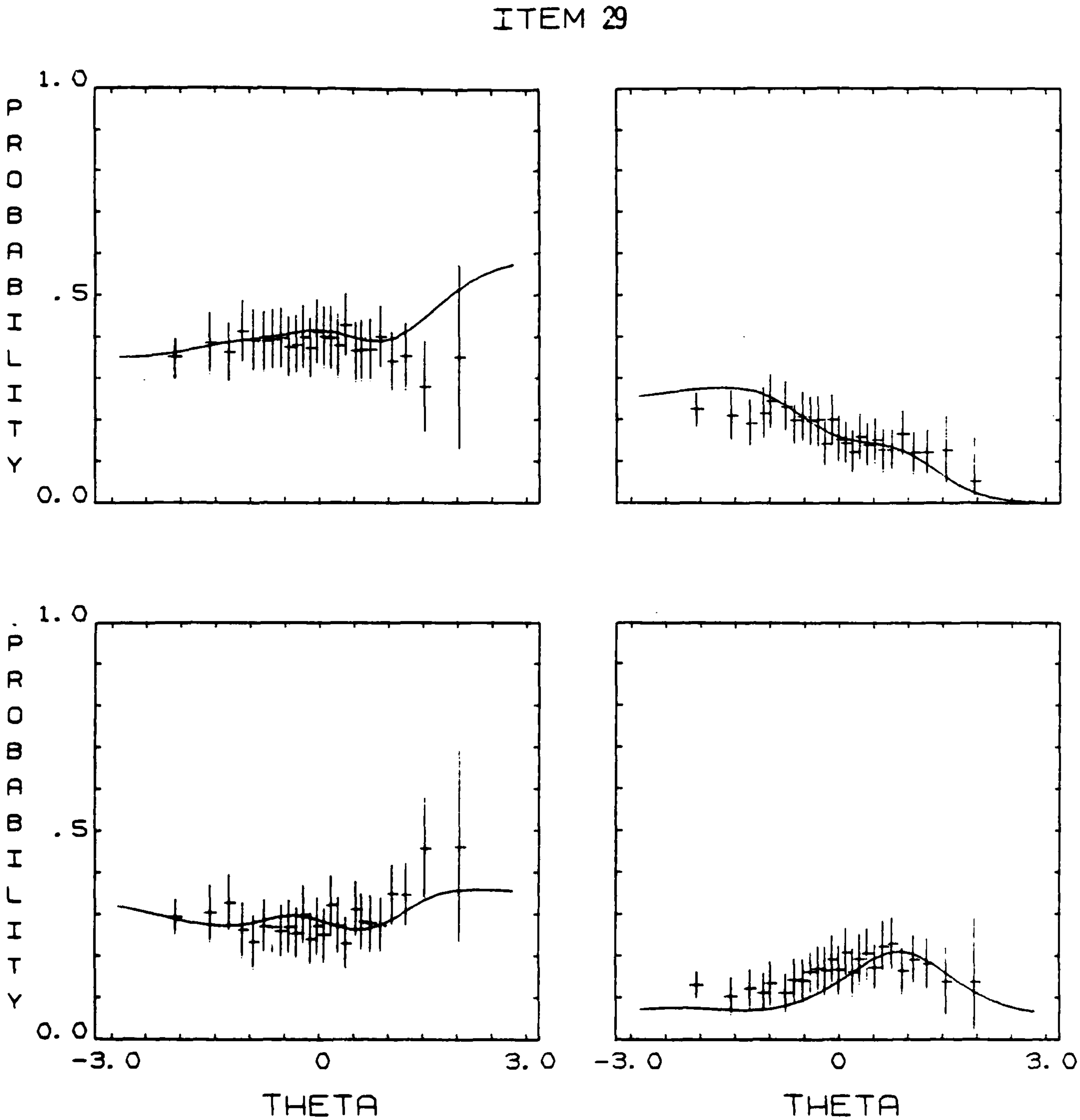


ITEM 27



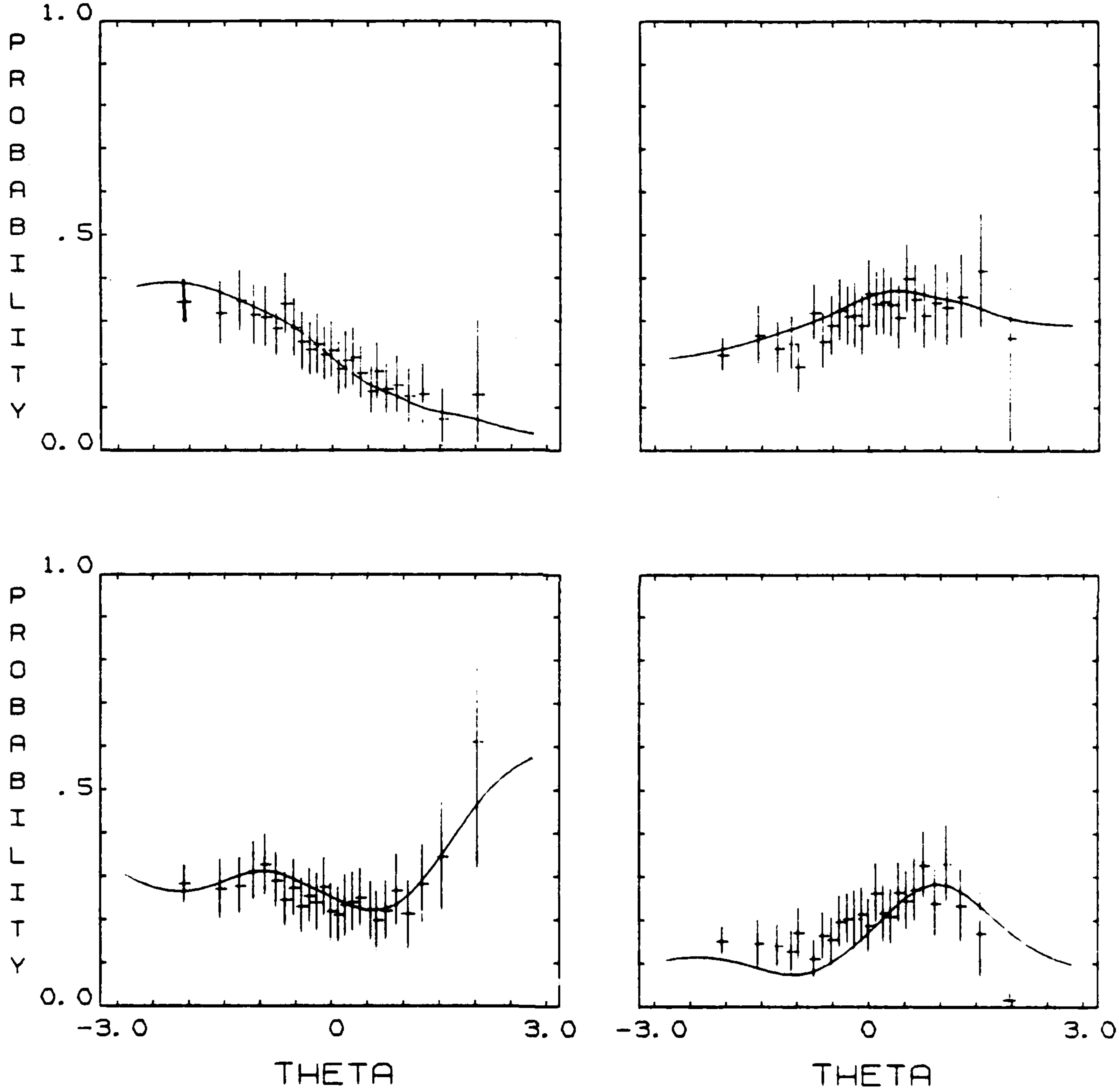
ITEM 28





ITEM 29
1.0
.5
0.0
PROBABILITY
-3.0
0
3.0
THETA

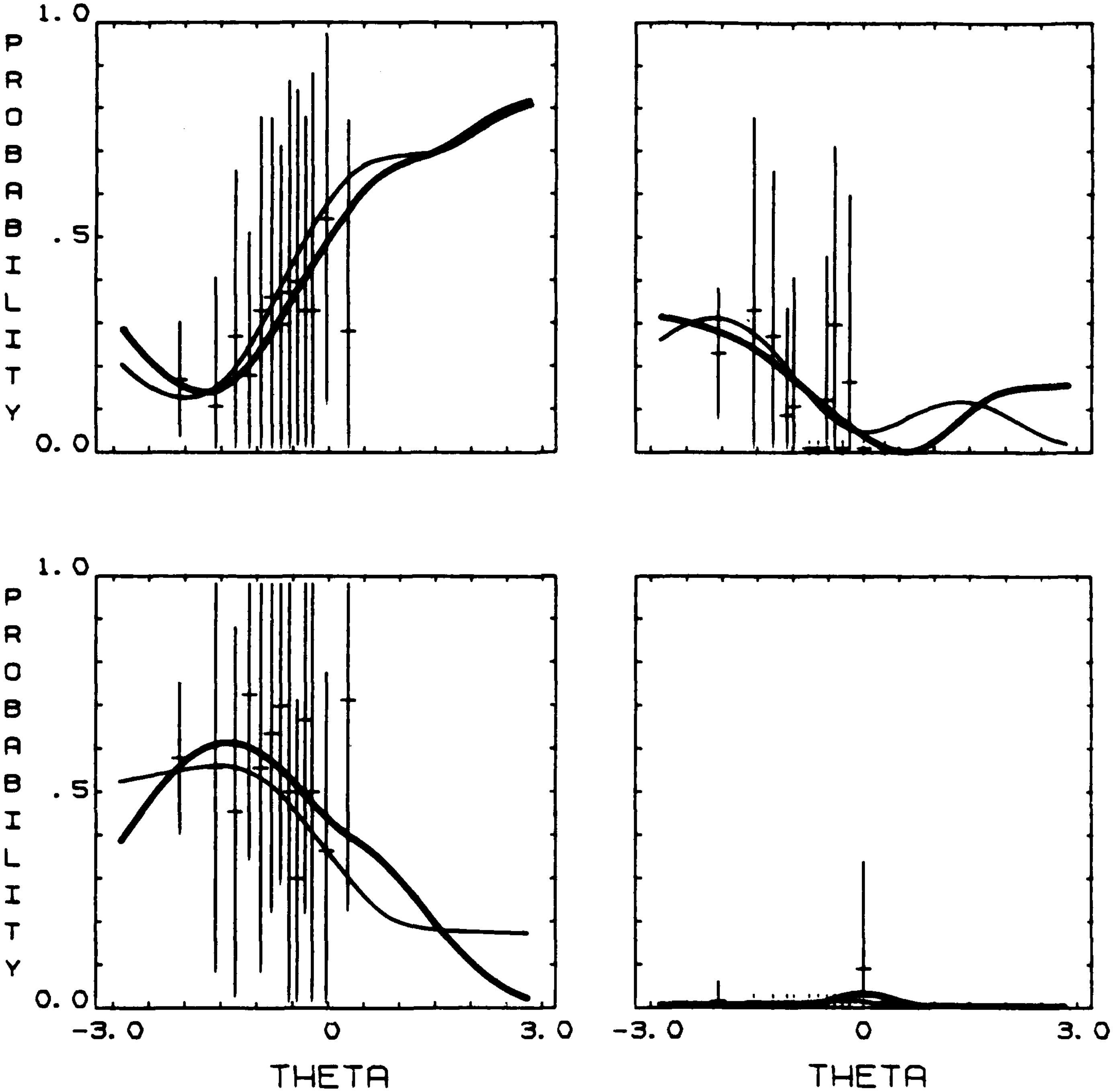
ITEM 30



Appendix 2

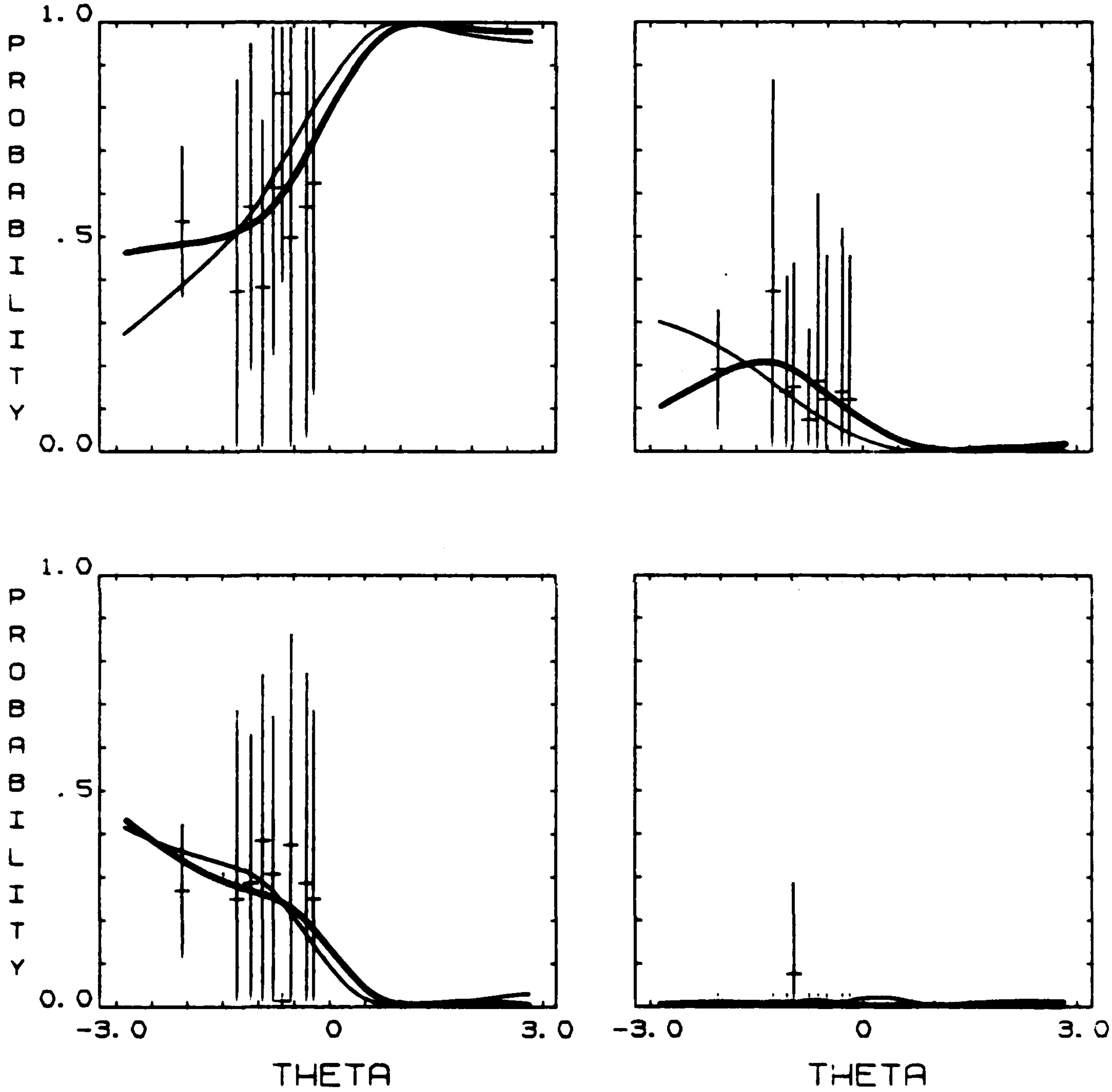
Estimated COCCs, Simulation COCCs, and Empirical Proportions from Estimation Sample.

ITEM 1

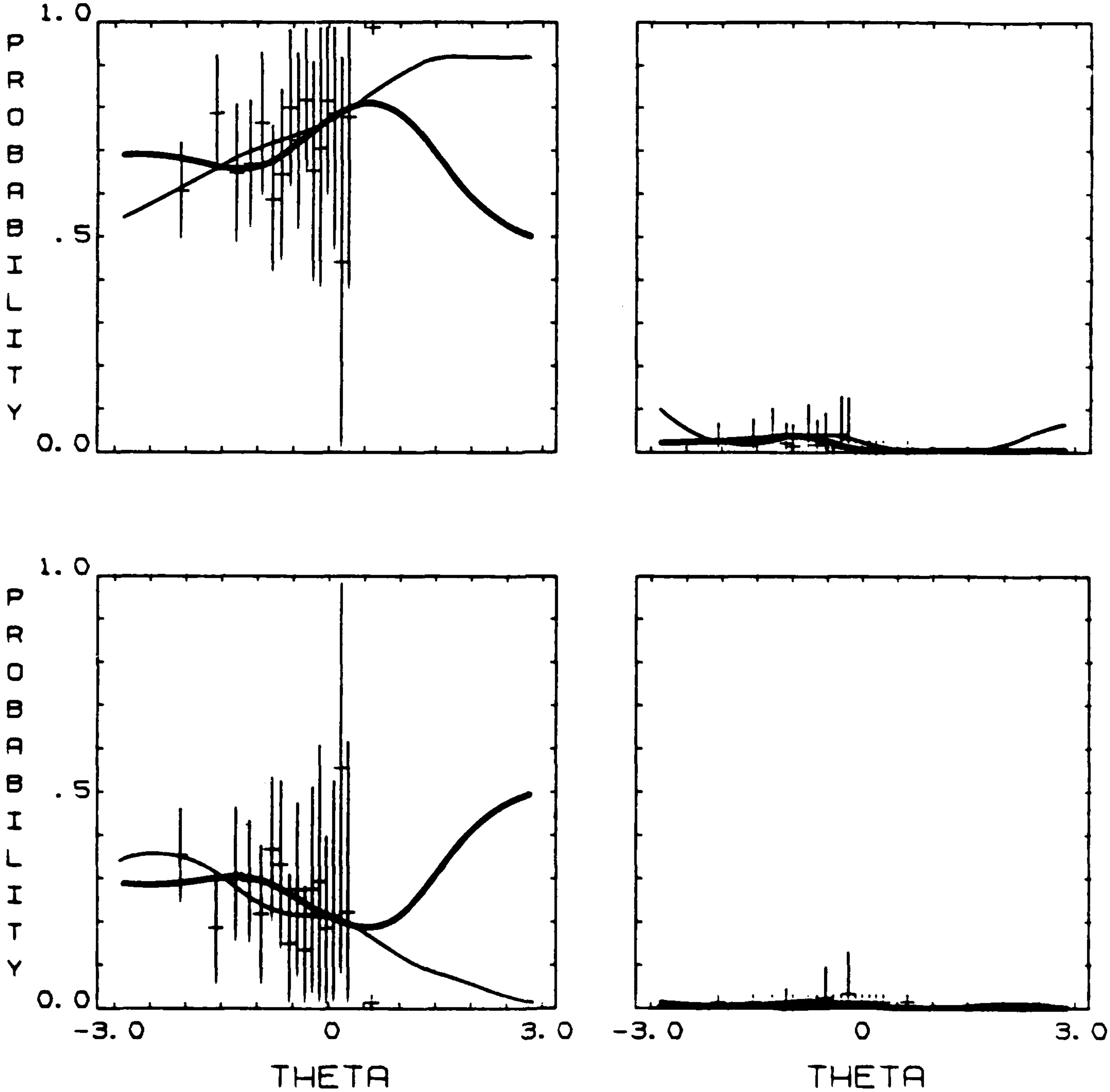


1.0
PROBABILITY
.5
0.0
1.0
PROBABILITY
.5
0.0
-3.0
0
3.0
-3.0
0
3.0
THETA
THETA

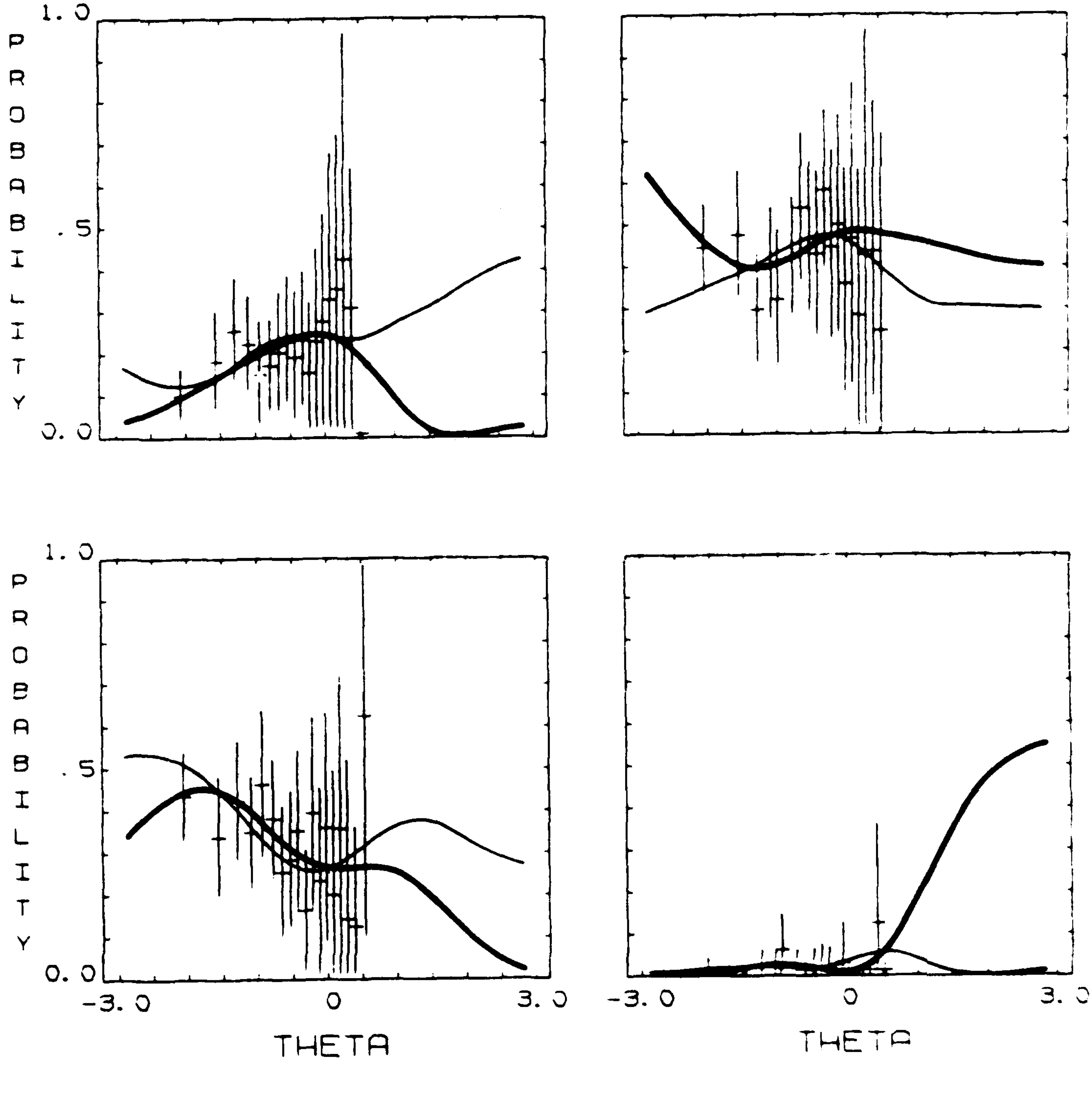
ITEM 2



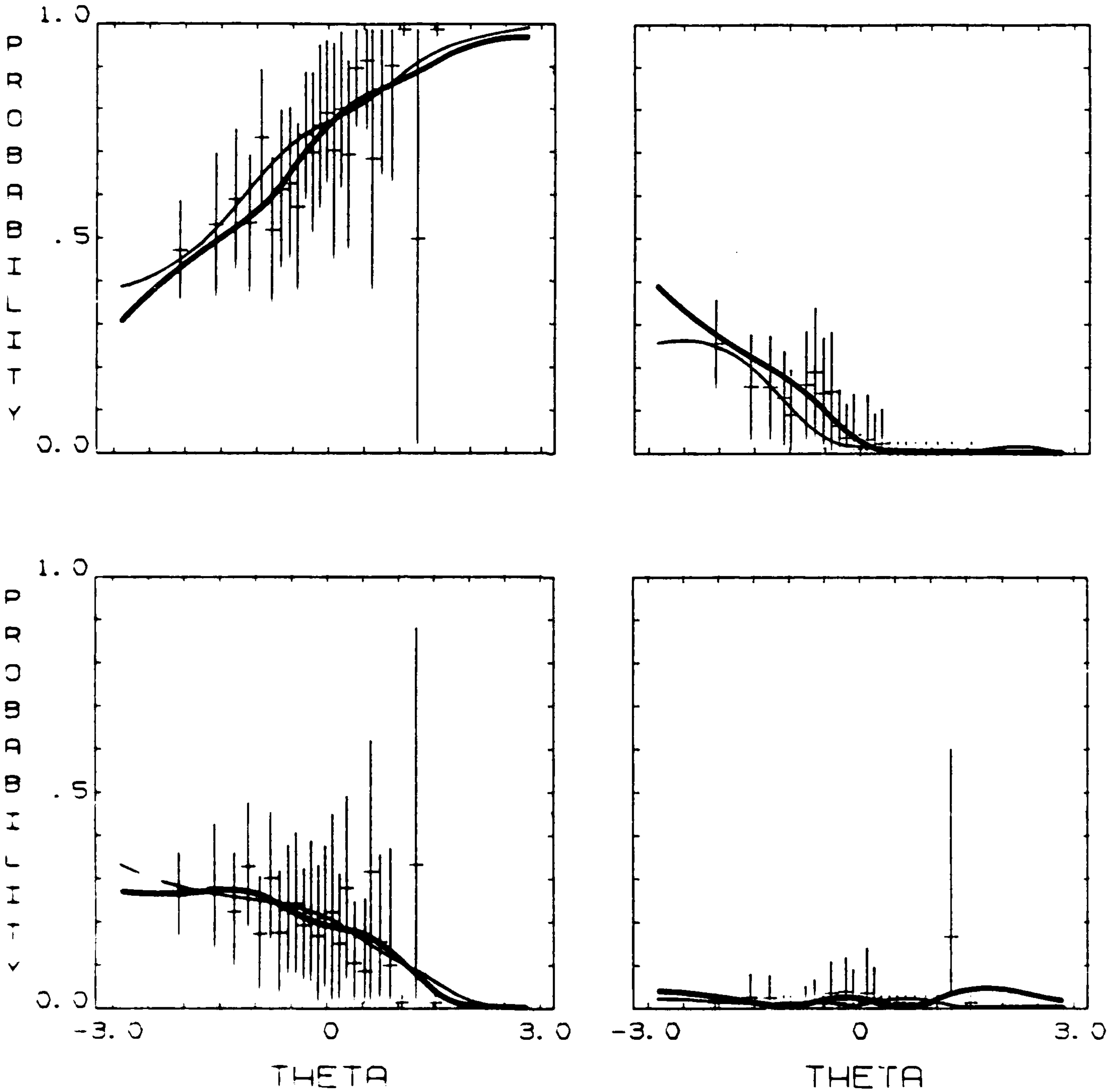
ITEM 3


1.0
PROBABILITY
.5
0.0
1.0
PROBABILITY
.5
0.0
-3.0
0
3.0
THETA
-3.0
0
3.0
THETA

ITEM 4

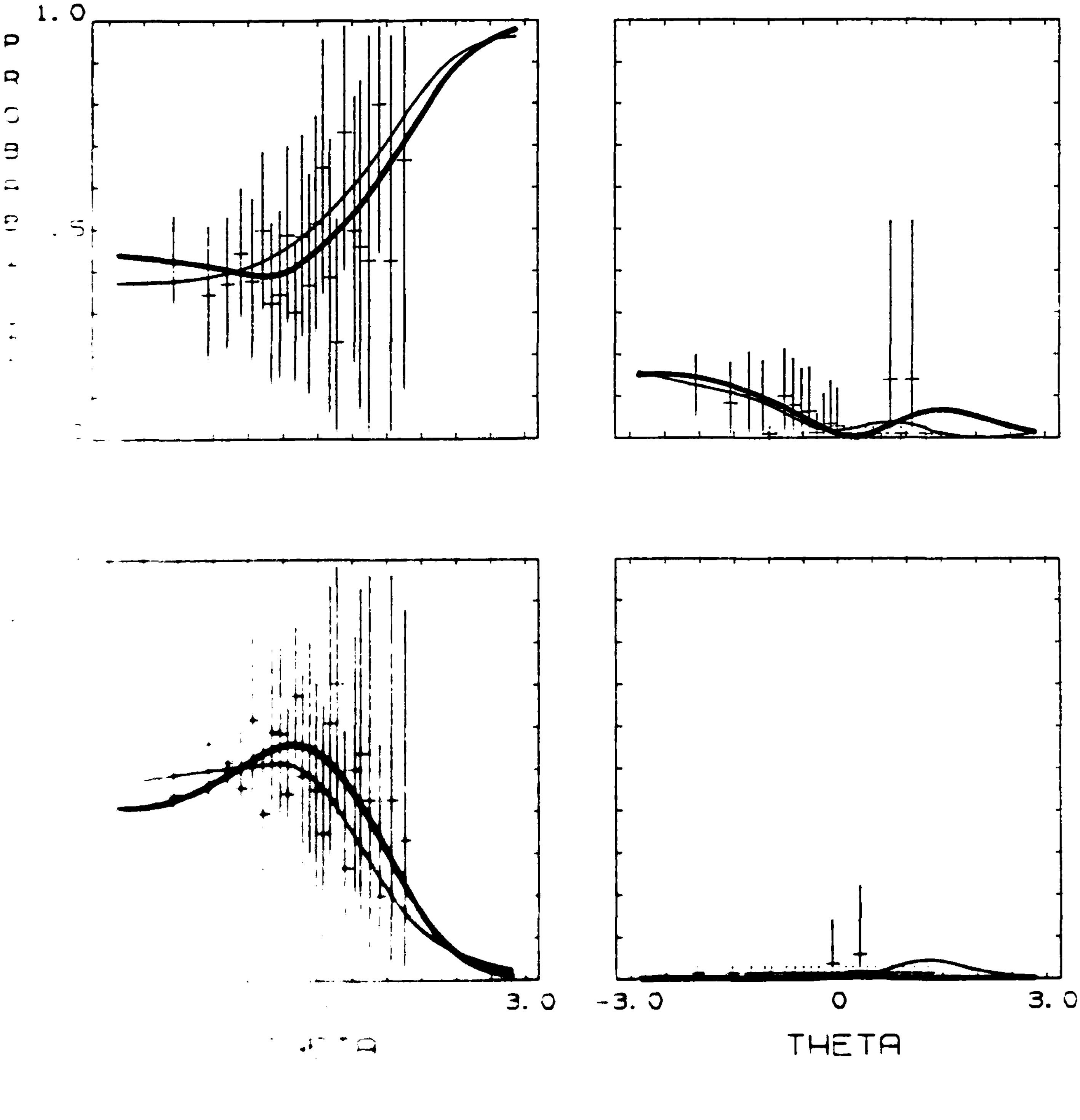

1.0
PROBABILITY
.5
0.0
1.0
PROBABILITY
.5
0.0
-3.0
0
3.0
THETA
-3.0
0
3.0
THETA

ITEM 5

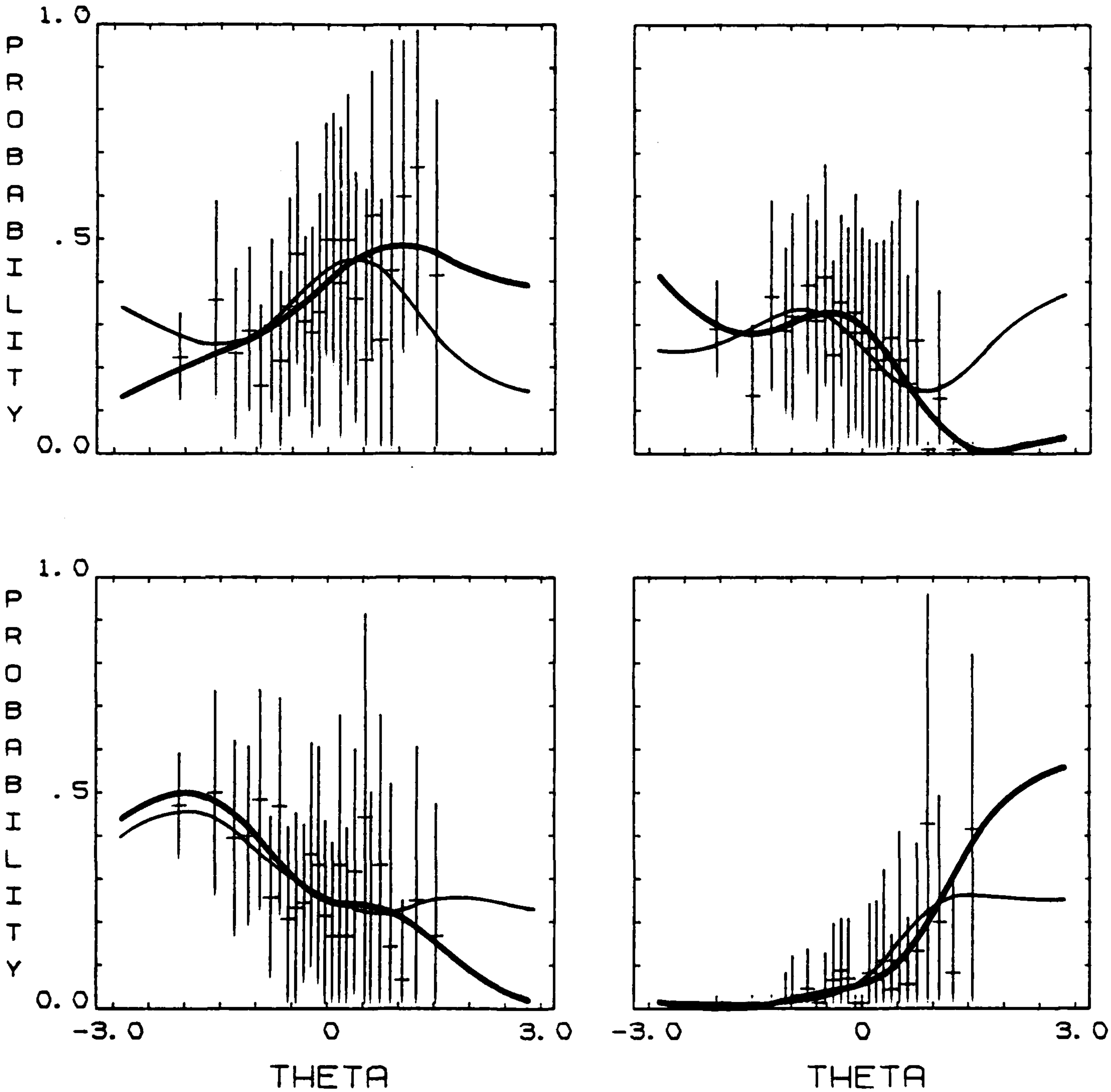


1.0
.5
0.0
PROBABILITY
1.0
.5
0.0
PROBABILITY
-3.0
0
3.0
-3.0
0
3.0
THETA
THETA

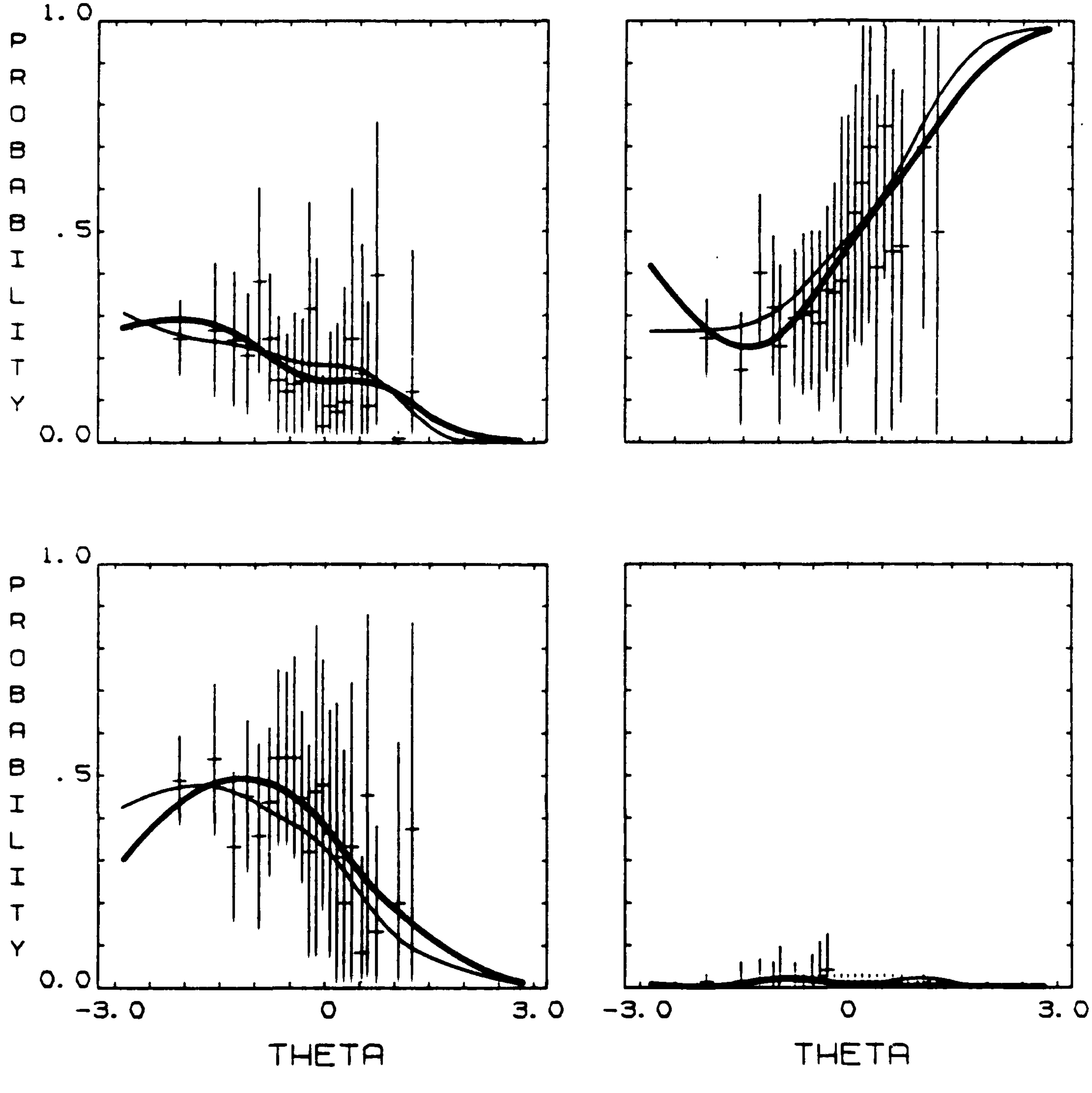
ITEM 6



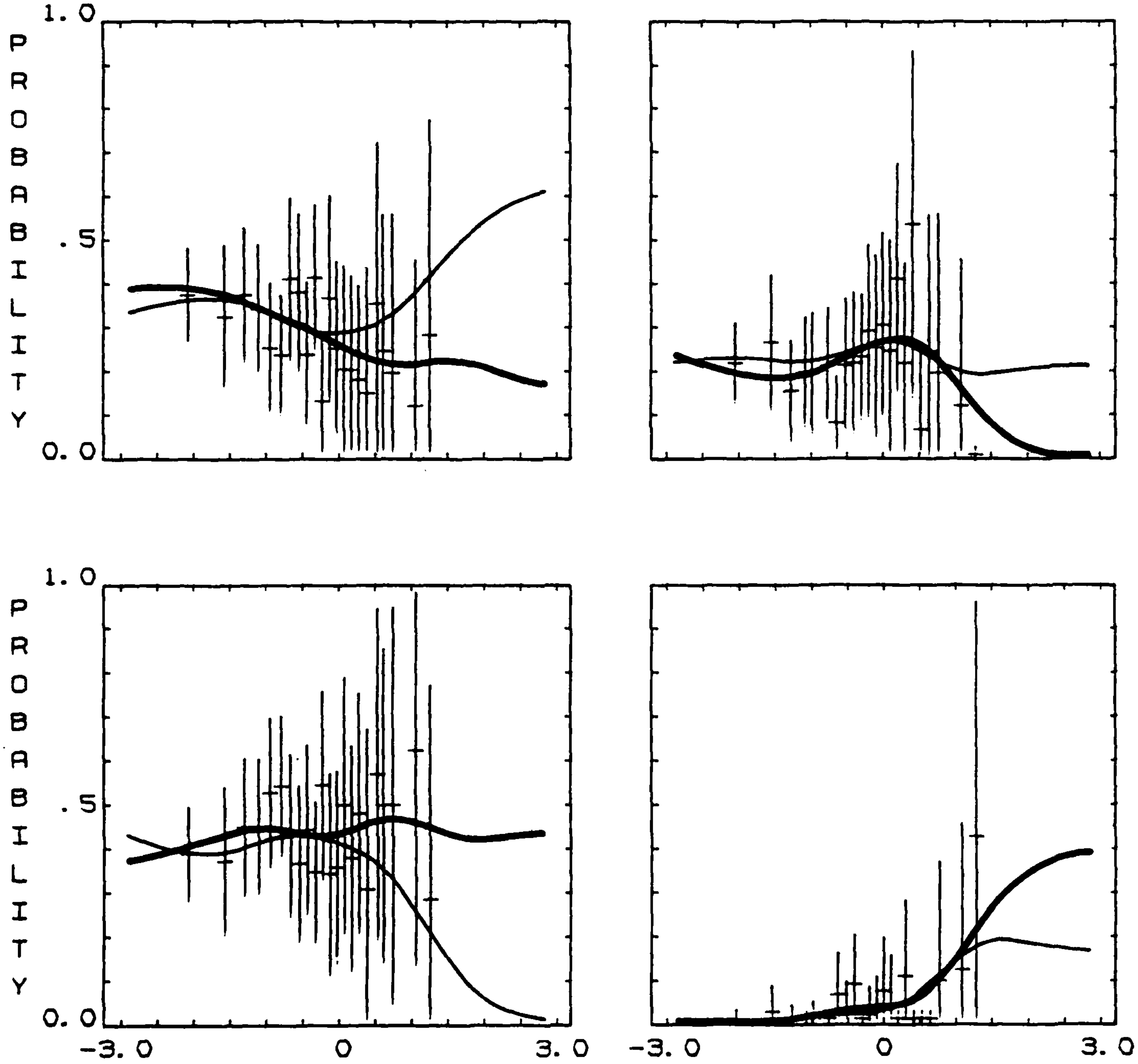
ITEM 7



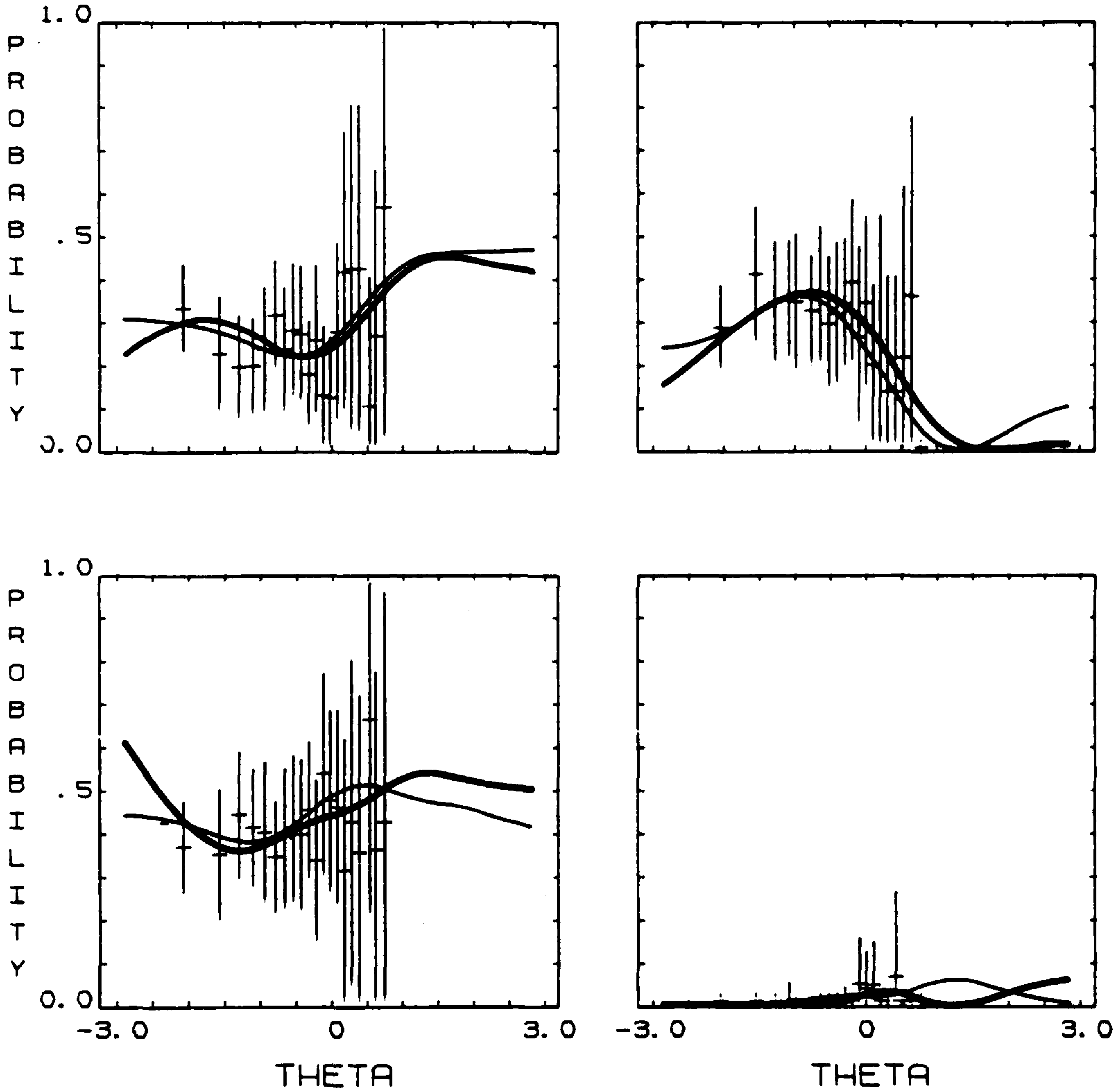
ITEM 8



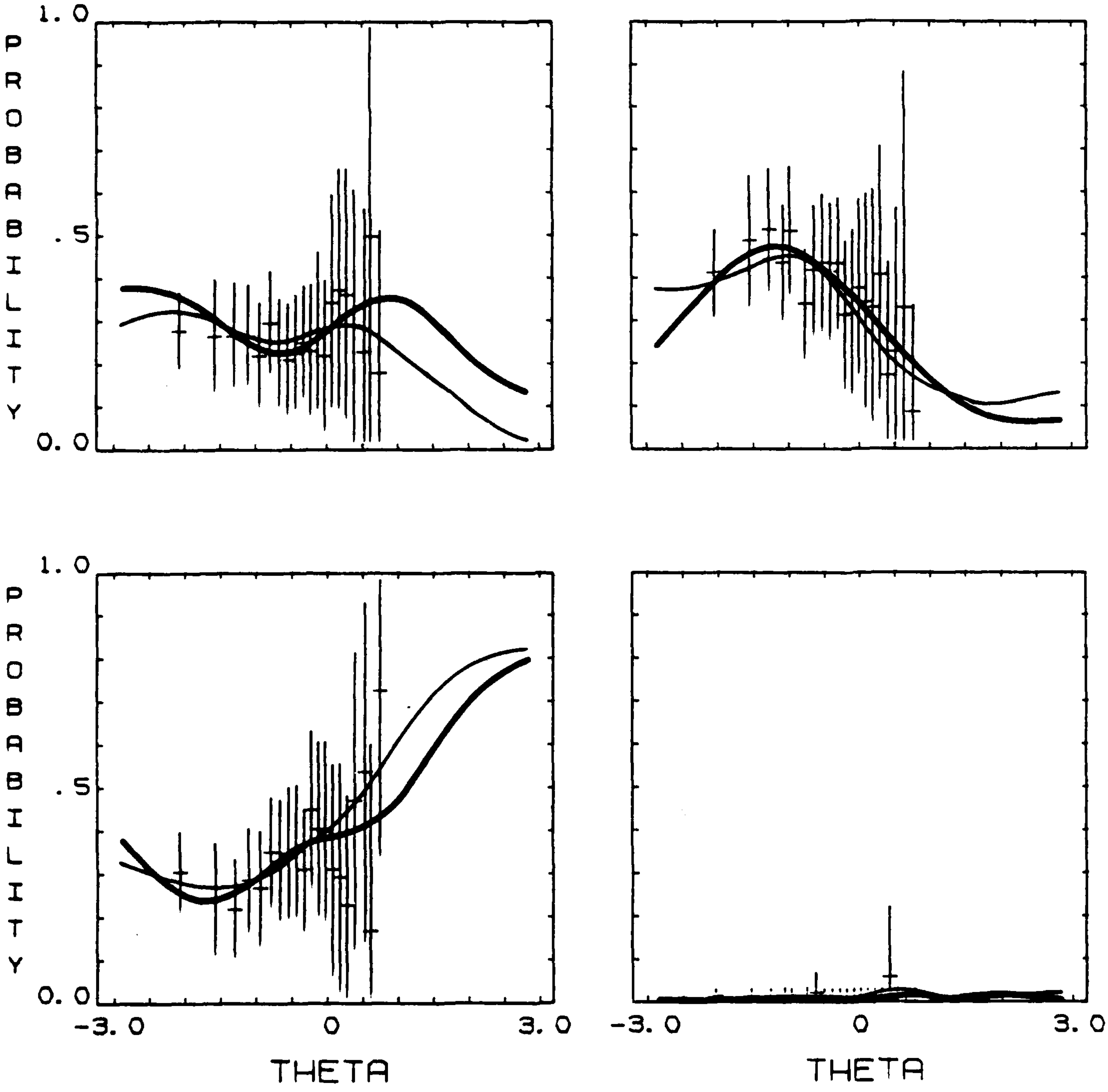
ITEM 9



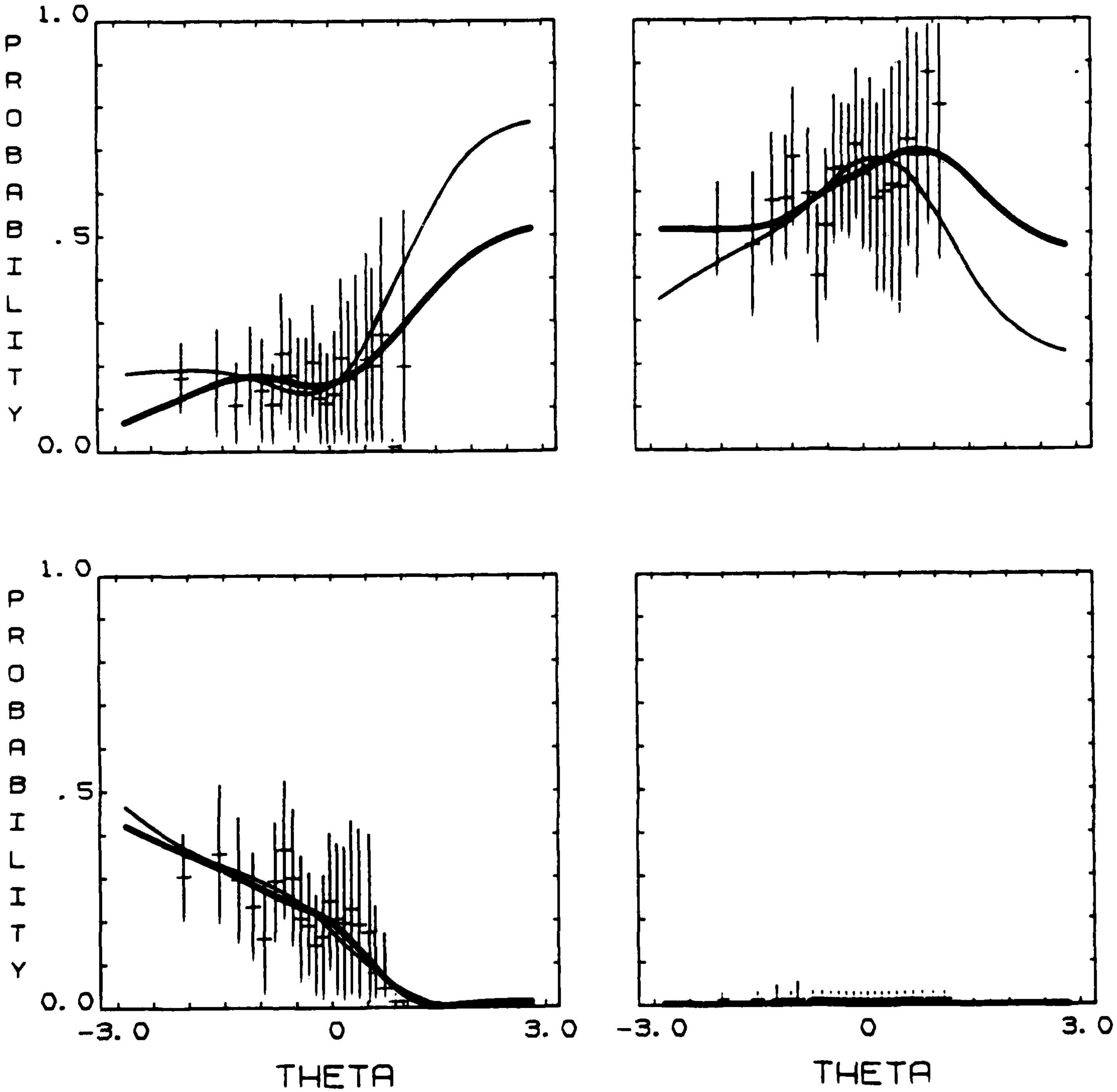
ITEM 10



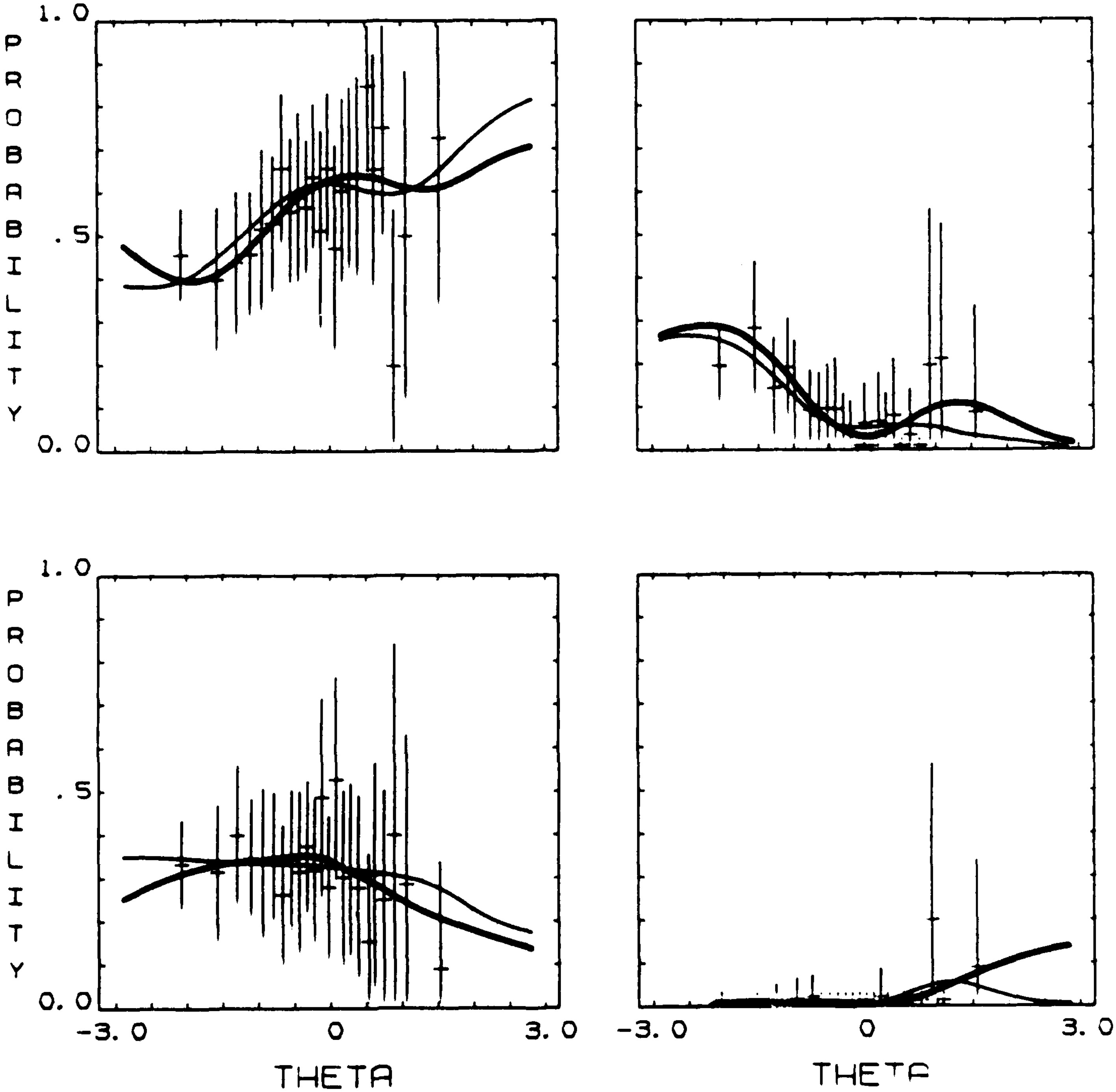
ITEM 11



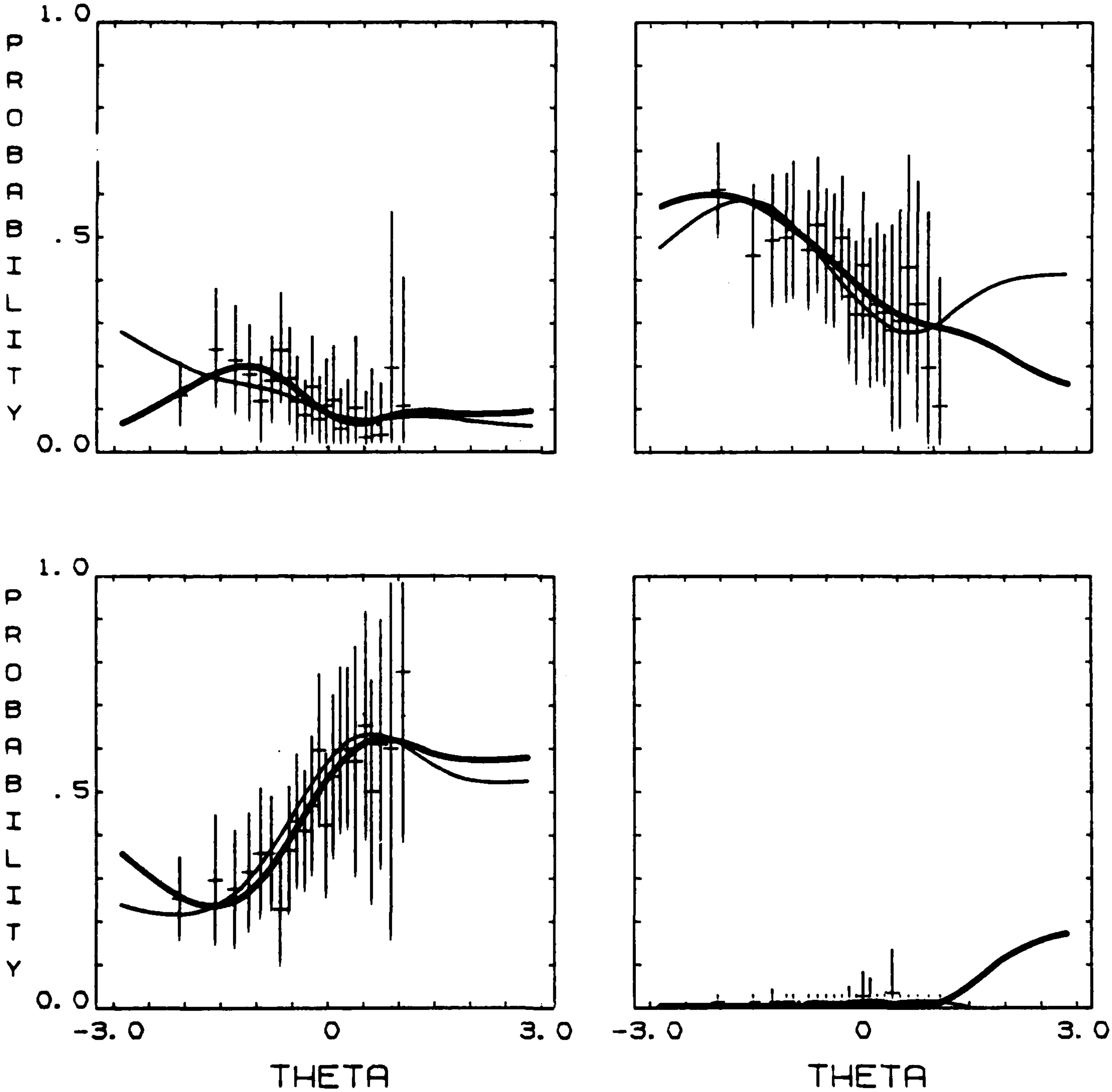
ITEM 12



ITEM 13

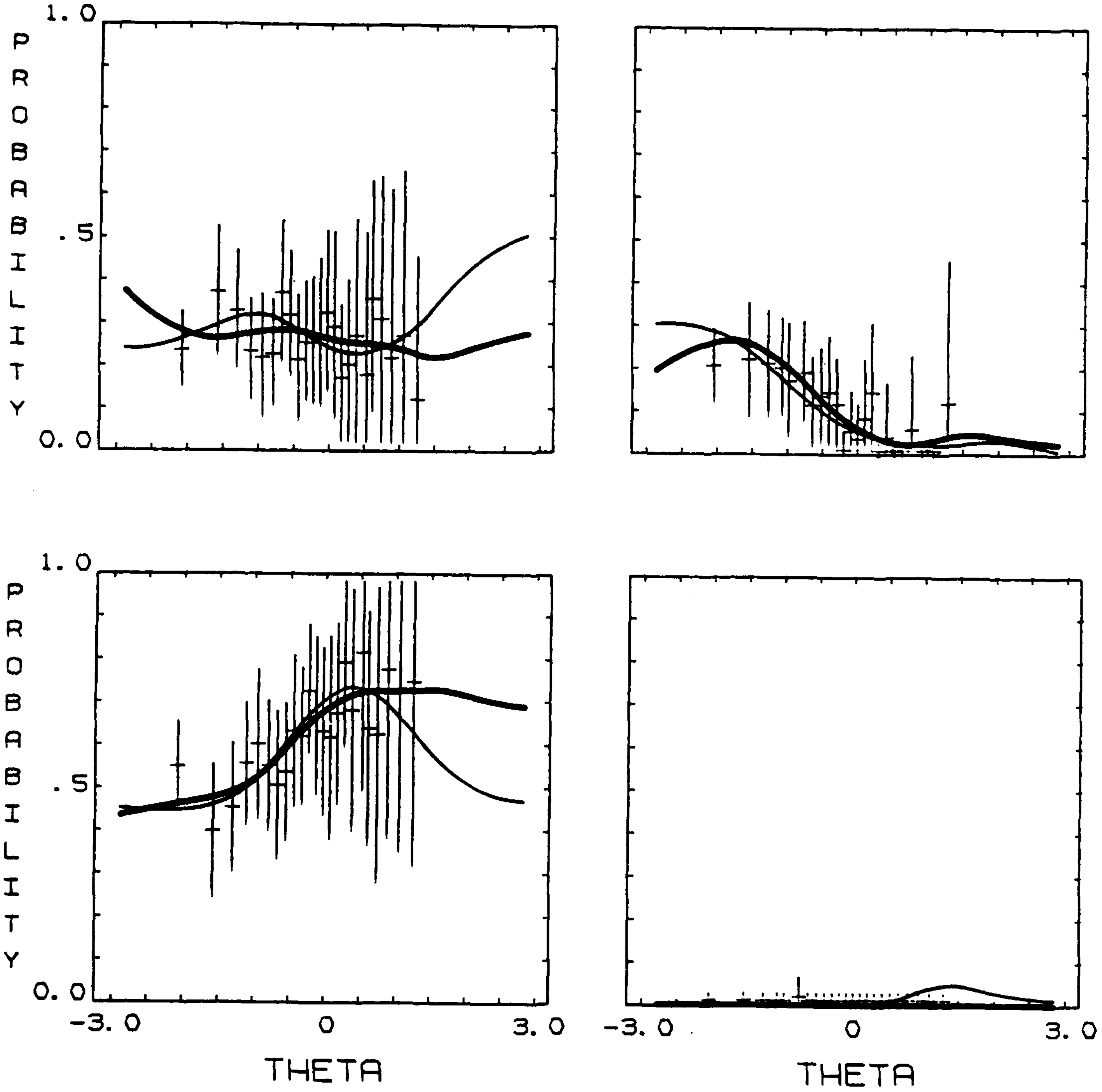


ITEM 14

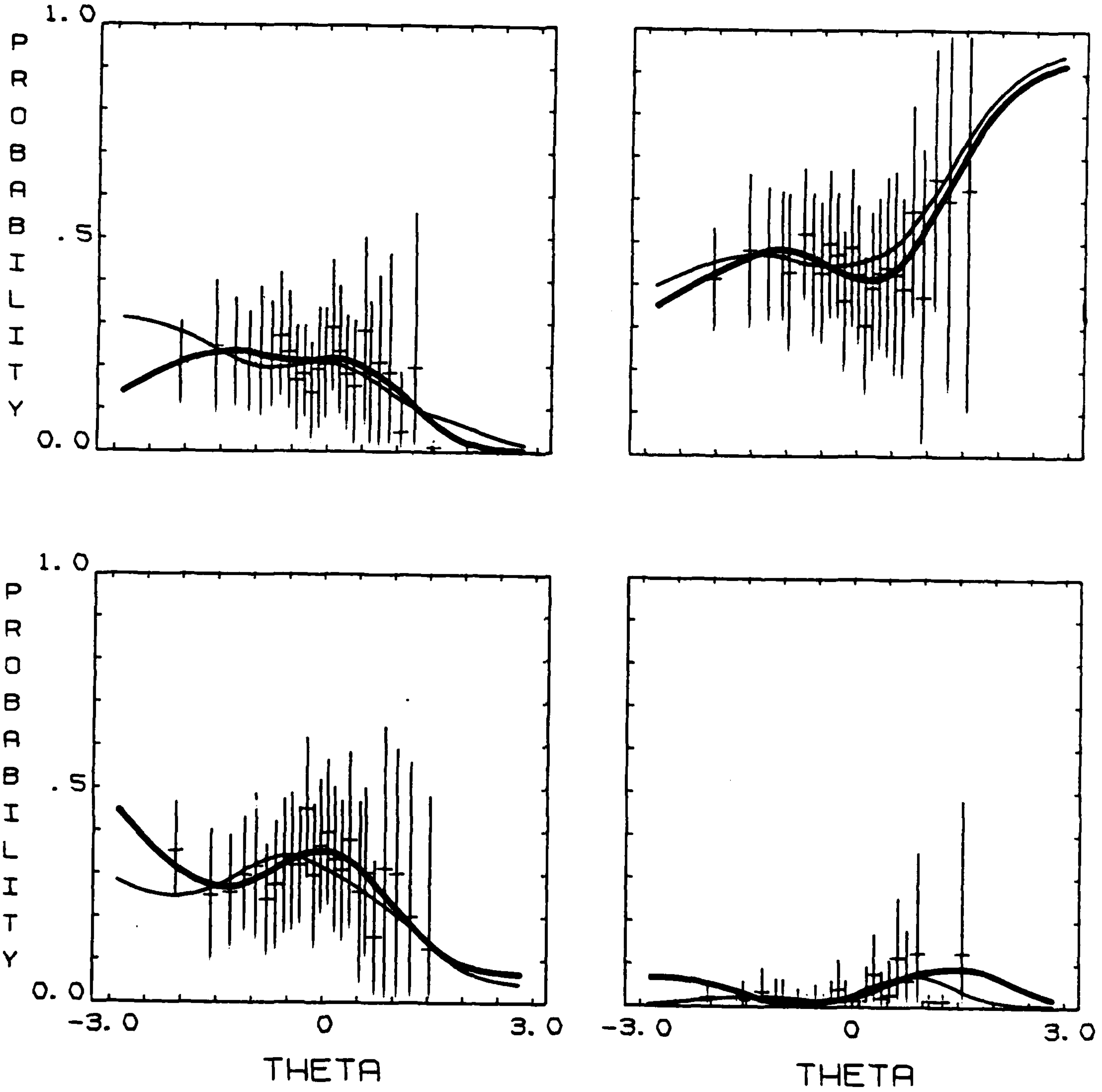


1.0
.5
0.0
PROBABILITY
1.0
.5
0.0
PROBABILITY
-3.0
0
3.0
THETA
-3.0
0
3.0
THETA

ITEM 15

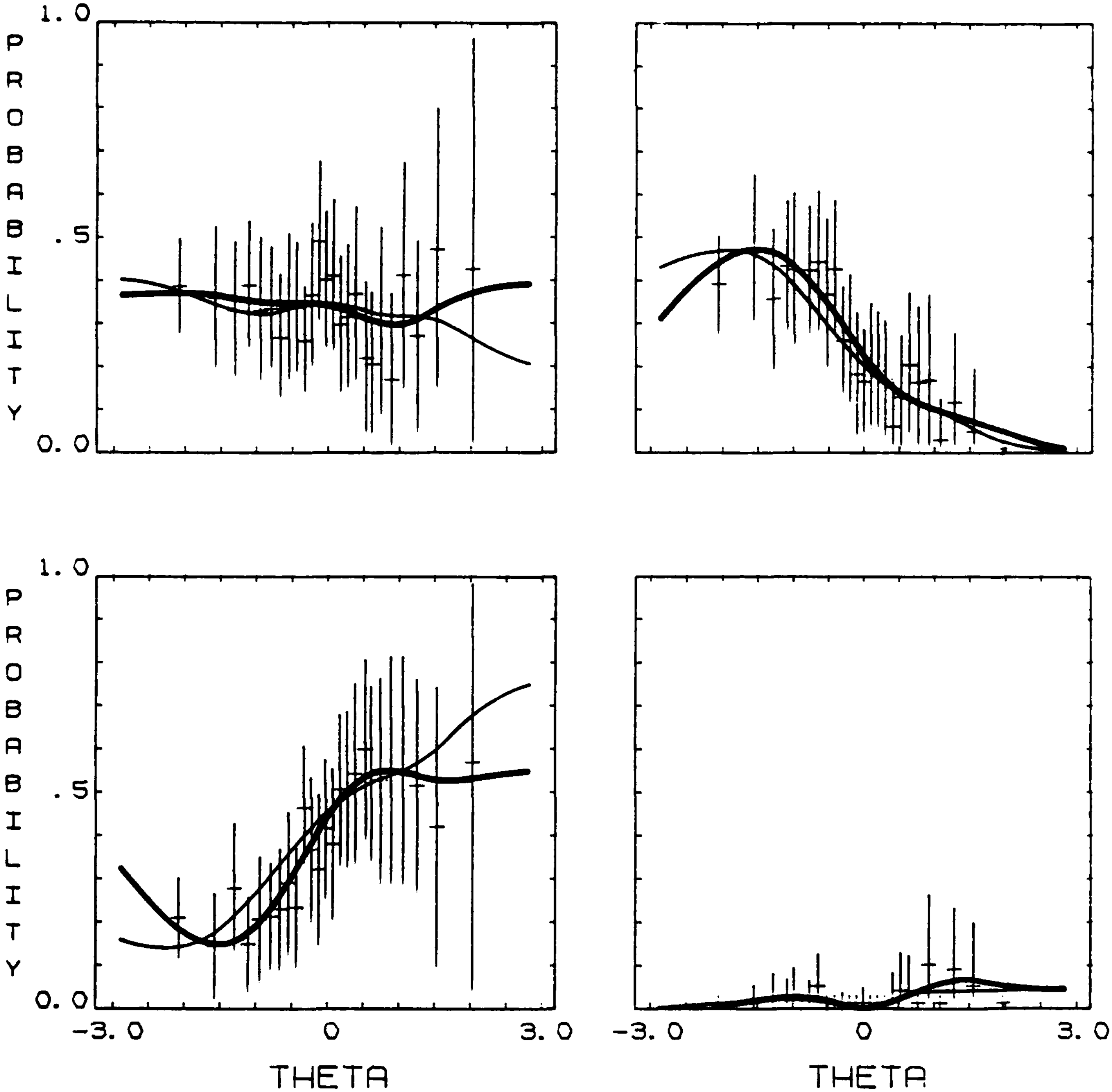


ITEM 16

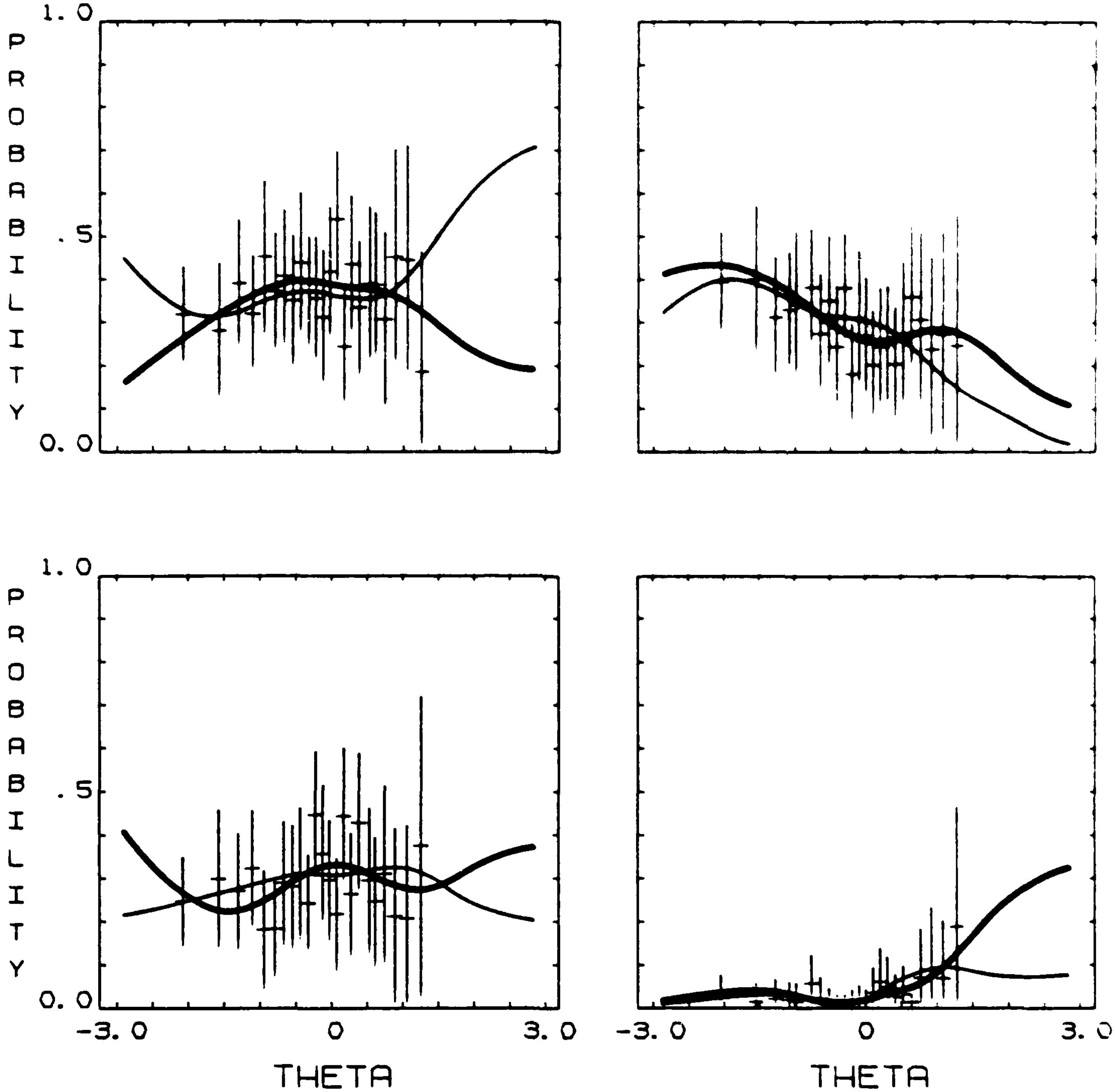


PROBABILITY
1.0
.5
0.0
PROBABILITY
1.0
.5
0.0
-3.0
0
3.0
THETA
-3.0
0
3.0
THETA

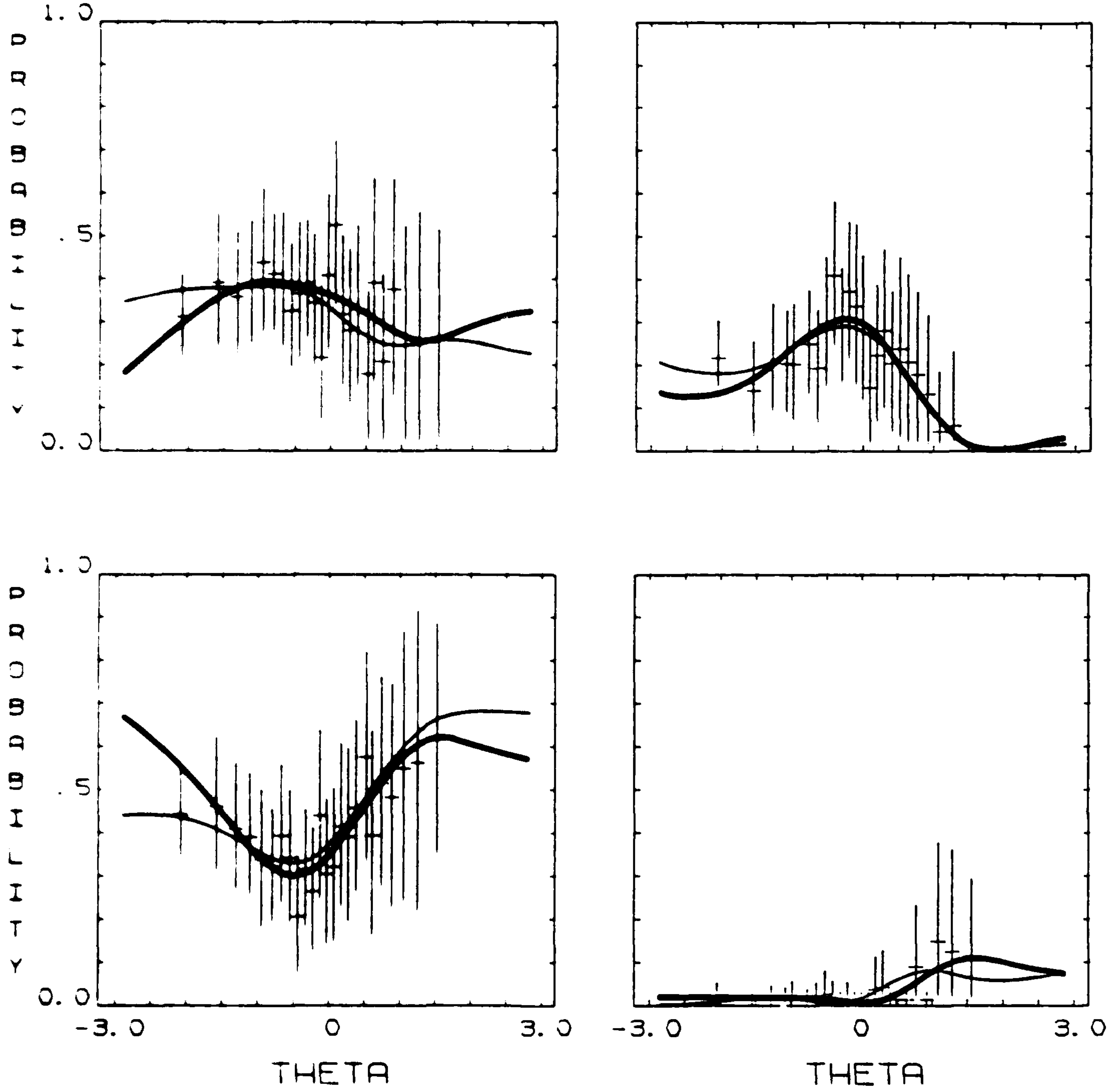
ITEM 17



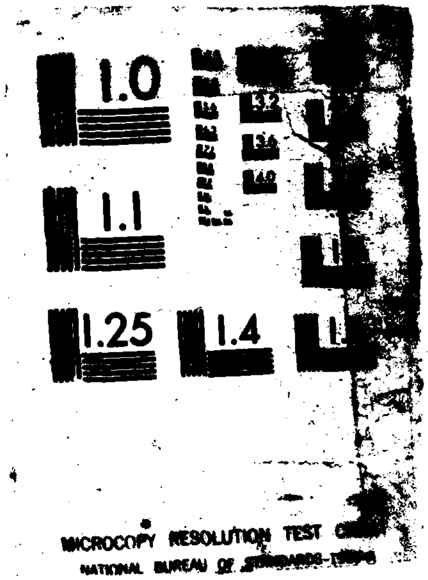
ITEM 18



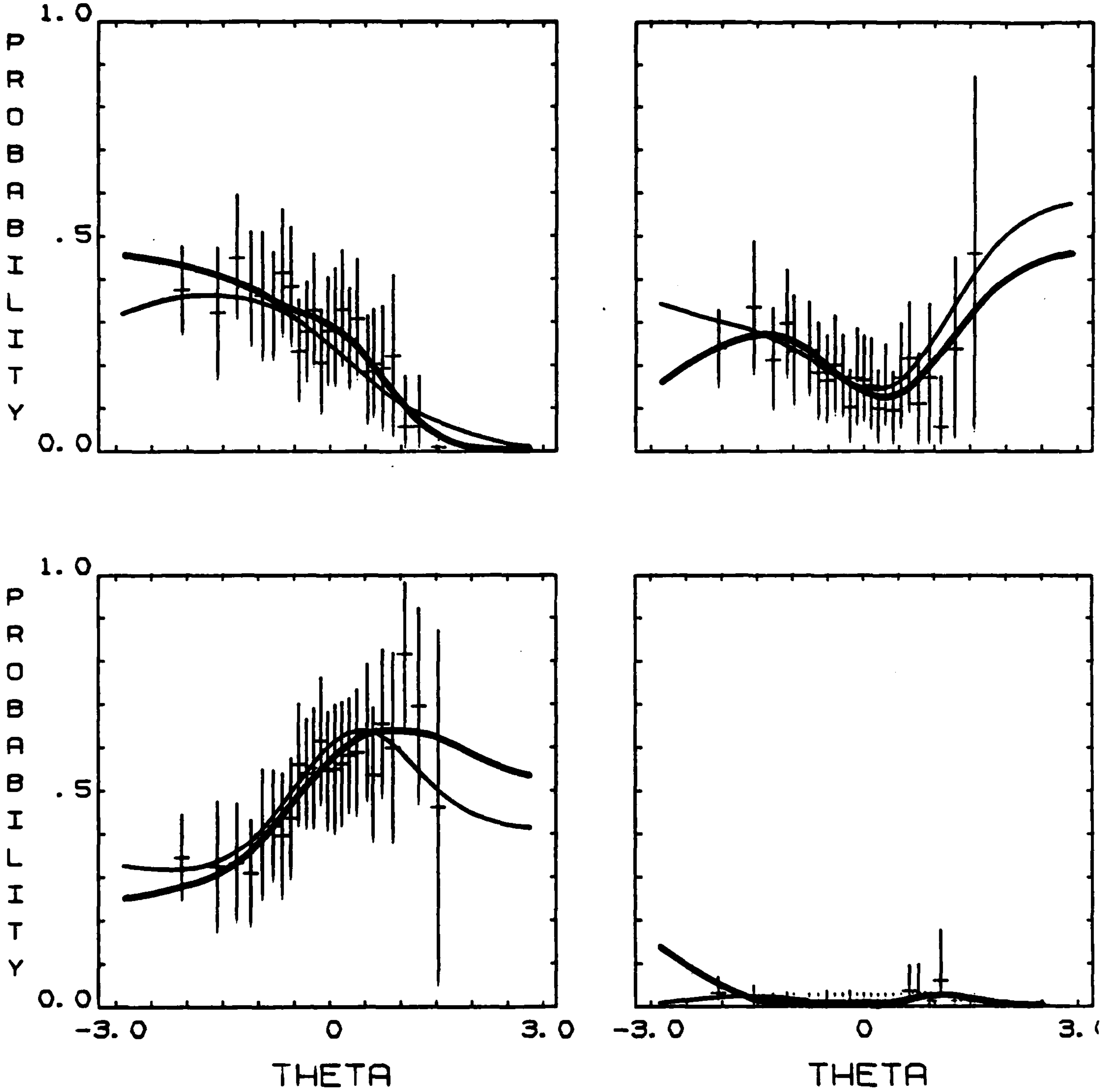
ITEM 19



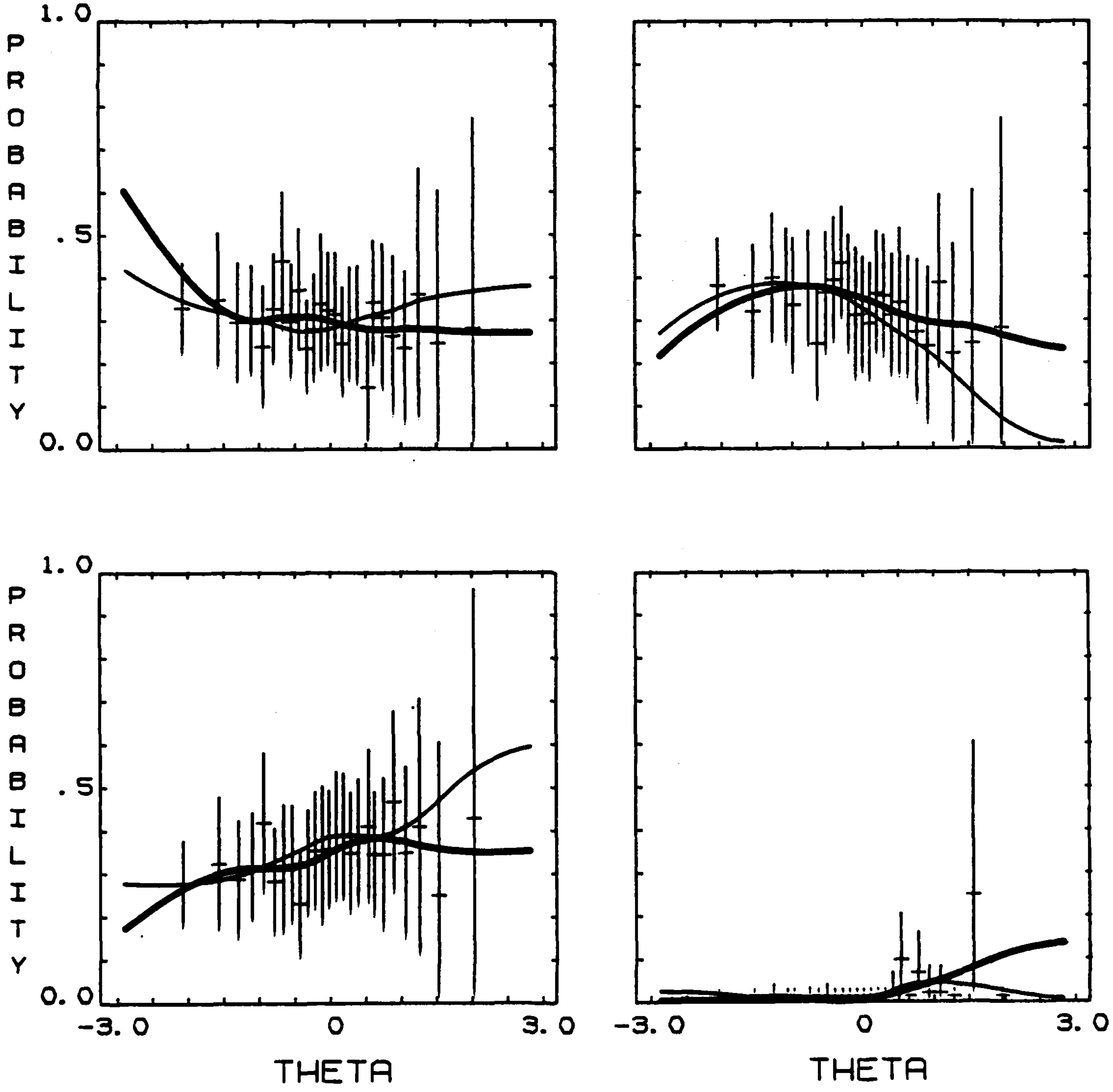
AD-A187 887   MODELING INCORRECT RESPONSES TO MULTIPLE-CHOICE ITEMS   2/2
WITH MULTILINEAR FO (U) ILLINOIS UNIV AT URBANA MODEL
BASED MEASUREMENT LAB F DRASGOW ET AL AUG 87
UNCLASSIFIED   MEASUREMENT SER-87-1 N00014-83-K-0397   F/G 5/8   NL


1.0
1.1
1.25
1.4
MICROCOPY RESOLUTION TEST
NATIONAL BUREAU OF

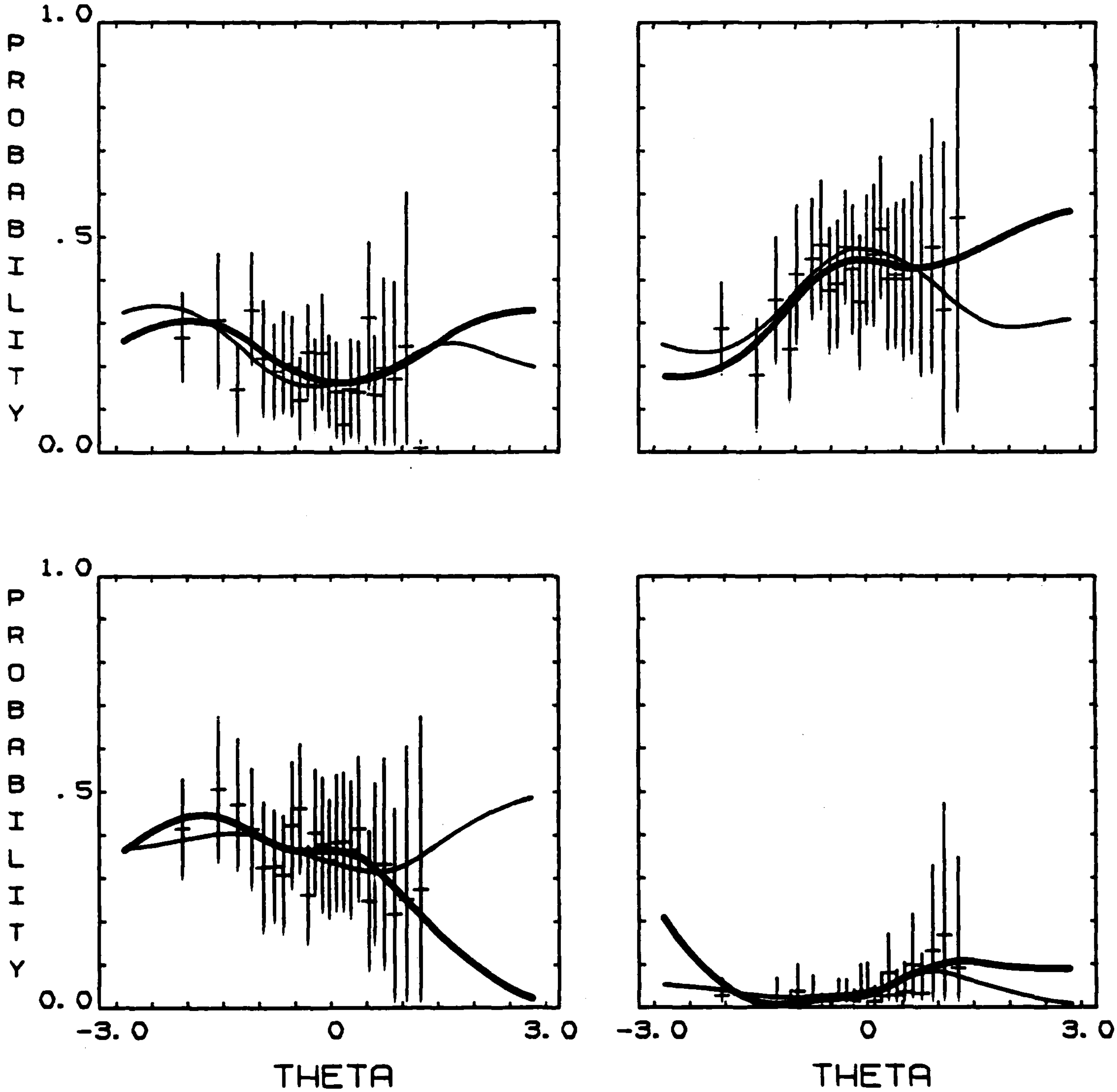
ITEM 20



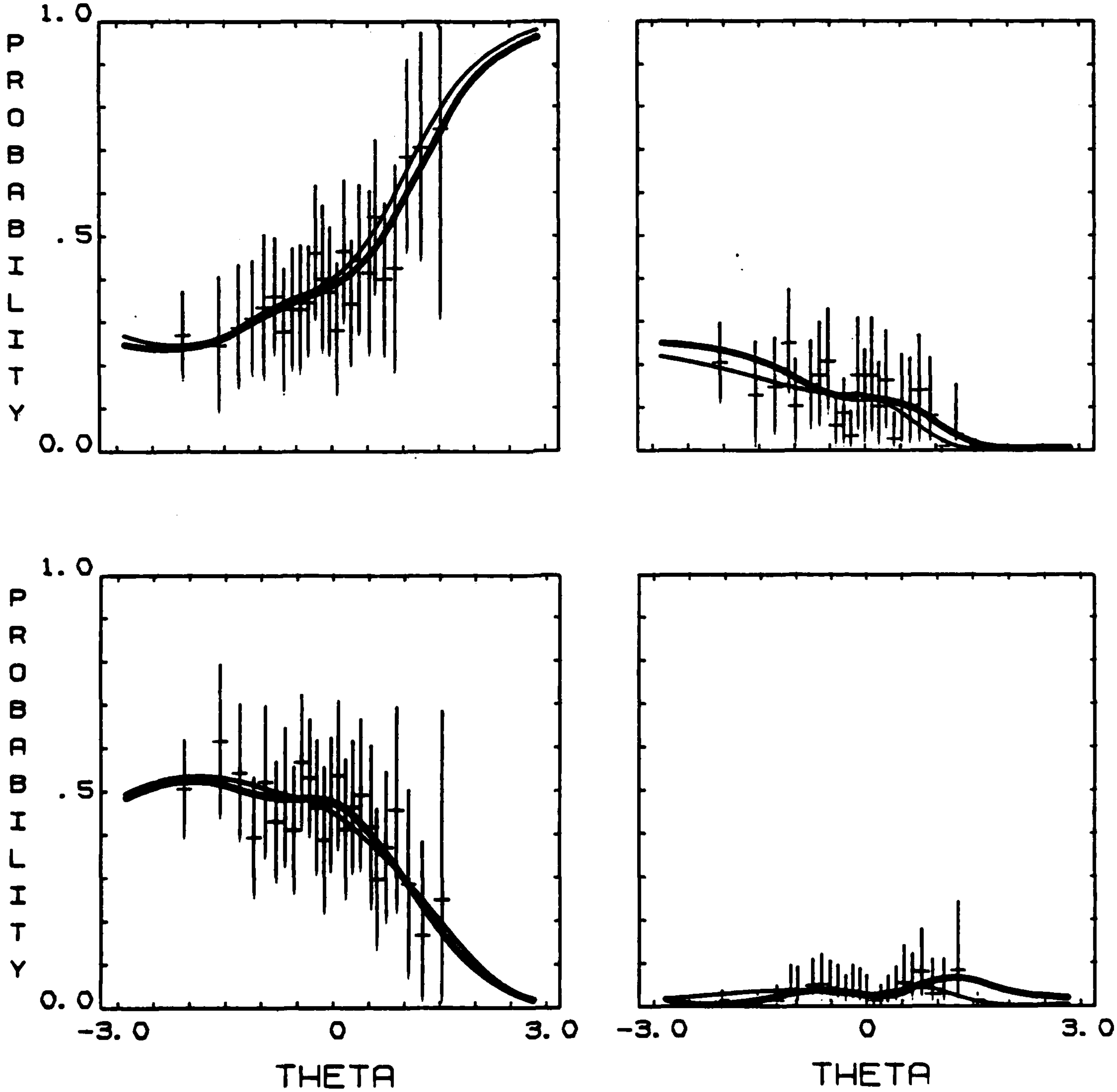
ITEM 21


1.0
PROBABILITY
.5
0.0
1.0
PROBABILITY
.5
0.0
-3.0
0
3.0
THETA
-3.0
0
3.0
THETA

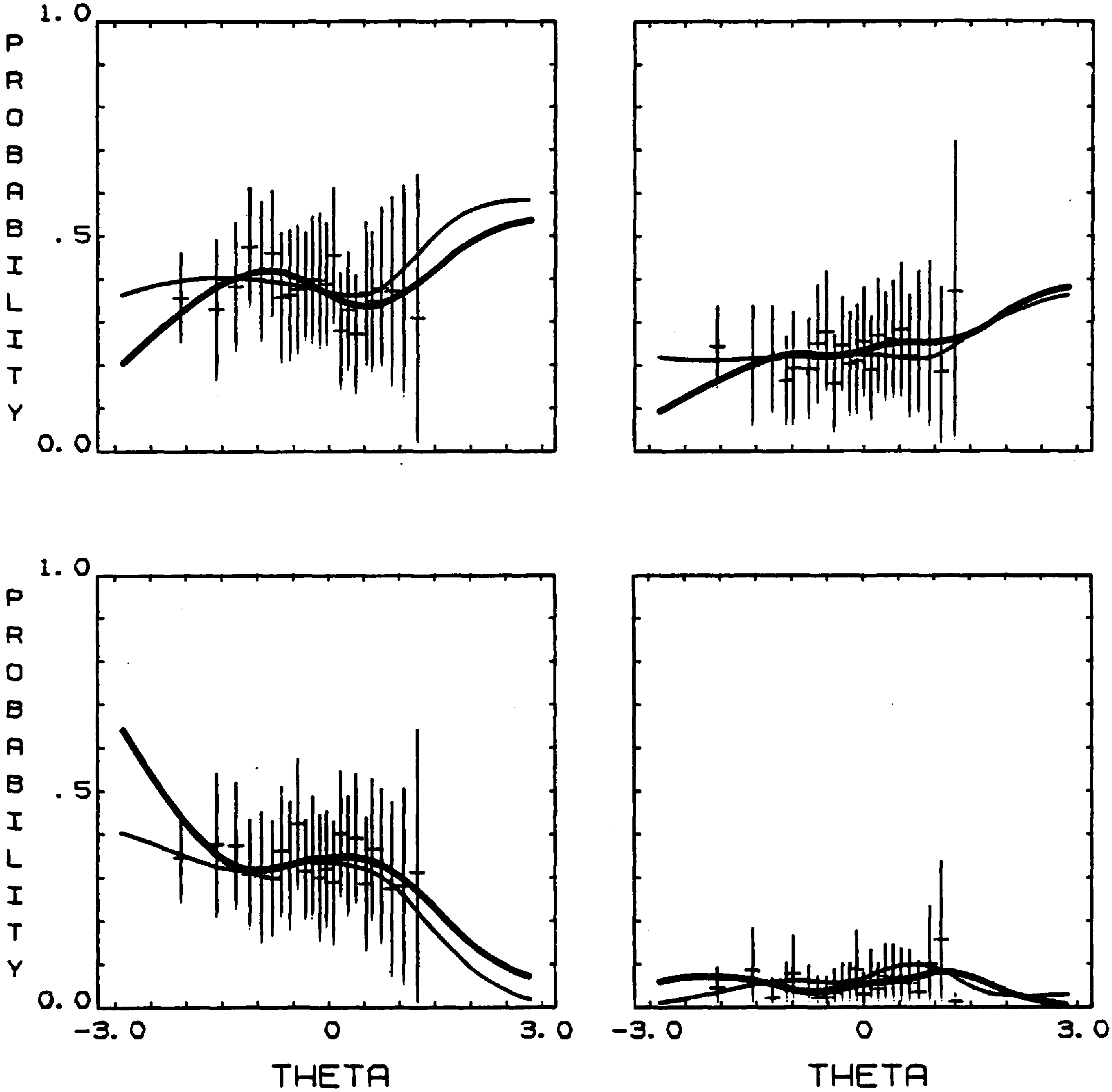
ITEM 22



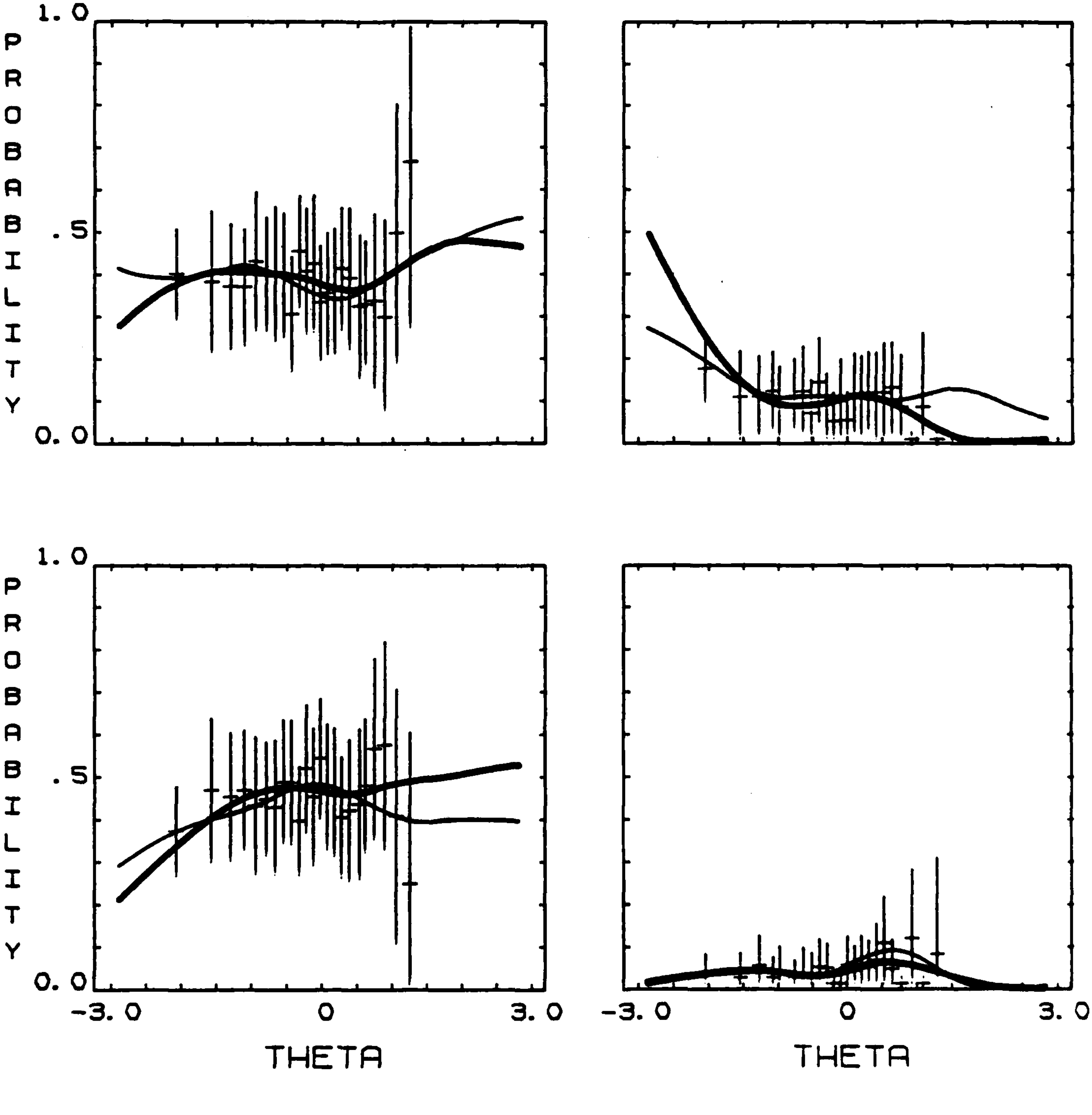
ITEM 23


1.0
.5
0.0
PROBABILITY
1.0
.5
0.0
PROBABILITY
-3.0
0
3.0
THETA
-3.0
0
3.0
THETA

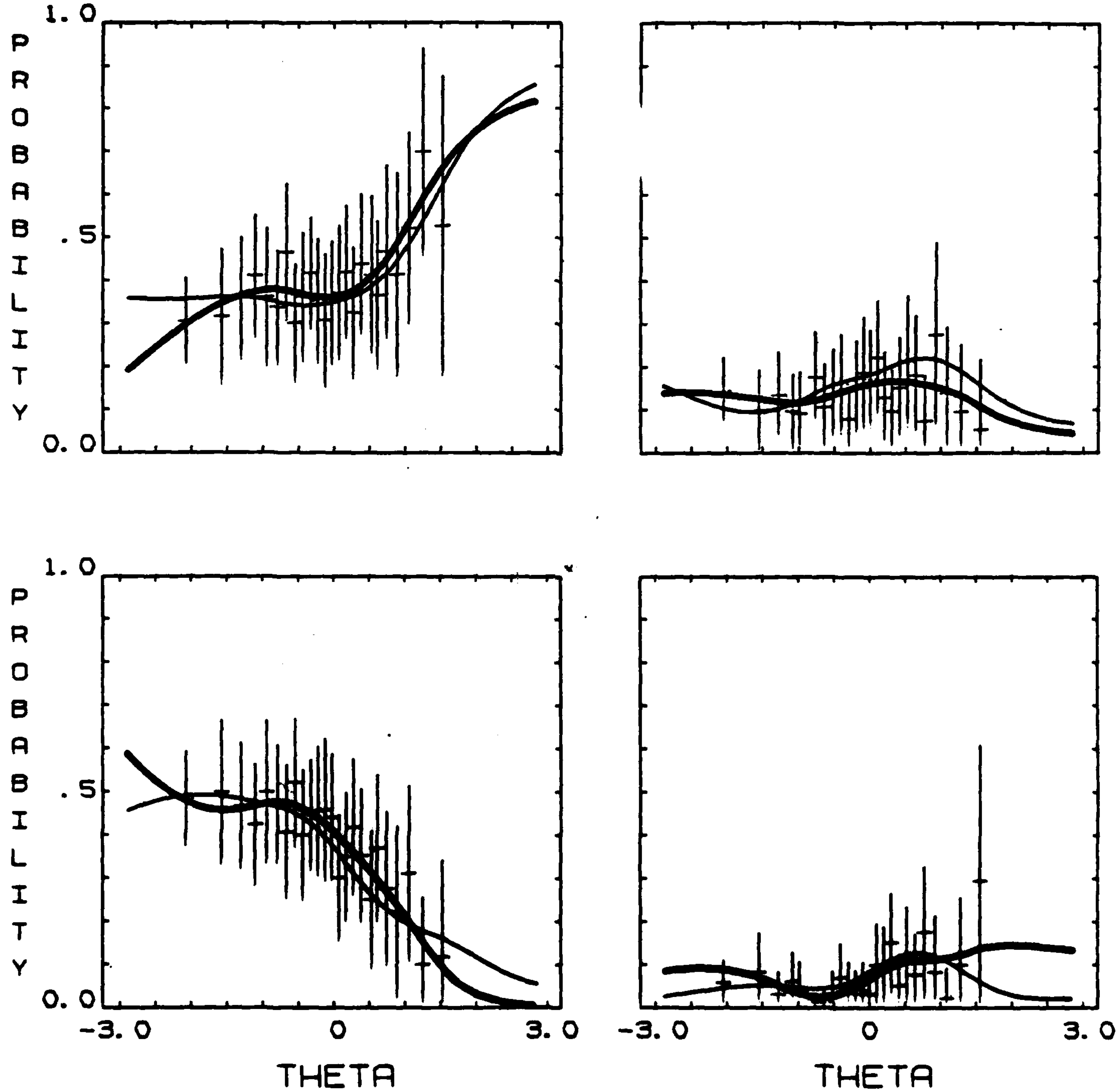
ITEM 24



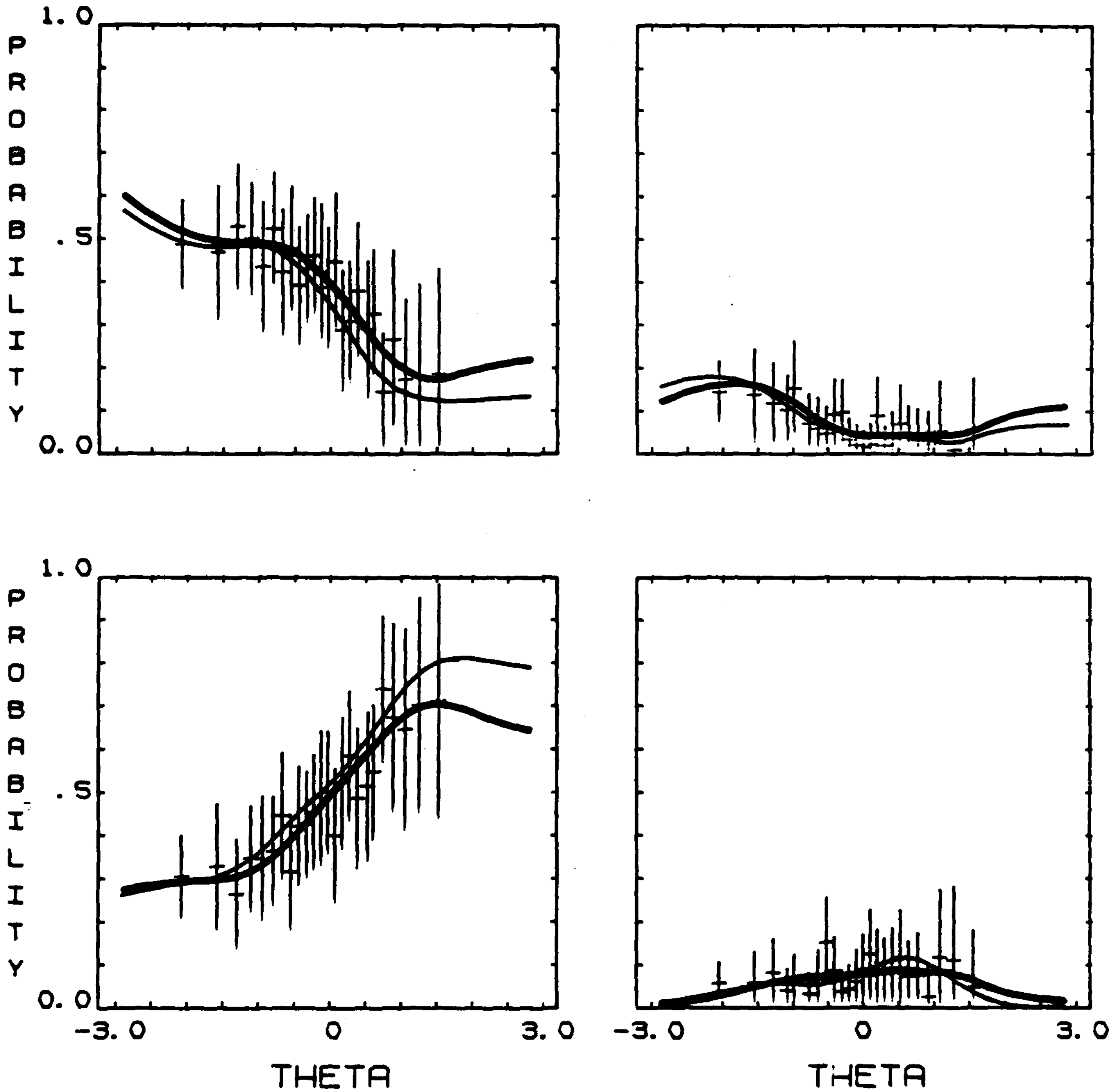
ITEM 25



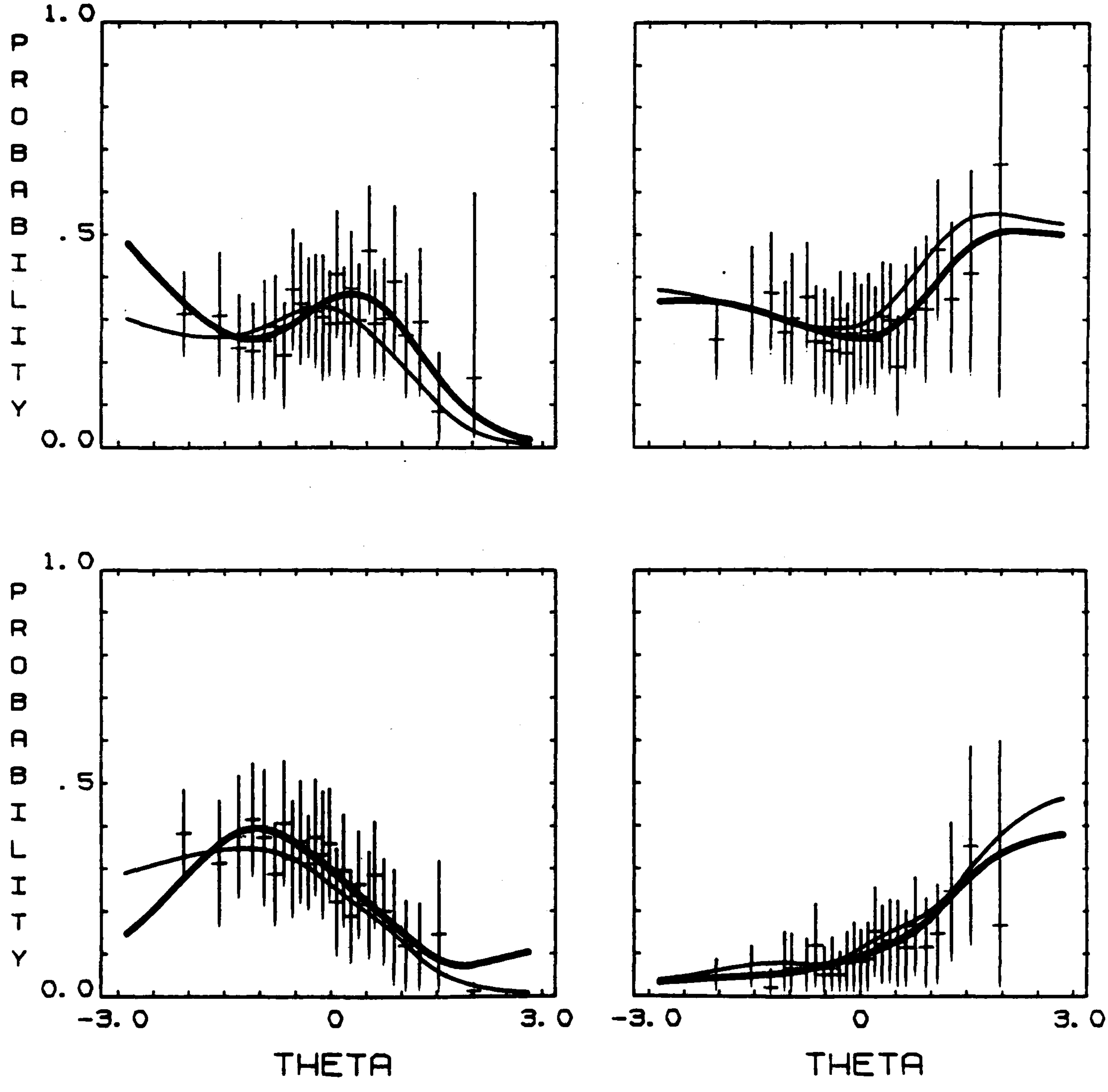
ITEM 26



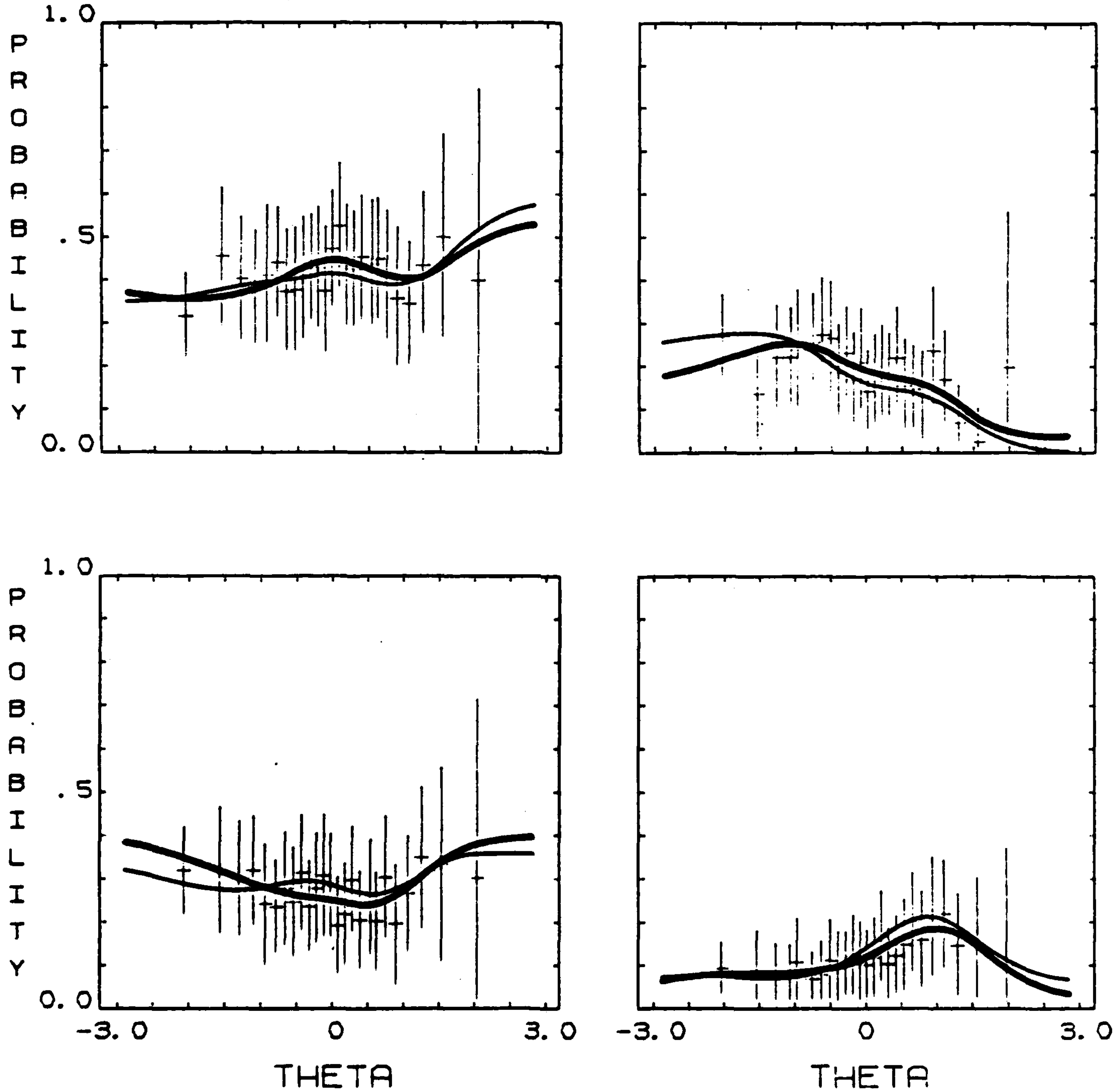
ITEM 27



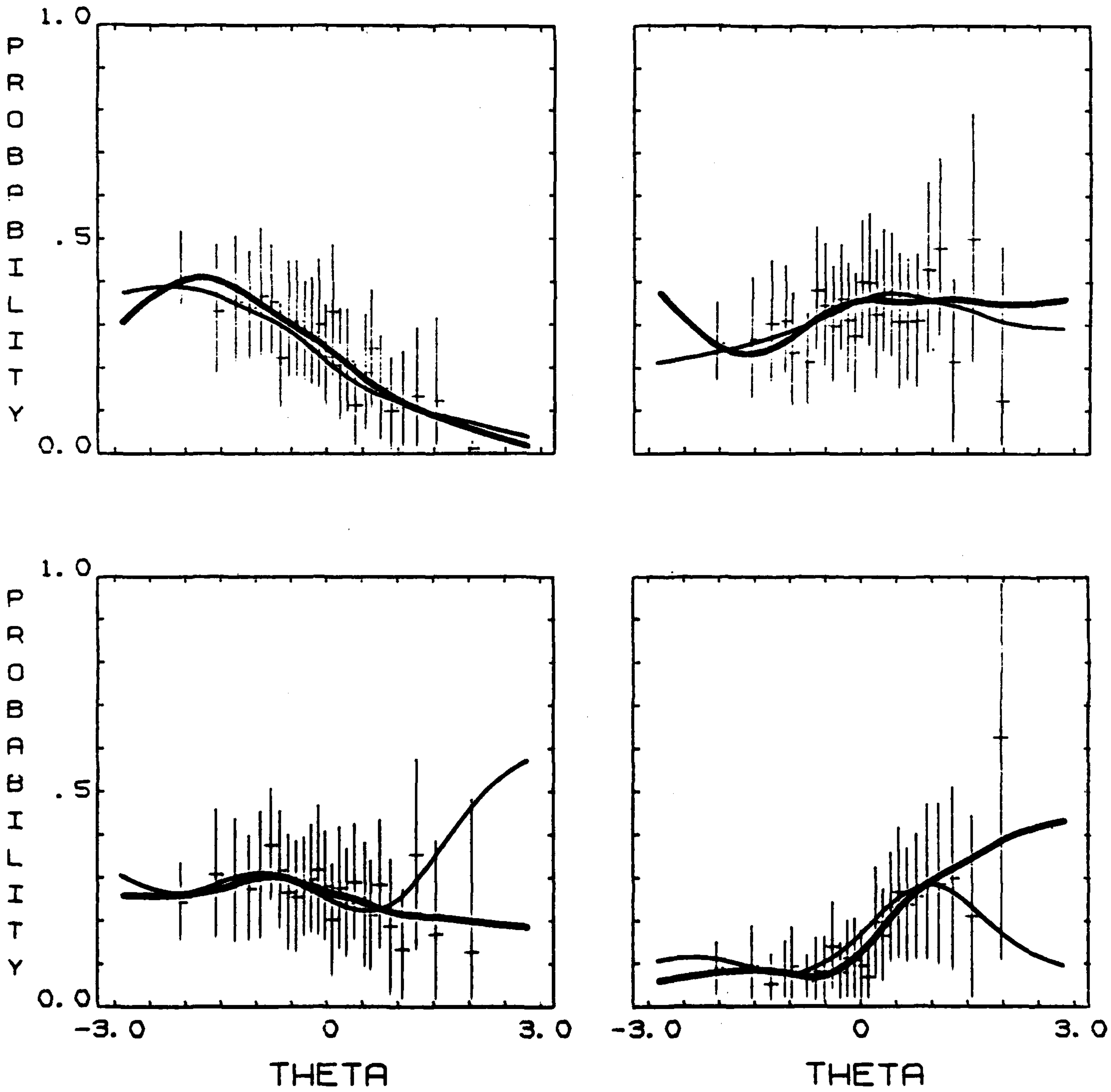
ITEM 28



ITEM 29



ITEM 30



Dr. Terry Ackerman
American College Testing Programs
P.O. Box 168
Iowa City, IA 52243

Dr. Robert Ahlers
Code N711
Human Factors Laboratory
Naval Training Systems Center
Orlando, FL 32813

Dr. James Algina
University of Florida
Gainesville, FL 32605

Dr. Erling B. Andersen
Department of Statistics
Studiestraede 6
1455 Copenhagen
DENMARK

Dr. Eva L. Baker
UCLA Center for the Study
of Evaluation
145 Moore Hall
University of California
Los Angeles, CA 90024

Dr. Isaac Bejar
Educational Testing Service
Princeton, NJ 08450

Dr. Menucha Birenbaum
School of Education
Tel Aviv University
Tel Aviv, Ramat Aviv 69978
ISRAEL

Dr. Arthur S. Blaiwes
Code N711
Naval Training Systems Center
Orlando, FL 32813

Dr. Bruce Bloxom
Defense Manpower Data Center
550 Camino El Estero,
Suite 200
Monterey, CA 93943-3231

Dr. R. Darrell Bock
University of Chicago
NORC
6030 South Ellis
Chicago, IL 60637

Cdt. Arnold Bohrer
Sectie Psychologisch Onderzoek
Rekruterings-En Selectiecentrum
Kwartier Koningen Astrid
Bruijnstraat
1120 Brussels, BELGIUM

Dr. Robert Breaux
Code N-095R
Naval Training Systems Center
Orlando, FL 32813

Dr. Robert Brennan
American College Testing
Programs
P. O. Box 168
Iowa City, IA 52243

Dr. Lyle D. Broemeling
ONR Code 1111SP
800 North Quincy Street
Arlington, VA 22217

Mr. James W. Carey
Commandant (G-PTE)
U.S. Coast Guard
2100 Second Street, S.W.
Washington, DC 20593

Dr. James Carlson
American College Testing
Program
P.O. Box 168
Iowa City, IA 52243

Dr. John B. Carroll
409 Elliott Rd.
Chapel Hill, NC 27514

Dr. Robert Carroll
OP 01B7
Washington, DC 20370

Mr. Raymond E. Christal
AFHRL/MOE
Brooks AFB, TX 78235

Dr. Norman Cliff
Department of Psychology
Univ. of So. California
University Park
Los Angeles, CA 90007

Director,
Manpower Support and
Readiness Program
Center for Naval Analysis
2000 North Beauregard Street
Alexandria, VA 22311

Dr. Stanley Collyer
Office of Naval Technology
Code 222
800 N. Quincy Street
Arlington, VA 22217-5000

Dr. Hans Crombag
University of Leyden
Education Research Center
Boerhaavelaan 2
2334 EN Leyden
The NETHERLANDS

Mr. Timothy Davey
University of Illinois
Educational Psychology
Urbana, IL 61801

Dr. Dattprasad Divgi
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Dr. Hei-Ki Dong
Bell Communications Research
6 Corporate Place
PYA-1k226
Piscataway, NJ 08854

Dr. Fritz Drasgow
University of Illinois
Department of Psychology
603 E. Daniel St.
Champaign, IL 61820

Defense Technical
Information Center
Cameron Station, Bldg 5
Alexandria, VA 22314
Attn: TC
(12 Copies)

Dr. Stephen Dunbar
Lindquist Center
for Measurement
University of Iowa
Iowa City, IA 52242

Dr. James A. Earles
Air Force Human Resources Lab
Brooks AFB, TX 78235

Dr. Kent Eaton
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. John M. Eddins
University of Illinois
252 Engineering Research
Laboratory
103 South Mathews Street
Urbana, IL 61801

Dr. Susan Embretson
University of Kansas
Psychology Department
426 Fraser
Lawrence, KS 66045

Dr. George Englehard, Jr.
Division of Educational Studies
Emory University
201 Fishburne Bldg.
Atlanta, GA 30322

Dr. Benjamin A. Fairbank
Performance Metrics, Inc.
5825 Callaghan
Suite 225
San Antonio, TX 78228

Dr. Pat Federico
Code 511
NPRDC
San Diego, CA 92152-6800

Dr. Leonard Feldt
Lindquist Center
for Measurement
University of Iowa
Iowa City, IA 52242

Dr. Richard L. Ferguson
American College Testing
Program
P.O. Box 168
Iowa City, IA 52240

Dr. Gerhard Fischer
Liebiggasse 5/3
A 1010 Vienna
AUSTRIA

Dr. Myron Fischl
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Prof. Donald Fitzgerald
University of New England
Department of Psychology
Armidale, New South Wales 2351
AUSTRALIA

Mr. Paul Foley
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Alfred R. Fregly
AFOSR/NL
Bolling AFB, DC 20332

Dr. Robert D. Gibbons
Illinois State Psychiatric Inst.
Rm 529W
1601 W. Taylor Street
Chicago, IL 60612

Dr. Janice Gifford
University of Massachusetts
School of Education
Amherst, MA 01003

Dr. Robert Glaser
Learning Research
& Development Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15260

Dr. Bert Green
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore, MD 21218

Dipl. Pad. Michael W. Habon
Universitat Dusseldorf
Erziehungswissenschaftliches
Universitatsstr. 1
D-4000 Dusseldorf 1
WEST GERMANY

Dr. Ronald K. Hambleton
Prof. of Education & Psychology
University of Massachusetts
at Amherst
Hills House
Amherst, MA 01003

Dr. Delwyn Harnisch
University of Illinois
51 Gerty Drive
Champaign, IL 61820

Ms. Rebecca Hetter
Navy Personnel R&D Center
Code 62
San Diego, CA 92152-6800

Dr. Paul W. Holland
Educational Testing Service
Rosedale Road
Princeton, NJ 08541

Prof. Lutz F. Hornke
Institut fur Psychologie
RWTH Aachen
Jaegerstrasse 17/19
D-5100 Aachen
WEST GERMANY

Dr. Paul Horst
677 G Street, #184
Chula Vista, CA 90010

Mr. Dick Hoshaw
OP-135
Arlington Annex
Room 2834
Washington, DC 20350

Dr. Lloyd Humphreys
University of Illinois
Department of Psychology
603 East Daniel Street
Champaign, IL 61820

Dr. Steven Hunka
Department of Education
University of Alberta
Edmonton, Alberta
CANADA

Dr. Huynh Huynh
College of Education
Univ. of South Carolina
Columbia, SC 29208

Dr. Robert Jannarone
Department of Psychology
University of South Carolina
Columbia, SC 29208

Dr. Dennis E. Jennings
Department of Statistics
University of Illinois
1409 West Green Street
Urbana, IL 61801

Dr. Douglas H. Jones
Thatcher Jones Associates
P.O. Box 6640
10 Trafalgar Court
Lawrenceville, NJ 08648

Dr. Milton S. Katz
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Prof. John A. Keats
Department of Psychology
University of Newcastle
N.S.W. 2308
AUSTRALIA

Dr. G. Gage Kingsbury
Portland Public Schools
Research and Evaluation Department
501 North Dixon Street
P. O. Box 3107
Portland, OR 97209-3107

Dr. William Koch
University of Texas-Austin
Measurement and Evaluation
Center
Austin, TX 78703

Dr. James Kraatz
Computer-based Education
Research Laboratory
University of Illinois
Urbana, IL 61801

Dr. Leonard Kroeker
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Daryll Lang
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Thomas Leonard
University of Wisconsin
Department of Statistics
1210 West Dayton Street
Madison, WI 53705

Dr. Michael Levine
Educational Psychology
210 Education Bldg.
University of Illinois
Champaign, IL 61801

Dr. Charles Lewis
Educational Testing Service
Princeton, NJ 08541

Dr. Robert Linn
College of Education
University of Illinois
Urbana, IL 61801

Dr. Robert Lockman
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Dr. Frederic M. Lord
Educational Testing Service
Princeton, NJ 08541

Dr. James Lumsden
Department of Psychology
University of Western Australia
Nedlands W.A. 6009
AUSTRALIA

Dr. Milton Maier
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Dr. William L. Maloy
Chief of Naval Education
and Training
Naval Air Station
Pensacola, FL 32508

Dr. Gary Marco
Stop 31-E
Educational Testing Service
Princeton, NJ 08451

Dr. Clessen Martin
Army Research Institute
5001 Eisenhower Blvd.
Alexandria, VA 22333

Dr. James McBride
Psychological Corporation
c/o Harcourt, Brace,
Javanovich Inc.
1250 West 6th Street
San Diego, CA 92101

Dr. Clarence McCormick
HQ, MEPCOM
MEPCT-P
2500 Green Bay Road
North Chicago, IL 60064

Dr. Robert McKinley
Educational Testing Service
20-P
Princeton, NJ 08541

Dr. James McMichael
Technical Director
Navy Personnel R&D Center
San Diego, CA 92152

Dr. Barbara Means
Human Resources
Research Organization
1100 South Washington
Alexandria, VA 22314

Dr. Robert Mislevy
Educational Testing Service
Princeton, NJ 08541

Dr. William Montague
NPRDC Code 13
San Diego, CA 92152-6800

Ms. Kathleen Moreno
Navy Personnel R&D Center
Code 62
San Diego, CA 92152-6800

Headquarters, Marine Corps
Code MPI-20
Washington, DC 20380

Dr. W. Alan Nicewander
University of Oklahoma
Department of Psychology
Oklahoma City, OK 73069

Deputy Technical Director
NPRDC Code 01A
San Diego, CA 92152-6800

Director, Training Laboratory,
NPRDC (Code 05)
San Diego, CA 92152-6800

Director, Manpower and Personnel
Laboratory,
NPRDC (Code 06)
San Diego, CA 92152-6800

Director, Human Factors
& Organizational Systems Lab,
NPRDC (Code 07)
San Diego, CA 92152-6800

Fleet Support Office,
NPRDC (Code 301)
San Diego, CA 92152-6800

Library, NPRDC
Code P201L
San Diego, CA 92152-6800

Commanding Officer,
Naval Research Laboratory
Code 2627
Washington, DC 20390

Dr. Harold F. O'Neil, Jr.
School of Education - WPH 801
Department of Educational
Psychology & Technology
University of Southern California
Los Angeles, CA 90089-0031

Dr. James Olson
WICAT, Inc.
1875 South State Street
Orem, UT 84057

Office of Naval Research,
Code 1142PT
800 N. Quincy Street
Arlington, VA 22217-5000
(6 Copies)

Office of Naval Research,
Code 125
800 N. Quincy Street
Arlington, VA 22217-5000

Assistant for MPT Research,
Development and Studies
OP 01B7
Washington, DC 20370

Dr. Judith Orasanu
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. Jesse Orlansky
Institute for Defense Analyses
1801 N. Beauregard St.
Alexandria, VA 22311

Dr. Randolph Park
Army Research Institute
5001 Eisenhower Blvd.
Alexandria, VA 22333

Wayne M. Patience
American Council on Education
GED Testing Service, Suite 20
One Dupont Circle, NW
Washington, DC 20036

Dr. James Paulson
Department of Psychology
Portland State University
P.O. Box 751
Portland, OR 97207

Administrative Sciences Department,
Naval Postgraduate School
Monterey, CA 93940

Department of Operations Research,
Naval Postgraduate School
Monterey, CA 93940

Dr. Mark D. Reckase
ACT
P. O. Box 168
Iowa City, IA 52243

Dr. Malcolm Ree
AFHRL/MP
Brooks AFB, TX 78235

Dr. Barry Riegelhaupt
HumRRO
1100 South Washington Street
Alexandria, VA 22314

Dr. Carl Ross
CNET-PDCD
Building 90
Great Lakes NTC, IL 60088

Dr. J. Ryan
Department of Education
University of South Carolina
Columbia, SC 29208

Dr. Fumiko Samejima
Department of Psychology
University of Tennessee
3108 AustinPeay Bldg.
Knoxville, TN 37916-0900

Mr. Drew Sands
NPRDC Code 62
San Diego, CA 92152-6800

Lowell Schoer
Psychological & Quantitative
Foundations
College of Education
University of Iowa
Iowa City, IA 52242

Dr. Mary Schratz
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Dan Segall
Navy Personnel R&D Center
San Diego, CA 92152

Dr. W. Steve Sellman
OASD(MRA&L)
2B269 The Pentagon
Washington, DC 20301

Dr. Kazuo Shigemasu
7-9-24 Kugenuma-Kaigan
Fujusawa 251
JAPAN

Dr. William Sims
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Dr. H. Wallace Sinaiko
Manpower Research
and Advisory Services
Smithsonian Institution
801 North Pitt Street
Alexandria, VA 22314

Dr. Richard E. Snow
Department of Psychology
Stanford University
Stanford, CA 94306

Dr. Richard Sorensen
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Paul Speckman
University of Missouri
Department of Statistics
Columbia, MO 65201

Dr. Judy Spray
ACT
P.O. Box 168
Iowa City, IA 52243

Dr. Martha Stocking
Educational Testing Service
Princeton, NJ 08541

Dr. Peter Stoloff
Center for Naval Analysis
200 North Beauregard Street
Alexandria, VA 22311

Dr. William Stout
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright St.
Champaign, IL 61820

Maj. Bill Strickland
AF/MPXOA
4E168 Pentagon
Washington, DC 20330

Dr. Hariharan Swaminathan
Laboratory of Psychometric and
Evaluation Research
School of Education
University of Massachusetts
Amherst, MA 01003

Mr. Brad Sympson
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. John Tangney
AFOSR/NL
Bolling AFB, DC 20332

Dr. Kikumi Tatsuoka
CERL
252 Engineering Research
Laboratory
Urbana, IL 61801

Dr. Maurice Tatsuoka
220 Education Bldg
1310 S. Sixth St.
Champaign, IL 61820

Dr. David Thissen
Department of Psychology
University of Kansas
Lawrence, KS 66044

Mr. Gary Thomasson
University of Illinois
Educational Psychology
Champaign, IL 61820

Dr. Robert Tsutakawa
University of Missouri
Department of Statistics
222 Math. Sciences Bldg.
Columbia, MO 65211

Dr. Ledyard Tucker
University of Illinois
Department of Psychology
603 E. Daniel Street
Champaign, IL 61820

Dr. Vern W. Urry
Personnel R&D Center
Office of Personnel Management
1900 E. Street, NW
Washington, DC 20415

Dr. David Vale
Assessment Systems Corp.
2233 University Avenue
Suite 310
St. Paul, MN 55114

Dr. Frank Vicino
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Howard Wainer
Division of Psychological Studies
Educational Testing Service
Princeton, NJ 08541

Dr. Ming-Mei Wang
Lindquist Center
for Measurement
University of Iowa
Iowa City, IA 52242

Dr. Thomas A. Warm
Coast Guard Institute
P. O. Substation 18
Oklahoma City, OK 73169

Dr. Brian Waters
Program Manager
Manpower Analysis Program
HumRRO
1100 S. Washington St.
Alexandria, VA 22314

Dr. David J. Weiss
N660 Elliott Hall
University of Minnesota
75 E. River Road
Minneapolis, MN 55455

Dr. Ronald A. Weitzman
NPS, Code 54Wz
Monterey, CA 92152-6800

Major John Welsh
AFHRL/MOAN
Brooks AFB, TX 78223

Dr. Douglas Wetzel
Code 12
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Rand R. Wilcox
University of Southern
California
Department of Psychology
Los Angeles, CA 90007

German Military Representative
ATTN: Wolfgang Wildegrube
Streitkraefteamt
D-5300 Bonn 2
4000 Brandywine Street, NW
Washington, DC 20016

Dr. Bruce Williams
Department of Educational
Psychology
University of Illinois
Urbana, IL 61801

Dr. Hilda Wing
Psychological Corporation
c/o Harcourt, Brace,
Javanovich Inc.
1250 West 6th Street
San Diego, CA 92101

Dr. Martin F. Wiskoff
Navy Personnel R & D Center
San Diego, CA 92152-6800

Mr. John H. Wolfe
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. George Wong
Biostatistics Laboratory
Memorial Sloan-Kettering
Cancer Center
1275 York Avenue
New York, NY 10021

Dr. Wallace Wulfeck, III
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Kentaro Yamamoto
Computer-based Education
Research Laboratory
University of Illinois
Urbana, IL 61801

Dr. Wendy Yen
CTB/McGraw Hill
Del Monte Research Park
Monterey, CA 93940

Dr. Joseph L. Young
Memory & Cognitive
Processes
National Science Foundation
Washington, DC 20550

Dr. Anthony R. Zara
National Council of State
Boards of Nursing, Inc.
625 North Michigan Ave.
Suite 1544
Chicago, IL 60611

END
FEb.
1988
DTIC