JOURNAL OF MATHEMATICAL PSYCHOLOGY 29, 443–473 (1985)

How Many Parameters Can a Model Have and Still Be Testable?

DONALD BAMBER*

VA Medical Center, St. Cloud, Minnesota

AND

JAN P. H. VAN SANTEN†

New York University

A standard rule of thumb states that a model has too many parameters to be testable if and only if it has at least as many parameters as empirically observable quantities. We argue that when one asks whether a model has too many parameters to be testable, one implicitly refers to a particular type of testability, which we call *quantitative testability*. A model is defined to be quantitatively testable if the model's predictions have zero probability of being correct by chance. Next, we propose a new rule of thumb, based on the rank of the Jacobian matrix of a model (i.e., the matrix of partial derivatives of the function that maps the model's parameter values onto predicted experimental outcomes). According to this rule, a model is quantitatively testable if and only if the rank of the Jacobian matrix is less than the number of observables. (The rank of his matrix can be found with standard computer algorithms.) Using Sard's theorem, we prove that the proposed new rule of thumb is correct provided that certain "smoothness" conditions are satisfied. We also discuss the relation between quantitative testability and reparameterization, identifiability, and goodness-of-fit testing.

I. INTRODUCTION

During model construction, often the issue comes up whether the model has "too many parameters to be testable."[1] This is an important question, because an incorrect assessment of a model's testability may result in the abandonment of a sound model. This can happen in either of two ways. If it is concluded that a model has "too many parameters to be testable," one may feel that the model is worthless and abandon it immediately. Alternatively, one may attempt to reduce the number

* Present address: Human Factors Department-Code 71, Navy Personnel Research and Development Center, San Diego, CA 92152.

† Present address: AT&T Bell Laboratories, Summit, NJ 07901.

[1] One can also ask whether a model has "too many parameters" to be possible to estimate. We discuss this question of *parameter identifiability* in Section VI.

0022-2496/85 $3.00

of parameters by adding assumptions to the model. For example, certain parameters may be set equal to each other while other parameters may be set equal to zero. When this modified model is empirically tested, it may be incompatible with the data because of these additional assumptions, even though the original model might have been compatible.

Thus, assessment of a model's testability is a preliminary screening device for judging the model's usefulness. It does not provide everything one wants to know about a model. In particular, knowing that a model is testable does not tell one how to test it. However, knowing whether a model is testable does help one decide whether to work more at figuring out how to test it. Thus, assessment of a model's testability helps one decide whether to (a) continue working with the model, (b) modify the model, or (c) abandon it. In addition, the particular type of testability that we will define is a prerequisite for commonly used goodness-of-fit tests.

When one says that a model has "too many parameters to be testable," one is implicitly referring to a particular type of testability, which we shall call *quantitative testability*. In this paper, we present a new method for assessing whether or not a model is quantitatively testable. Moreover, we show that the commonly used assessment method based upon counting parameters and observations is invalid.

Quantitative Testability

We give the reader here an *intuitive feel* for the concept of quantitative testability. A *precise definition* is presented in Section II. Consider an experiment and a model that applies to the experiment. The experiment measures the values of n quantities. Consider the point in Euclidean n-space, R^n, whose coordinates are these n quantities' *true* values (which may differ from their *measured* values). This point is called the *true outcome of the experiment*. Any point in R^n that could logically be the true outcome of the experiment is called a *potential outcome*. The set of all potential outcomes of an experiment is called the *outcome space* and is denoted $\mathbf{O}$. The set of all potential outcomes that are consistent with the model is called the *prediction range* and is denoted $\mathbf{P}$. Thus, $\mathbf{P} \subset \mathbf{O} \subset R^n$.

The defining condition for testability (i.e., falsifiability) is that, in a given experiment, there exist potential outcomes that are inconsistent with the model. Formally, there must exist some $y \in \mathbf{O} - \mathbf{P}$.

We distinguish between two degrees of testability, *qualitative* and *quantitative*, that may be loosely differentiated as follows. Qualitatively testable models make *inexact* predictions, whereas quantitatively testable models make *exact* predictions. Put otherwise: If the model's prediction range $\mathbf{P}$ is defined exclusively by inequalities (that do not imply equalities), then the model is qualitatively testable. If $\mathbf{P}$ involves at least one equality, then the model is quantitatively testable. For example, suppose three observations y_1, y_2, and y_3 are made in an experiment. If the model predicts that $y_3 = y_1 + y_2$, then it is quantitatively testable; however, if the model predicts that $y_3 \leqslant y_1 + y_2$, then it is qualitatively testable.

A further way to differentiate between qualitatively and quantitatively testable models is by looking at the "volume" of the model's prediction range $\mathbf{P}$. If $\mathbf{P}$ is a

solid portion of **O**, that fills "volume," then the model is qualitatively testable. However, if **P** is an infinitesimally thin film in **O** that has zero "volume," then the model is quantitatively testable.

A final way to differentiate between the two is by probability statements. Suppose Y is a random vector that can be characterized by a probability density on **O**. If $P(Y \in \mathbf{P}) = 0$, then the model is quantitatively testable. If $P(Y \in \mathbf{P}) > 0$, then the model is qualitatively testable. In other words, suppose one performs an experiment with outcome space **O** to test a model with prediction range **P**. A computer is programmed to run the experiment and report the outcome, but due to a programming error, the computer reports a randomly selected point in **O** rather than the true outcome. How likely is it that the computer will report a point in **P** and thus lead one to conclude that the model has been confirmed? If **P** is so small that the probability of this happening is zero, then the model is quantitatively testable.

An example of a quantitatively testable model is provided by Luce's (1959) *Choice Model* applied to a pair comparison situation in which a subject has to decide, for each pair of objects, which of the two is more preferable. Let M_{ij} denote the probability with which the subject prefers object i over object j. Thus, $0 \leqslant M_{ij} \leqslant 1$. Assume that $M_{ii} = 0.5$ and $M_{ij} = 1 - M_{ji}$. Hence, when there are N objects, **O** is $[0, 1]^{N(N-1)/2}$, i.e., **O** is the set of the above-diagonal entries of $N \times N$ matrices $[M_{ij}]$ whose elements are between 0 and 1. According to the Choice Model, there exist scale values $v_1, ..., v_N$ such that $M_{ij} = v_i/(v_i + v_j)$. It can be shown that the existence of this scale implies that **P** is defined by equalities. Specifically, **P** consists of those matrices in **O** where, for all $i < k < j$, $M_{ij}/(1 - M_{ij}) = [M_{ik}/(1 - M_{ik})][M_{kj}/(1 - M_{kj})]$.

An example of a qualitatively testable "model" is given by the property of *strong stochastic transitivity*, which states that if M_{ik} and $M_{kj} \geqslant 0.5$, then $M_{ij} \geqslant M_{ik}$ and $M_{ij} \geqslant M_{kj}$.

Assessing Whether a Model Has "Too Many Parameters"

Important as it is to know whether a model is quantitatively testable, it has typically been assumed that the assessment of quantitative testability is trivial. Indeed, common opinion holds that all one has to do is to apply the following rule:

COUNTING RULE. *Count, first, the number of "independent" parameters in the model and, second, the number of "independent" quantities measured in the experiment. The model is quantitatively testable if and only if the latter exceeds the former.*

In the above, the number of "independent" quantities in a possibly larger set is the size of the smallest subset that determines the entire set. To be precise, consider m variables $x_1, ..., x_m$. Let $\mathbf{A} \subset R^m$ consist of all allowable combinations of $x_1, ..., x_m$. Given some $k \leqslant m$, and some $i_1, ..., i_k$, where $1 \leqslant i_1 < \cdots < i_k \leqslant m$, a *coordinate-deleting function* is a function from R^m to R^k that maps $(x_1, ..., x_m)$ onto $(x_{i_1}, ..., x_{i_k})$. Then, the number of "independent" variables is defined as the smallest value of k for which there exists a coordinate-deleting function that is 1–1 when restricted to

A. For example, consider a model that has m parameters that may take on any values subject to the constraint that they always sum to one (i.e., **A** is the subset of R^m given by $\{(x_1,..., x_m) | x_1 + \cdots + x_m = 1\}$). This model has only $m-1$ "independent" parameters, for we may take $k = m-1$, $i_1,..., i_k = 1,..., m-1$, and let the coordinate-deleting function map $(x_1,..., x_{m-1}, x_m)$ to $(x_1,..., x_{m-1})$. This function is 1–1 when restricted to **A**.

However, as we shall see in this article, the matter of determining whether a model is testable is substantially more complicated than the counting rule. Briefly, the complication is the following. For each possible combination of parameter values, the model predicts an outcome of the experiment. Now, this correspondence between parameter values and predicted outcomes need not be one-to-one. If the correspondence is infinitely many-to-one, then the set **P** of predicted outcomes may be much smaller than if the correspondence were one-to-one. In fact, the correspondence may map the set of possible parameter values onto a set **P** that has zero probability measure even when there are *more* "independent" parameters than "independent" quantities observed in the experiment.

Goal

The goal of this paper is to investigate how a model's testability depends on properties of (a) the set of possible parameter values and (b) the equations that define the model. More specifically, our goal is to provide a criterion for quantitative testability that is relatively easy to apply (although not quite as easy as the Counting Rule). The criterion we develop can be applied mechanically to most models using standard computer programs.

This brings us to the outline of the present paper. First, we provide a rigorous definition of quantitative testability. Second, we show that the Counting Rule is incorrect—generally, it is too conservative (i.e., falsely categorizes models as not being quantitatively testable). Third, we provide our criterion for assessing quantitative testability. Fourth, we discuss a slightly modified version of quantitative testability, which is concerned with imprecision of empirical measurement. When measurement imprecision is taken into account, the Counting Rule can be too liberal as well as too conservative. Finally, we discuss relations between quantitative testability, parameter identifiability, reparameterization, and statistical tests of goodness-of-fit.

Remark. This paper is concerned with experiments of the finite-dimensional type only. By "finite-dimensional" the following is meant. Suppose that there exist a finite number n of quantities whose true values are to be measured. An experiment that measures these n quantities (to a greater or lesser accuracy) is finite-dimensional. For example, an experiment that measures the probability of each of n events is finite dimensional. On the other hand, an experiment that measures the cumulative distribution function of a continuous random variable X is not finite-dimensional. In such an experiment, the experimenter wants to know the values of an infinite number of quantities. Specifically, for each real number a, the experimen-

ter wants to know $P(X \leqslant a)$. Note that, with a sufficiently large but finite sample size, this infinite set of quantities can be measured to any degree of accuracy (Rao, 1973, Sect. 6f.1). Thus, it is not sample size but number of quantities to be measured that determines whether an experiment is finite-dimensional.

In practice, potentially infinite-dimensional experiments are frequently converted to finite-dimensional experiments. For example, in a reaction-time experiment, the data are often reduced to the mean and variance, or to the parameters that characterize an assumed family of cumulative distribution functions. In this paper, we limit ourselves to finite-dimensional experiments.

II. Quantitative Testability Defined

Models

A model, in the most general sense of the word, can be tested in a variety of experiments. In fact, one often constructs a model in the absence of any notion of how to test it. However, once one has decided on an experiment, one deals with a specific version of the model, namely, the model as applied to the particular experiment. In what follows, we use the term "model" in this narrow sense of experiment-specific model.

Consider an experiment where n observable quantities are measured. These quantities' *true* values (as distinct from their measured values) define a point in R^n which is called the true outcome of the experiment. (For example, in experiments in which we observe relative frequencies of events, the true outcome consists of the underlying probabilities, not the observed relative frequencies.) Any point in R^n that could logically be the true outcome is called a potential outcome; the set of all potential outcomes is called the outcome space and is denoted **O**.

Suppose that a model that applies to this experiment has m parameters. Then, any combination of parameter values may be represented as a point in R^m. Let the set of all *possible* parameter value combinations be termed the *parameter domain* and let it be denoted **D**. The parameter domain may be a proper subset of R^m. For example, if the parameters represent the probability of each of m mutually exclusive and exhaustive events, then the parameter domain consists of all points in R^m whose coordinates are non-negative and sum to one.

For each possible combination of parameter values, the model predicts the true value of the n quantities to be measured. In effect, the model specifies a function (called the *prediction function*) that maps possible parameter combinations onto predicted experimental outcomes.[2] A point in the outcome space is consistent with

[2] Prediction functions should not be confused with *theoretical relations*. The former are tied to a particular experiment; the latter are not. An example will make this clearer. Suppose we have a theory of springs that states that the length y of a spring is a linear function of the weight w suspended from the spring. Thus $y = x_1 + wx_2$, where x_1 and x_2 are parameters. This is an example of a *theoretical relation*. In an experiment, weights (w) of 1, 2, and 3 kg are suspended from the spring and the its length is measured. Then, the prediction function *for this experiment* is $f(x_1, x_2) = (x_1 + x_2, x_1 + 2x_2, x_1 + 3x_2)$. We shall not deal with theoretical relations in this paper.

the model if and only if it lies in the range of the prediction function, i.e., the image of the parameter domain under the prediction function. We call the range of the prediction function the prediction range and denote it by $\mathbf{P}$.

The above considerations are summarized by the following formal definition:

DEFINITION 1. $(\mathbf{D}, f, \mathbf{O})$ is a *model* if $\mathbf{D} \subset R^m$, $\mathbf{O} \subset R^n$, and $f: \mathbf{D} \to \mathbf{O}$. In this definition, we call $\mathbf{D}$ the *parameter domain*, f the *prediction function*, and $\mathbf{O}$ the *outcome space*. The *prediction range* $\mathbf{P}$ is defined as $f[\mathbf{D}]$.

Before using these definitions (which, by the way, are unrelated to the meaning of *model* in formal logic), we give a few illustrations.

EXAMPLES 2. (a) Consider, again, Luce's (1959) Choice Model. With N objects, the parameters are the scale values $v_1, \ldots, v_N$. Hence, $m = N$ and $\mathbf{D} = (R^+)^N$. As discussed earlier, $\mathbf{O} = [0, 1]^{N(N-1)/2}$. The prediction function f maps each $(v_1, \ldots, v_N) \in \mathbf{D}$ onto the above-diagonal elements of the $N \times N$ matrix $[v_i/(v_i + v_j)]$.

(b) Suppose that, in testing Estes' (1959) *Small Element Model*, the experimenter reduces the data to *learning curves*. That is, the raw data, which consist of error–success strings, are transformed into curves that indicate for each trial the probability of a correct response. Let N be the number of trials, then the number of observables $n = N$, and $\mathbf{O} = [0, 1]^N$. Since there are only two parameters (a learning parameter c and a guessing parameter g) that both range between 0 and 1, $m = 2$ and the parameter domain $\mathbf{D}$ is $[0, 1]^2$. The prediction function may be written in the form $f(c, g) = [f_1(c, g), \ldots, f_N(c, g)]$, where each $f_i(c, g)$ is the predicted probability correct on trial i and equals $1 - (1 - g)(1 - c)^{i-1}$.

(c) In *Factor Analysis*, with r respondents and v variables, the number of outcomes $n = rv$ and $\mathbf{O} = R^n$. With q factors $(q < r - 1, v)$, the parameters may be represented by the $r \times q$ matrix $S = [S_{ik}]$ of respondents' factor scores, the $q \times v$ matrix $L = [L_{kj}]$ of variables' factor loadings, and the $r \times v$ matrix $M = [M_{ij}]$ of variables' intercepts (r identical rows). Hence, the number of parameters $m = rq + qv + v$ and $\mathbf{D} = R^{rq + qv + v}$. The prediction function f maps matrix triples $(S, L, M) \in \mathbf{D}$ onto the $r \times v$ matrix $SL + M$. Consequently, $\mathbf{P} = f[\mathbf{D}]$ is the set of all $r \times v$ matrices that have the property that if one subtracts from each element the corresponding column mean, the resulting matrix has rank q or less. (The rank of a matrix H is the number of linearly independent columns (or, equivalently, rows) in the matrix and is denoted $\rho(H)$.)

Quantitative Testability

We now define quantitative testability. The starting point for our definition of quantitative testability is our desire that there should be zero probability that a point randomly sampled from the outcome space $\mathbf{O}$ will lie in $\mathbf{P}$. Of course, a definition of quantitative testability must define this concept in a way that is independent of the sampling distribution on $\mathbf{O}$. Fortunately, it is known that any

measurable set that has *Lebesgue measure* zero also has a sampling probability of zero provided that the sampling distribution can be characterized by a density function. (All commonly used continuous probability distributions can be characterized by density functions.)

Roughly speaking, Lebesgue measure corresponds to line length in the one-dimensional case, to area in the two-dimensional case, to volume in the three-dimensional case, etc.

We define weak and strong forms of quantitative testability, the difference between which we shall elucidate below.

Weak Quantitative Testability

DEFINITION 3. A model $(\mathbf{D}, f, \mathbf{O})$ is *weakly quantitatively testable* (WQT)[3] if (1) $\lambda(\mathbf{O}) > 0$ and (2) $\lambda(\mathbf{P}) = 0$. Here, $\lambda(A)$ denotes the Lebesgue measure of a Lebesgue-measurable set $A \subset R^n$.

Note the requirement that $\mathbf{O}$ cannot have measure zero. The reason for this requirement is that, without it, the definition could sometimes be vacuous: If $\lambda(\mathbf{O}) = 0$, then $\lambda(\mathbf{P}) = 0$ for any $\mathbf{P} \subset \mathbf{O}$, even if $\mathbf{P} = \mathbf{O}$. To illustrate, consider an experiment with N mutually exclusive and exhaustive response categories, in which the quantities to be measured are response probabilities. Let y_i denote the *probability* of a response in the ith category. Then $\mathbf{O}$ is $\{(y_1, \ldots, y_N) | \sum_{i=1}^{N} y_i = 1,$ and $y_i \geqslant 0\}$. The measure of $\mathbf{O}$ in R^N is zero; in fact, $\mathbf{O}$ is a subset of an $(N-1)$-dimensional flat. Then, $\mathbf{P}$ would also have measure zero even if $\mathbf{P}$ were the same as $\mathbf{O}$ and the model, thus, would not even be qualitatively testable. However, we can resolve this difficulty by *transforming* $\mathbf{O}$ into $\mathbf{O}'$ as follows: Map each $(y_1, \ldots, y_N) \in \mathbf{O} \subset R^N$ to $(y_1, \ldots, y_{N-1}) \in \mathbf{O}' \subset R^{N-1}$ so that $\mathbf{O}' = \{(y_1, \ldots, y_{N-1}) | \sum_{i=1}^{N-1} y_i \leqslant 1,$ and $y_i \geqslant 0\}$. Then $\mathbf{O}'$ has positive measure in R^{N-1}, and no model for which $\mathbf{P}' = \mathbf{O}'$ would be quantitatively testable.

Strong Quantitative Testability

In real life, tests of models are complicated by the existence of measurement error. While in this paper we are not concerned with statistical issues, there is one measurement-error issue that needs to be addressed for our concept of testability to be a reasonable "screening device" for models. In particular, we have tacitly assumed that if the true outcome t of an experiment is not in $\mathbf{P}$, then, by measuring t with sufficient accuracy, we can ultimately determine that t lies outside of $\mathbf{P}$. Surprisingly, there are situations where, no matter how accurately t is measured, it cannot be demonstrated that t lies outside of $\mathbf{P}$.

EXAMPLE 4 (Lissajous figure). Let $(\mathbf{D}, f, \mathbf{O})$ be a model where $\mathbf{D} = R$, $\mathbf{O} = [-1, 1]^2 \subset R^2$, and $f(x_1) = (\sin x_1, \sin \vartheta x_1)$, where ϑ is an irrational constant.

[3] The abbreviations WQT and SQT will be used both as adjectival and as noun phrases. Thus, WQT can mean either weakly quantitatively testable or weak quantitative testability.

It follows readily from Theorem 9 later in this paper that this model is WQT. However, it is shown in Appendix A that the model's prediction range is *dense* in the experiment's outcome space. What does this mean? Suppose that it is known that the true outcome t of the experiment lies within distance α of the measured outcome y. Then, no matter how small α is, there are points of $\mathbf{P}$ that are within distance α of y. Thus, no matter how small the measurement error α, it can never be demonstrated that t lies outside $\mathbf{P}$. I.e., if the mode is false, it can never be proven so.

The above example shows that measurement error (no matter how small) can render some WQT models untestable. This situation provides a rationale for defining a new and stronger form of quantitative testability which does not allow models to be rendered untestable by small measurement errors. We do this as follows. First, for reasons that will soon be evident, we define a model's *apparent* prediction range as $\mathrm{Cl}(\mathbf{P}) \cap \mathbf{O}$, where $\mathrm{Cl}(\mathbf{P})$ denotes the (topological) closure of $\mathbf{P}$. Then, we define *strong* quantitative testability by substituting the apparent for the actual prediction range in the WQT definition.

DEFINITION 5. A model $(\mathbf{D}, f, \mathbf{O})$ is *strongly quantitatively testable* (*SQT*) (see footnote 3) if (1) $\lambda(\mathbf{O}) > 0$ and (2) $\lambda(\mathrm{Cl}(\mathbf{P}) \cap \mathbf{O}) = 0$.

We give now a more detailed rationale for the use of $\mathrm{Cl}(\mathbf{P}) \cap \mathbf{O}$ in this definition. Let γ denote a confidence level (e.g., 0.95). An event will be said to *usually* occur if its probability is at least γ. For the sake of simplicity, suppose that measurement error operates as follows. For fixed γ, any experiment can be characterized by an imprecision level $\alpha > 0$. Let t and y be the true and measured outcomes, respectively. Then y is *usually* a point in R^n such that $|y - t| < \alpha$. (Here, $|y - t|$ denotes the Euclidean length of the vector $y - t$.) The imprecision level α can be made arbitrarily small[4] by devoting more resources to the experiment (e.g., by building more precise equipment or by increasing the size of random samples). Thus, for any $\beta > 0$, the experiment can be conducted with imprecision α where $0 < \alpha \leqslant \beta$.

Suppose the following model-testing strategy is adopted. In an experiment with imprecision level α, the model is accepted whenever the measured value y of the outcome is less than distance α from any point $\zeta \in \mathbf{P}$; otherwise the model is rejected.

A point $p \in \mathbf{O}$ is said to *appear* to be in $\mathbf{P}$ if, given that the true outcome is p, the model will *usually* be accepted no matter how small the experimental imprecision is. It is evident that, if in fact $p \in \mathbf{P}$, then p will appear to be in $\mathbf{P}$. What is not so evident is that some points outside of $\mathbf{P}$ can also appear to be in $\mathbf{P}$. It will now be shown that the set of points in $\mathbf{O}$ that appear to be in $\mathbf{P}$ is $\mathrm{Cl}(\mathbf{P}) \cap \mathbf{O}$.

Suppose $p \in \mathrm{Cl}(\mathbf{P}) \cap \mathbf{O}$. If p is the true outcome and y the measured outcome and α the imprecision, then *usually* $\alpha - |y - p| > 0$. So, since p is in the closure of $\mathbf{P}$,

[4] In some quantum mechanical experiments, Heisenberg's uncertainty principle prevents such unlimited accuracy from being achieved.

there *usually* exists some $\zeta \in \mathbf{P}$ such that $|p-\zeta| < \alpha - |y-p|$. Hence, $|y-\zeta| \leqslant |y-p| + |p-\zeta| < \alpha$, *usually*. Thus, no matter how small α is, the model will *usually* be accepted.

Suppose p is not in $\mathrm{Cl}(\mathbf{P}) \cap \mathbf{O}$. Since p is not in the closure of $\mathbf{P}$, there exists some $\beta > 0$ such that $|p-\zeta| > \beta$ for all $\zeta \in \mathbf{P}$. So, if p is the true outcome and if the experiment has imprecision $\beta/2$ or less, then the model will *usually* be rejected.

This result explains why we call $\mathrm{Cl}(\mathbf{P}) \cap \mathbf{O}$ the apparent prediction range and, thus, provides a rationale for the previously given definition of SQT. Note that the Lissajous figure in Example 4 is not SQT since its apparent prediction range is all of $\mathbf{O}$.

Remark. If a model $(\mathbf{D}, f, \mathbf{O})$ is SQT and if $\mathbf{O}$ is closed, then $\mathbf{P}$ is nowhere dense in R^n. (Proof: if $\mathbf{P}$ were *not* nowhere dense, then $\mathrm{Cl}(\mathbf{P}) = \mathrm{Cl}(\mathbf{P}) \cap \mathbf{O}$ would have a non-empty interior and, hence, $\lambda(\mathrm{Cl}(\mathbf{P}) \cap \mathbf{O}) > 0$.) The converse does not hold. A model with $\mathbf{P}$ nowhere dense need not be SQT since $\mathbf{P}$ could have positive Lebesgue measure (Gelbaum and Olmsted, 1964, Chap. 8, Example 4).

III. Why We Cannot Count Parameters

Note that the Lissajous-figure model (Example 4) violates the Counting Rule. Since the model has two independent observable quantities but only one parameter, it should be quantitatively testable according to the Counting Rule. However, in a real experiment, this model would be rendered completely untestable by the imprecision of measurement.

In this example, the Counting Rule fails by being too *liberal*; it falsely classifies an untestable model as being quantitatively testable. However, this is an unusual type of failure for the Counting Rule. Its typical failure is to be too *conservative*; i.e., to falsely classify quantitatively testable models as being not quantitatively testable. We now present an example of this type of failure.

Elsewhere (Bamber and van Santen, 1980; in preparation; and van Santen and Bamber, 1981), we have discussed a class of models that share the assumption that perception is mediated by a finite number of internal states. We call these models *discrete state models.* Also in other areas, such as psychophysics [e.g., low- and high-threshold models (Swets, Tanner, and Birdsall, 1961; Luce, 1963a, 1963b; Krantz, 1969)], and attention research (Sperling, 1983), discrete state models are common. For many discrete state models, the Counting Rule is conservative in the sense of falsely indicating that they are not quantitatively testable. To illustrate this conservativeness of the Counting Rule, we discuss the simplest of the discrete state models, namely, the *Threshold Model.* This model includes the low- and high-threshold models as special cases.

Example 6 (*The Threshold Model*). Consider a psychophysical experiment where, on each trial, the subject is presented with one of three possible stimuli: (a) a

strong signal, (b) a weak signal, or (c) no signal at all. To each stimulus, the subject responds with one of three possible ratings (a) confident that a signal was presented, (b) unsure, or (c) confident that no signal was presented.

According to the Threshold Model, presentation of one of the three stimuli stochastically causes the subject to enter either a "detect" state D or a "non-detect" state ND. The subject's confidence rating is stochastically determined by his internal state. Thus, the conditional probability of the subject responding with rating r given stimulus s is

$$P(r|s) = P(D|s)\,P(r|D) + P(ND|s)\,P(r|ND). \tag{1}$$

There is one such equation for each combination of three stimuli and three ratings, a total of nine equations.

These nine versions of Eq. (1) may be summarized by a simple matrix equation. Let C denote the 3×3 matrix of conditional probabilities $P(r|s)$ with each row representing a different stimulus s and each column representing a different rating r. Let A denote the 3×2 matrix where each row represents a different stimulus s and contains $P(D|s)$ on the left and $P(ND|s)$ on the right. Let B denote the 2×3 matrix where each column represents a different rating r and contains $P(r|D)$ on the top and $P(r|ND)$ on the bottom. Then, the nine versions of Eq. (1) can be summarized by the matrix equation

$$C = AB.$$

Matrices A and B contain the model's parameters. Matrix C contains the predicted values of the conditional probabilities observed in the experiment. Of course, the matrices A, B, and C must all be *probability matrices*, i.e., all matrix entries must be non-negative and the sum of each row must be one. (Some writers use the term *row stochastic* to denote probability matrices.) Other than requiring that they be probability matrices, we place no restrictions on the parameter matrices A and B.

The Threshold Model makes the following prediction. Since A and B have two columns and two rows, respectively, their *ranks* are two or less. Since C is their product, its rank cannot be more than theirs. Thus, the Threshold Model predicts that the rank of C will be two or less.

This prediction is falsifiable since not all 3×3 probability matrices have rank two or less. In fact, *most* 3×3 probability matrices have rank three. This leads one to suspect that the Threshold Model is not merely testable, but is quantitatively testable. Indeed, it can be shown, using Theorem 17 from later in this paper, that the Threshold Model is SQT.

An interesting aspect of the Threshold Model is that, despite its quantitative testability, its parameters cannot be uniquely estimated. To see this, suppose C is any 3×3 probability matrix that is consistent with the threshold model. In general, there will not be a unique pair of 3×2 and 2×3 probability matrices A and B

whose product is C. For suppose that A and B are such a pair. Then it can easily be shown that generally there are (uncountably) many 2×2 nonsingular probability matrices Θ which have the property that $A^* = A\Theta^{-1}$ and $B^* = \Theta B$ are probability matrices. Obviously, $A^*B^* = A\Theta^{-1}\Theta B = AB = C$.

It will now be shown that the Counting Rule gives the wrong result for the Threshold Model. The matrix C is 3×3 and, so, contains 9 observable quantities. However, the entries in each row must sum to one; so, only $3(3-1)$ or 6 of these observables are "independent." Similarly, A contains $3(2-1)$ or 3 "independent" parameters, and B contains another $2(3-1)$ or 4. So, the total number of "independent" parameters $3+4=7$ is greater than the number of "independent" observables 6. Thus, according to the Counting Rule, the Threshold Model should not be quantitatively testable. However, we have just seen that, in fact, it is SQT. Thus, the Counting Rule is overly conservative in this case.

The same situation occurs in most other discrete state models studied by Bamber and van Santen (1980; in preparation). Moreover, this situation is, of course, not confined to Discrete State models. For example, in Luce's (1959) Choice Model, the Counting Rule is also too conservative. As discussed in Example 2, the number of parameters equals N, the number of objects. Moreover, the parameters are all "independent" since no proper subset of them determines the entire set. The number of "independent" observable quantities is $N(N-1)/2$. Hence, when $N=3$, the number of "independent" parameters N and the number of "independent" observables $N(N-1)/2$ both equal 3, so that, according to the Counting Rule, the model should not be testable. However, it is easy to show that for $N=3$, the Choice Model is SQT. (In this case, the reason why the models is quantitatively testable is intuitively obvious: Since $v_i/(v_i+v_j) = \alpha v_i/(\alpha v_i + \alpha v_j)$ for any non-zero α, we can always eliminate one parameter from the model by arbitrarily setting $v_N = 1$, so that we have only $N-1=2$ parameters left.)

We conclude that the Counting Rule can give wrong answers that are too conservative. This leaves two options. First, one may attempt to modify the Counting Rule, by applying it to a model only after one has made certain changes in the model that reduce the number of "independent" parameters while leaving **P** intact (i.e., a *reparameterization*; see Sect. VI). This was demonstrated above in the case of the Choice Model, but not in the case of the Threshold Model; in fact, in the latter case it is quite difficult to find a way to make the appropriate changes. Second, one may develop an entirely new rule for assessing quantitative testability. This is what we have done.

IV. A New Criterion for Weak Quantitative Testability

The main conclusion from the preceding section is that a model can be WQT (i.e., its prediction range **P** can have measure zero) even when there are as many or more "independent" parameters than "independent" observables. Apparently, the

measure of **P** depends in some manner on properties of the prediction function f. What is the central mathematical principle that underlies this dependency?

Relevance of the Jacobian Matrix

The central mathematical principle is that the *Jacobian matrix of the prediction function* should have *rank* smaller than the number n of independent observable quantities. The prediction function's Jacobian matrix is constructed as follows. For any point x in the model's parameter domain **D**, we may write

$$x = (x_1, ..., x_m)',$$
$$f(x) = [f_1(x), ..., f_n(x)]'.$$

(The functions $f_1, ..., f_n$ are called the *scalar components* of f.) The Jacobian matrix of f evaluated at $x^0 \in \mathbf{D}$ is the $n \times m$ matrix of partial derivatives

$$\mathbf{J}_{x^0 f} = [\partial f_j(x^0)/\partial x_i]$$

In other words, the value of the entry in column i and row j is $\partial f_j(x^0)/\partial x_i$. We shall attempt to give the reader some intuition for the role of this matrix's rank.

Suppose that f is differentiable at x^0. Consider the function α: $R^m \to R^n$ defined by

$$\alpha(x) = f(x^0) + \mathbf{J}_{x^0 f}[x - x^0]. \tag{3}$$

Note that, in this equation, $\alpha(x)$ and $f(x^0)$ are $n \times 1$ vectors, x and x^0 are $m \times 1$ vectors, and $\mathbf{J}_{x^0 f}$ is an $n \times m$-matrix. Near the point x^0, the function α is an approximation to the function f (Flett, 1966, Sect. 10.4). The range of α is the flat obtained by (a) taking the linear subspace spanned by the column vectors of the Jacobian matrix and (b) translating this subspace by $f(x^0)$. Now, the range of α has measure zero in R^n if and only if its dimension—which, of course, equals the Jacobian matrix rank $\rho(\mathbf{J}_{x^0 f})$—is less than n. Thus, if $\rho(\mathbf{J}_{x^0 f}) < n$, then f is approximated near x^0 by a function whose range has measure zero. Naturally, this leads one to suspect that, if $\rho(\mathbf{J}_{xf}) < n$, for *all* $x \in \mathbf{D}$, then the range **P** of f will have measure zero.

Indeed, this turns out to be true provided that f is sufficiently "smooth." "Smoothness" is here defined in terms of differentiability. Specifically, a function f is said to be C^q at a point x^0 in its domain if all qth order partial derivatives of each of the scalar components of f exist in a neighborhood of x^0 and are continuous at x^0. A function is said to be C_q if it is C^q at each point in its domain. The larger the value of q for which a function is C^q, the "smoother" the function is. (Clearly, if a function is C^q and if p is a positive integer less than q, then the function is also C^p.)

The above definitions imply that a function can be C^q at x^0 only if x^0 is an *interior* point of the function's domain; hence, a function can be C^q only if its domain is *open*. The restriction of a C^q-function to an *open* subset of its domain is also C^q.

Sard's Theorems

We now formalize these intuitive notions. Sard (1942) proved a number of theorems that provide the critical link between the measure of **P** and properties of **D**, f, and **O**. For convenience, we state here the relevant theorems from Sard (1942). Sard's theorems deal with sets in R^n having "M-dimensional measure" zero. We shall not define this measure here. Suffice it to say that, if a subset of R^n has M-dimensional zero and $M \leqslant n$, then the subset has Lebesgue measure zero.

We shall use the following notation. If f is differentiable on an open set **S**, let

$$\mathbf{S}_t^r = \{x \mid x \in \mathbf{S} \text{ and } t \leqslant \rho(\mathbf{J}_{xf}) \leqslant r\}.$$

LEMMA 7 (Theorem 7.1 in Sard, 1942). *Let $r < n < m$. Suppose* **S** *is a region (i.e., connected open subset) of R^m and $f: \mathbf{S} \to R^n$ is C^q for some $q \geqslant (m-r)/(n-r)$. Then $f[\mathbf{S}_r^r]$ has n-dimensional measure zero in R^n.*

LEMMA 8 (Theorem 4.1 in Sard, 1942). *Let $n \geqslant m$. Suppose* **S** *is a region of R^m and $f: \mathbf{S} \to R^n$ is C^1. Then $f[\mathbf{S}_0^{m-1}]$ has m-dimensional measure zero in R^n.*

The First Main Result: the WQT Theorem

These two Lemmas allow us to prove the following central theorem.

THEOREM 9 (WQT Theorem). *Suppose* **S** *is an open subset of R^m and the function f maps* **S** *into R^n. Suppose* $\mathbf{D} \subset \mathbf{S}$ *and that* $(\mathbf{D}, f, \mathbf{O})$ *is a model, where* $\lambda(\mathbf{O}) > 0$. *Let* $r = \max_{x \in \mathbf{D}} \rho(\mathbf{J}_{xf})$. *If $r < n$ and if f is C^q on* **S** *for some* $q \geqslant \max\{1, (m-r)/(n-r)\}$, *then the model is WQT.*

Proof. Obviously, $f[\mathbf{D}] \subset f[\mathbf{S}_0^r]$. It will be shown that the latter and, hence, the former has Lebesgue measure zero.

Case 1. **S is connected.** This case is treated as three subcases. In each subcase, $f[\mathbf{S}_0^r]$ is seen to have Lebesgue measure zero.

Subcase (1a). $(m > n)$. Now, $f[\mathbf{S}_0^r] = f[\mathbf{S}_0^0] \cup \cdots \cup f[\mathbf{S}_r^r]$. Apply Lemma 7 to each term on the right.

Subcase (1b). $(m = n)$. Apply Lemma 8.

Subcase (1c). $(m < n)$. Convert this to Subcase (1b) by defining the function $F: \mathbf{S} \times R^{n-m} \to R^n$ such that $F(x, \zeta) = f(x)$ for all $x \in \mathbf{S}$ and $\zeta \in R^{n-m}$.

Case 2. **S is not connected.** For any $x \in \mathbf{S}$, there exists an open ball **B** such that $x \in \mathbf{B} \subset \mathbf{S}$. Thus, **S** may be covered by an uncountable set of open balls. By the Lindelöf theorem (Kelley, 1955, pp. 49–50), a countable subset of these open balls covers **S**. Applying Case 1 to each such open ball **B**, it is seen that $f[\mathbf{D} \cap \mathbf{B}]$ has measure zero. Since $f[\mathbf{D}]$ is the union of a countable number of such $f[\mathbf{D} \cap \mathbf{B}]$, it follows that $f[\mathbf{D}]$ has measure zero. Q.E.D.

The WQT Theorem states that, *for sufficiently smooth functions f, we can assess whether a model is WQT by determining whether the rank of the Jacobian matrix is everywhere less than the dimension of the space containing* $\mathbf{O}$. This is our new criterion for WQT. Before we further discuss issues concerning the practical use of this criterion, we comment on a few mathematical aspects of this theorem.

Comments on the WQT Theorem

Our first comment concerns independence of both observations and parameters. Recall that, in the Counting Rule, one counts only "independent" observations and only "independent" parameters. The WQT Theorem contains similar independence requirements albeit in disguised forms. The requirement that $\lambda(\mathbf{O}) > 0$ is a type of independence requirement for, if some observable quantities were (smooth) functions of the remaining observable quantities, then one would have $\lambda(\mathbf{O}) = 0$. Thus, it may be necessary to transform some models (as discussed after Definition 3) to guarantee that $\lambda(\mathbf{O}) > 0$. Likewise, the requirement that the prediction function be C^q is (among other things) a type of independence requirement. If one parameter were a function of the remaining parameters, then it would be impossible to vary that parameter while holding the remaining parameters constant. Hence, partial derivatives with respect to that parameter would be undefined and, consequently, the prediction function would not be C^q. In such cases, it is necessary to reparameterize the model in order to apply the WQT Theorem. For example, suppose the parameters must always sum to one: $x_1 + \cdots + x_m = 1$. Then x_m can be eliminated as a parameter by expressing the prediction function in the form $f(x_1, \ldots, x_{m-1}, 1 - x_1 - \cdots - x_{m-1})$.

Our second comment is that the smoothness (i.e., C^q) requirement cannot be dropped from the WQT Theorem. Sard (1942, p. 890) and Gelbaum and Olmsted (1964, Chap. 10, Example 8) give examples showing this. Note that f is required to be C^q at every point of $\mathbf{S}$, not just the points of $\mathbf{D}$.

The third comment is that, since, like any other $n \times m$ matrix, $\rho(\mathbf{J}_{xf}) \leqslant \min\{n, m\}$, a model is always WQT when $m < n$, provided that f is C^1. This establishes that, for models whose prediction function is C^1 (on $\mathbf{S}$), and for which $m < n$, the Counting Rule leads to the correct conclusion that the model is WQT. Of course, Example 6 shows that the Counting Rule can give the wrong result when $m \geqslant n$.

The fourth comment concerns the question of whether the reverse of the WQT Theorem is true. Is it true that a model for which $\max_{x \in \mathbf{D}} \rho(\mathbf{J}_{xf})$ equals n is necessarily not WQT? The following proposition asserts that this is indeed the case, provided that f is C^1 at least one interior point where the Jacobian matrix has rank n (This is a far weaker requirement than the requirement that f be C^q everywhere in an open set $\mathbf{S}$ containing $\mathbf{D}$.)

PROPOSITION 10. *Let* $(\mathbf{D}, f, \mathbf{O})$ *be a model with* $\mathbf{O} \subset R^n$. *Suppose that* $\mathbf{D}$ *contains an interior point* x^0 *such that* (a) f *is* C^1 *at* x^0 *and* (b) *the Jacobian matrix has rank n at* x^0. *Then the model is not* WQT.

Proof. According to (10.11.1 Corollary 3) of Flett (1966), there exists an open set $\mathbf{U}$ ($x^0 \in \mathbf{U} \subset \mathbf{D}$) such that $f[\mathbf{U}]$ is open in R^n. Hence, $\lambda(f[\mathbf{U}]) > 0$. Q.E.D.

Finding the Maximum Rank of the Jacobian Matrix

To apply the WQT Theorem, it is necessary to show for the model under consideration that the maximum rank r of the Jacobian matrix is less than n. Since this matrix is $n \times m$, $r \leqslant n, m$; so, if $m < n$—i.e., there are fewer parameters than observable quantities—it follows trivially that $r < n$. However, in the case where $m \geqslant n$, which is of focal interest for this paper, it is not so trivial to verify whether $r < n$.

One possible approach is to *prove a theorem* that asserts an upper bound for the rank of the Jacobian matrix. However, since our methods are targeted for a quick screening of a model under development, this approach is not pragmatic.

Fortunately, a much simpler approach is possible for a large class of models. These are models whose prediction functions are *analytic*. Under these circumstances, the Jacobian matrix has maximum rank almost everywhere in $\mathbf{D}$ and has submaximum rank almost nowhere. Consequently, the maximum rank may be found by randomly selecting a point in $\mathbf{D}$, calculating the rank at that point, and assuming that that rank is the maximum. Although it is theoretically possible for this procedure to give an incorrect value for the maximum, the probability of this happening is zero. The precise details are given in the following theorem which is proven in Appendix B.

THEOREM B.5. *Suppose* $\mathbf{S}$ *is a connected open subset of* R^m *and the analytic function* f *maps* $\mathbf{S}$ *into* R^n. *Suppose* $\mathbf{D} \subset \mathbf{S}$ *and* $\lambda(\mathbf{D}) > 0$. *Let* X *be a random point in* $\mathbf{D}$ *whose distribution* P *is generated by a probability density function on* $\mathbf{D}$. *Then,*

$$P[\rho(\mathbf{J}_{Xf}) = \max_{x \in \mathbf{D}} \rho(\mathbf{J}_{xf})] = 1.$$

Appendix B also shows that, if the Jacobian matrix has maximum rank at a point x, then it will also have maximum rank at all points that lie near x (i.e., in a neighborhood of x). Thus, rounding off the coordinates of the random point X by a small amount will not cause an error in determining the maximum rank.

Analytic functions are extremely smooth; in fact, analytic functions are C^q for all positive integers q (Dieudonné, 1960, Result 9.3.6). Thus, if a model's prediction function is analytic, it necessarily satisfies the smoothness requirement of the WQT Theorem.

When a model's prediction function is analytic and certain other conditions are met, the following algorithm may be used to determine whether the model is WQT. This algorithm is justified by the WQT Theorem, Proposition 10, and Theorem B.5.

RANDOM-POINT ALGORITHM (WQT Version). *Conditions.* Suppose $\mathbf{S}$ is a connected open subset of R^m and the analytic function f maps $\mathbf{S}$ into R^n. Suppose $\mathbf{D} \subset \mathbf{S}$ and $\mathbf{O} \subset R^n$ and $(\mathbf{D}, f, \mathbf{O})$ is a model. Suppose $\mathbf{D}$ and $\mathbf{O}$ have positive

Lebesgue measure in R^m and R^n, respectively. *Actions.* Sample a random point X from a distribution generated by a probability density function on $\mathbf{D}$. Calculate the Jacobian matrix rank $\rho(\mathbf{J}_{Xf})$ at X. *Conclusions.* (a) If $\rho(\mathbf{J}_{Xf}) < n$, then one may be 100 percent confident that the model is WQT. (b) If $\rho(\mathbf{J}_{Xf}) = n$ and if X is an interior point of $\mathbf{D}$, then the model is definitely not WQT.

Later, it will be shown how to add further conditions to the Random Point Algorithm so that it may also be used to determine whether models are SQT rather than merely WQT.

To apply the Random Point Algorithm with a given model, it is necessary to (a) recognize whether the model's prediction function is analytic, (b) recognize whether $\lambda(\mathbf{D}) > 0$ and $\lambda(\mathbf{O}) > 0$, and (c) calculate the rank of the Jacobian matrix at a point. These three steps will now be considered in turn.

Recognizing Analytic Functions

Dieudonné (1960, Sect. 9.3) gives a formal definition of analytic functions. Roughly speaking, a function from an open subset of R^m into R^n is analytic if, at every point, it can be locally expressed as the sum of an absolutely convergent power series in $x_1, \ldots, x_m$ (i.e., the coordinates of R^m). This formal definition does not provide easy means of recognizing analytic functions. Fortunately, the following facts enable one to recognize a wide variety of analytic functions: (a) A vector-valued function is analytic if and only if each of its scalar-valued component function is analytic. (b) Common analytic functions of one variable include polynomials, exponential and logarithmic functions, power functions, and the standard trigonometric functions and their inverses. (c) Whenever defined, the sum, difference, product, ratio, and composition of analytic functions is analytic. (d) Polynomials of more than one variable are analytic. (e) All partial derivatives of analytic functions are analytic. (f) The integral of an analytic function of one variable is analytic. (g) If a function of one real variable can be extended to an invertible analytic function of one *complex* variable, then its inverse is analytic (Hoffman, 1975, Chap. 5, Theorem 20).

It is clear from the above facts that analytic prediction functions are very common in the models that one encounters.

Recognizing whether $\mathbf{D}$ *and* $\mathbf{O}$ *Have Positive Lebesgue Measure*

To show that a set has positive Lebesgue measure, one shows, first, that the set is Lebesgue measurable, and, second, that the measure is not zero.

In practice, one not need worry about Lebesgue measurability of $\mathbf{D}$ and $\mathbf{O}$. Virtually any set one is liable to construct will be Lebesgue measurable. The known methods for constructing sets that are not Lebesgue measurable are exotic and utilize the axiom of choice. To demonstrate that a set's measure is not zero, it suffices to show that the set contains a nonempty open set.

Finding the Rank of the Jacobian Matrix

The calculation of the Jacobian matrix at a given point can be accomplished by a variety of both analytic and numerical methods. Once this matrix has been obtained, its rank can be evaluated most conveniently using a computer program that performs the *singular value decomposition* (Stewart, 1973, pp. 317–318). Specifically, a matrix's rank equals the number of its nonzero singular values. The EISPACK and LINPACK packages of computer programs contain subroutines for performing this decomposition. Moreover, Nash (1979, pp. 30–31) provides a decomposition algorithm suitable for use on microcomputers.

When using a computer to determine the rank of a matrix, one needs to be concerned with round-off error. In order to represent each matrix element with a fixed number of memory bits, the computer will round off each matrix element (The amount by which a computer internally rounds off numbers may be determined using an algorithm described by Malcolm (1972) and Nash (1979, pp. 4–6).) Fortunately, the singular value decomposition is numerically stable. When a matrix is perturbed (by rounding it off, for example), its singular values are perturbed less than the matrix itself (Lawson and Hanson, 1974, p. 25; Stewart, 1973, pp. 321–322). Thus, after the singular values of a matrix have been computed, one can determine which singular values are definitely nonzero and which ones are so close to zero that they are in all likelihood perturbations of true zeros.

Maximum Jacobian Matrix Rank for Some Models

We conclude this section by giving a few examples of models and the ranks of their Jacobian matrices. Examples 11(a) and (b) serve to demonstrate that the weak quantitative testability of rather forbidding-looking models can be assessed without great difficulty, and that, in addition, seemingly slight changes in f (in this case, the change of an exponent of -2 into an exponent of 2) may have a impact on the measure of $\mathbf{P}$ that could no be anticipated by cursory inspection of the models' formulae. Example (c) gives the rank of the Threshold Model (Example 6).

EXAMPLES 11. (a) Let $\mathbf{D} = (R^+)^2$, $\mathbf{O} = R^2$, and let

$$f(x_1, x_2) = (2 \log x_1 - \log x_2, x_2[x_1 + x_2^{1/2}]^{-2}).$$

It can be shown that $\max_{x \in \mathbf{D}} \rho(\mathbf{J}_{xf}) = 1$.

(b) Let $\mathbf{D} = (R^+)^2$, $\mathbf{O} = R^2$, and let

$$f(x_1, x_2) = (2 \log x_1 - \log x_2, x_2[x_1 + x_2^{1/2}]^{2}).$$

It can be shown that $\max_{x \in \mathbf{D}} \rho(\mathbf{J}_{xf}) = 2$.

(c) The Threshold Model (Example 6). It can be shown that (see Bamber and van Santen, 1980; in preparation) $\max_{x \in \mathbf{D}} \rho(\mathbf{J}_{xf}) = 5$. Since the number of independently observable quantities is 6, it follows from Theorem 9 that the model is WQT.

V. Conditions for Strong Quantitative Testability

The WQT Theorem states conditions that guarantee that a model is WQT. In this section, we present what we shall call *SQT conditions*, i.e., conditions which, if added to those of the WQT Theorem, guarantee that a model is SQT.

Preliminaries

One of the properties of the Lissajous figure (Example 4) that causes **P** to be dense in **O**, with the consequence that the model is not SQT, is that **O** is bounded whereas **D** is not. Thus, **P** consists of a curve of infinite length that has to be fitted in the bounded square **O**, and thus "passes near" every point in this square. While it is obvious that boundedness of **O** in conjunction with unboundedness of **D** is not sufficient by itself to cause a model not to be SQT (e.g., consider the Lissajous figure with rational ϑ), we surmised that the requirement that a model have a bounded parameter domain could be a SQT condition. It turns out that, together with one more condition, this condition indeed is a SQT condition. The following *within–within* condition is a generalization of this idea.

Definition 12. A model $(\mathbf{D}, f, \mathbf{O})$ is said to be *within–within* or satisfy the *within–within* condition if and only if for each $\beta > 0$ there exists an $\alpha > 0$ having the following property: If $y \in f(\mathbf{D})$ and $|y| \leqslant \beta$, then there exists an $x \in \mathbf{D}$ such that $|x| \leqslant \alpha$ and $f(x) = y$. As before, $|\ |$ denotes the Euclidean norm (i.e., length) of the enclosed vector.

If a model is within–within, then, for any bounded set $B \subset f[\mathbf{D}]$, there exists a bounded set $A \subset f^{-1}[B] \subset \mathbf{D}$ such that $f[A] = B$ (Note that, although A is bounded, $f^{-1}[B]$ may be unbounded.)

One can verify that the Lissajous figure indeed does not satisfy the within–within condition.

The following proposition provides some easy methods for verifying that a model satisfies the within–within condition.

Proposition 13. *Suppose* $\mathbf{D} \subset R^m$ *and* $f\colon \mathbf{D} \to \mathbf{O} \subset R^n$, *then statement* (a) *implies statement* (b) *implies statement* (c), *where*:

(a) **D** *is bounded.*

(b) *If* $\{x_i\}$ *is a sequence of points in* **D** *and if* $|x_i| \to \infty$ *then* $|f(x_i)| \to \infty$.

(c) *The model* $(\mathbf{D}, f, \mathbf{O})$ *satisfies the within–within condition.*

Proof. (a) → (b). Suppose **D** is bounded. Then (b) is trivially true.

not-(c) → not-(b). Suppose (c) is not true. Then there exists a $\beta > 0$ such that, for every $\alpha > 0$, there exists a $y_\alpha \in f[\mathbf{D}]$ such that $|y_\alpha| \leqslant \beta$, but there is no $x \in \mathbf{D}$ such that $|x| \leqslant \alpha$ and $f(x) = y_\alpha$. Choose x_α such that $f(x_\alpha) = y_\alpha$. Then $|x_\alpha| > \alpha$.

Now, for $\alpha = 1, 2, 3,...$, consider the sequence $\{x_\alpha\}$. Then $|x_\alpha| \to \infty$ but $|f(x_\alpha)| = |y_\alpha| \leqslant \beta$. Hence (b) does not hold. Q.E.D.

Note that the statements (a), (b), and (c) of Proposition 13 are not equivalent. Obviously, (b) does not imply (a). A trivial example of a model $(\mathbf{D}, f, \mathbf{O})$ that satisfies (c) but not (b) is: $\mathbf{D} = R^2$, $\mathbf{O} = R^1$, and $f(x_1, x_2) = x_1$.

We shall now prove that, roughly speaking, a WQT model satisfying the within–within condition (and one additional assumption) is SQT.

DEFINITION 14. Suppose $(\mathbf{D}, f, \mathbf{O})$ is a model. Define the set $\mathbf{D}^*$ by letting $x \in \mathbf{D}^*$ if and only if there exists a sequence $\{x_i\}$, where (a) each $x_i \in \mathbf{D}$, (b) $x_i \to x$, and (c) for some $y \in \mathbf{O}$, $f(x_i) \to y$. [Obviously, $\mathbf{D} \subset \mathbf{D}^* \subset \mathrm{Cl}(\mathbf{D})$.]

DEFINITION 15. Suppose $(\mathbf{D}, f, \mathbf{O})$ is a model. Then a **-extension* of $(\mathbf{D}, f, \mathbf{O})$ is a model $(\mathbf{D}^*, h, \mathbf{O})$ such that $h(x) = f(x)$ for all $x \in \mathbf{D}$. This *-extension is said to be *continuous* if h is continuous on $\mathbf{D}^*$.

A continuous *-extension of $(\mathbf{D}, f, \mathbf{O})$ need not exist, but if it does exist, it is unique.

LEMMA 16. *Let* $(\mathbf{D}, f, \mathbf{O})$ *be a model which* (a) *is within–within and* (b) *has a continuous *-extension. Then* $(\mathbf{D}, f, \mathbf{O})$ *is SQT if and only if its continuous *-extension is WQT.*

Proof. Let $(\mathbf{D}^*, h, \mathbf{O})$ denote the continuous *-extension. It will be shown that $\mathrm{Cl}(f[\mathbf{D}]) \cap \mathbf{O} = h[\mathbf{D}^*]$. From this, the lemma follows.

Suppose $x \in \mathbf{D}^*$. Let $x_i \to x$, where each $x_i \in \mathbf{D}$ and $f(x_i) \to y \in \mathbf{O}$. Since h is continuous on $\mathbf{D}^*$, $f(x_i) = h(x_i) \to h(x)$. Thus, $h(x) \in \mathrm{Cl}(f[\mathbf{D}])$. Moreover, $h(x) = y \in \mathbf{O}$. So, $h[\mathbf{D}^*] \subset \mathrm{Cl}(f[\mathbf{D}]) \cap \mathbf{O}$.

Suppose $y \in \mathrm{Cl}(f[\mathbf{D}]) \cap \mathbf{O}$. Then, there exists a sequence $\{y_i\}$ such that $y_i \to y$, where each $y_i \in f[\mathbf{D}]$. Since this sequence converges, it is bounded. Since $(\mathbf{D}, f, \mathbf{O})$ is within–within, there exists a bounded sequence $\{x_i\}$ such that, for each i, $f(x_i) = y_i$ where $x_i \in \mathbf{D}$. Since $\{x_i\}$ is bounded, it contains a convergent subsequence $\{x_{i_j}\}$. Let x denote the limit of this subsequence. Then, $x \in \mathbf{D}^*$. Since h is continuous on $\mathbf{D}^*$, $h(x_{i_j}) \to h(x)$. Since it is already known that $h(x_{i_j}) = f(x_{i_j}) = y_{i_j} \to y$, it follows that $y = h(x)$. Hence, $y \in h[\mathbf{D}^*]$. So $\mathrm{Cl}(f[\mathbf{D}]) \cap \mathbf{O} \subset h[\mathbf{D}^*]$. Q.E.D.

The Second Main Result: the SQT Theorem

THEOREM 17 (SQT Theorem). (a) *Suppose* $\mathbf{S}$ *is an open subset of* R^m *and the function* f *maps* $\mathbf{S}$ *into* R^n. *Suppose that* $\mathbf{D} \subset \mathbf{S}$ *and that* $(\mathbf{D}, f, \mathbf{O})$ *is a model where* $\lambda(\mathbf{O}) > 0$. *Let*

$$r = \max_{x \in \mathbf{D}} \rho(\mathbf{J}_{xf}).$$

If $r < n$ *and if* f *is* C^q *on* $\mathbf{S}$ *for some* $q \geqslant \max\{1, (m-r)/(n-r)\}$, *then the model is*

WQT. (b) *If, in addition to the above,* $(\mathbf{D}, f, \mathbf{O})$ *is within–within and* $\mathbf{D}^* \subset \mathbf{S}$, *then the model is SQT.*

Proof. Part (a) is merely a restatement of the WQT Theorem. To prove (b), let

$$r^* = \max_{x \in \mathbf{D}^*} \rho(\mathbf{J}_{xf}).$$

Obviously, $r^* \geqslant r$. It will now be shown that $r^* \leqslant r$. Suppose $x \in \mathbf{D}^*$. Since f is continuously differentiable at x, there exists a neighborhood $\mathbf{M}$ of x such that $\rho(\mathbf{J}_{xf}) \leqslant \rho(\mathbf{J}_{wf})$ for all $w \in \mathbf{M}$ (Flett, 1966, p. 412, third paragraph). Since $x \in \mathbf{D}^* \subset \mathrm{Cl}(\mathbf{D})$, some of these $w \in \mathbf{M}$ must lie in $\mathbf{D}$. Hence, for all $x \in \mathbf{D}^*$, $\rho(\mathbf{J}_{xf}) \leqslant r$. Hence, $r^* \leqslant r$ and, so, $r^* = r$. Since f is C^q on $\mathbf{S} \supset \mathbf{D}^*$, the model $(\mathbf{D}^*, f, \mathbf{O})$ is a continuous *-extension of $(\mathbf{D}, f, \mathbf{O})$. By Theorem 9, $(\mathbf{D}^*, f, \mathbf{O})$ is WQT. Then, by Lemma 16 $(\mathbf{D}, f, \mathbf{O})$ is SQT. Q.E.D.

Comments on the SQT Theorem

Consider the model (a) $(\mathbf{D}, f, \mathbf{O})$. If $\mathbf{O}$ is not closed, it may be easier to work with the model (b) $[\mathbf{D}, f, \mathrm{Cl}(\mathbf{O})]$ rather than with model (a). Thus, if $\mathrm{Cl}(\mathbf{O}) - \mathbf{O}$ has Lebesgue measure zero, then model (a) is SQT if and only if model (b) is SQT. Model (b) is easier to work with because it is easier to characterize the set $\mathbf{D}^*$ for model (b) than for model (a). For model (b), $x \in \mathbf{D}^*$ if and only if, for some sequence $\{x_i\}$ of points in $\mathbf{D}$, $x_i \to x$ and the sequence $\{f(x_i)\}$ is bounded.

The SQT Theorem contains two requirements not found in the WQT Theorem, namely, that $\mathbf{D}^* \subset \mathbf{S}$ and that the model be within–within. Now, any model can be reparameterized so that it has a bounded parameter domain and, hence, by Proposition 13, is within–within. (This may be accomplished by replacing the old parameters $x_1, \ldots, x_m$ with the new parameters $w_k = \exp x_k/(1 + \exp x_k)$ so that $0 < w_k < 1$ $(k = 1, \ldots, m)$.) However, we know from the Lissajous model (Example 4) that there exist models which satisfy the WQT Theorem but which are not SQT. This shows, then, that the $\mathbf{D}^* \subset \mathbf{S}$ requirement cannot be dropped from the SQT Theorem.

The Random-Point Algorithm discussed earlier is a procedure for determining whether a point is WQT. By adding the conditions that $\mathbf{D}^* \subset \mathbf{S}$ and that the model be within–within, this procedure may also be used to determine whether a model is SQT.

The Threshold Model discussed in Example 6 satisfies the SQT conditions stated in Theorem 17. We may take $\mathbf{S} = \mathbf{R}^7$ (the model has 7 independent parameters). Thus, $\mathbf{D}^* \subset \mathbf{S}$. The parameters of this model are all conditional probabilities and, thus, lie between zero and one. Hence, the model's parameter domain is bounded and, so, by Proposition 13, this model is within–within. (In fact, it can be shown that this model satisfies the within–within condition even after it has been modified by dropping the restriction that the parameters lie between zero and one.) Thus, the model is not only WQT but also SQT.

Our SQT conditions are sufficient but not necessary. To see that they are not

necessary, consider the following trivial example. Let $(\mathbf{D}, f, \mathbf{O})$ be a model, where $\mathbf{D} = (1, \infty) \subset R^1$, $\mathbf{O} = R^2$, and $f(x) = (1/x, 1/x)$. This model is obviously SQT, because $\mathbf{P} = \{(y_1, y_2) | 0 < y_1 = y_2 < 1\}$ is a line segment in Euclidean two-space. However, it is *not* within–within because, although $f[\mathbf{D}]$ is bounded, there is not bounded $\mathbf{A} \subset \mathbf{D}$ such that $f[\mathbf{A}] = f[\mathbf{D}]$.

Thus, there do exist SQT models that cannot be shown to be SQT by the SQT Theorem. However, a reparameterization will sometimes make a model susceptible to the theorem. In the above example, we can reparameterize as follows: $\mathbf{D} = (0, 1) \subset R^1$, $\mathbf{O} = R^2$, and $f(x) = (x, x)$.

Since our SQT conditions are sufficient but not necessary, it might be useful to search for SQT conditions different from those we have presented. Consider the following *generalized monotonicity condition* that would require that each scalar component of $f(x)$ have a finite number of zero-crossings of the first partial derivative with respect to each component of x, holding the other components of x constant. We speculate that this condition is an SQT condition.

It is difficult to construct examples of models that are WQT but not SQT. Hence, if one can prove that a model is WQT but cannot prove that it is SQT, it seems reasonable to nevertheless assume the model is SQT.

VI. Some Remaining Issues

In this section, we discuss how the notion of quantitative testability and our criteria for assessing quantitative testability are related to *reparameterizations*, *identifiability*, and *degrees of freedom*.

Quantitative Testability and Reparameterizations

Definition 18. A model $(\mathbf{D}', f', \mathbf{O}')$ is a *reparameterization* of another model $(\mathbf{D}, f, \mathbf{O})$ if the two models have the same observation space and the same prediction range, i.e., if $\mathbf{O}' = \mathbf{O}$ and $f'[\mathbf{D}'] = f[\mathbf{D}]$.

In the above definition, the parameter domains $\mathbf{D}$ and $\mathbf{D}'$ need not have the same number of parameters. Thus, it is possible that $\mathbf{D} \subset R^m$ and $\mathbf{D}' \subset R^{m'}$ where $m \neq m'$.

The relevance of reparameterizations for assessing quantitative testability is the following. One might conjecture that a model is WQT if and only if it has a reparameterization with fewer parameters than independently observable quantities. We call this conjecture the *Modified Counting Rule*.

Take literally, this conjecture is false. Unless the reparameterization's domain is required to have a nonempty interior, the prediction function is required to be "smooth," and $\lambda(\mathbf{O})$ is required to be non-zero, counterexamples to the conjecture can be constructed. With the addition of these requirements, the conjecture might be correct; but we do not know.

What we do know is that, even if some form of it is correct, the Modified Counting Rule is not a *practical* method for assessing whether a model is WQT. To

show that a model is WQT using the Modified Counting Rule, one must find an appropriate reparameterization. However, failure to find such a reparameterization does not prove that it does not exist. Perhaps, one has not been clever enough in searching for a reparameterization.

For example, Bamber and van Santen (1980; in preparation) found that a general class of discrete-state models can be reparameterized in a manner consistent with the Modified Counting Rule. However, the reparameterization is rather tricky. We had worked with this class of models for a few years before we discovered the reparameterization. In the meantime, we had long known, from considering their Jacobian matrices, under which conditions the models are WQT.

In contrast to the Modified Counting Rule, the WQT Theorem and the SQT Theorem provide a practical, straightforward method of determining that a model is WQT or SQT. Provided that the model's prediction function is analytic, this method may be applied mechanically using a computer.

Quantitative Testability and Identifiability

DEFINITION 19. A model $(\mathbf{D}, f, \mathbf{O})$ is *identifiable* if its prediction function is one-to-one.

In other words, a model is identifiable if, given any experimental outcome y consistent with the model (i.e., $y \in \mathbf{P}$), the values of the model's parameters are uniquely determined.

There is no relationship between identifiability and testability. On the one hand, a model is testable if there is at least one $y \in \mathbf{O}$ such that $y \notin f[\mathbf{D}]$; however, f does not have to be one-to-one. On the other hand, a model is identifiable if for any $y \in f[\mathbf{D}]$ there is a *unique* $x \in \mathbf{D}$ such that $y = f(x)$; however, $f[\mathbf{D}]$ may be equal to $\mathbf{O}$. Thus, it is neither true that testability implies identifiability, nor vice versa. In fact, it is not even true that *quantitative* testability implies identifiability: many discrete-state models (Bamber and van Santen, 1980; in preparation), including the Threshold Model, are quantitatively testable but not identifiable.

The following proposition shows that, roughly speaking, a model is *not* identifiable if the rank of its Jacobian matrix is everywhere less than the number of its parameters. This result also casts light on the difference between quantitative testability and identifiability. In the former case, the Jacobian matrix rank is compared with n, in the latter case with m. The Threshold Model, for example, has more parameters than the rank of its Jacobian matrix and, as discussed earlier, is not identifiable.

PROPOSITION 20. *Let* $\mathbf{S}$ *be an open set in* R^m *and let* $f: \mathbf{S} \to R^n$ *be* C^1. *Let* $\mathbf{D} \subset \mathbf{S}$ *consist of a nonempty open set plus none, some, or all of its boundary points, and let* $\mathbf{O} \supset f[\mathbf{D}]$. *If* $\rho(\mathbf{J}_{xf}) < m$ *everywhere in* $\mathbf{D}$, *then the model* $(\mathbf{D}, f, \mathbf{O})$ *is not identifiable.*

Proof. Let r denote the Jacobian matrix's maximum rank within the *interior* of

D. Select an interior point x^0 where this rank is attained. Then **D** contains an open neighborhood **M** of x^0 in which the Jacobian matrix has everywhere rank at least r (Flett, 1966, p. 412, third paragraph). Hence, the Jacobian matrix has rank r everywhere in **M**. Hence, by the Rank Theorem (Flett, 1966, 10.14.3 Corollary), there exists an open neighborhood $\mathbf{V}^*$ of x^0 such that $\mathbf{V}^* \subset \mathbf{D}$ and $f^{-1}[\{f(x^0)\}] \cap \mathbf{V}^*$ is homeomorphic with an open set in R^{m-r}. This shows, since $m > r$, that *uncountably many* points in $\mathbf{V}^* \subset \mathbf{D}$ map onto $f(x^0)$. Q.E.D.

Identifiability and the Counting Rule

While there is no relationship between identifiability and testability, there is a relationship between identifiability and the performance of the Counting Rule. The next proposition shows, roughly speaking, that the Counting Rule is valid for identifiable models.

PROPOSITION 21. *Let* **S**, **D**, f, *and* **O** *be as in Proposition* 20 *and let* $\lambda(\mathbf{O}) > 0$. *Suppose the model* (**D**, f, **O**) *is identifiable. Then the model is WQT if and only if* $m < n$.

Proof. Let r be as in the proof of Proposition 20. The argument of the SQT Theorem shows that r is also the maximum rank over *all* of **D**. Since the Jacobian matrix is $n \times m$, $r \leqslant n, m$. Since the model is identifiable, it follows from Proposition 20 that $r = m$. Hence, either (a) $m = r < n$ or (b) $m = r = n$. If (a), then the model is WQT by the WQT Theorem. If (b), then there exists an *interior* point x^0 of **D** at which the Jacobian matrix has rank $r = n$. Then, by Proposition 10, the model is not WQT. Q.E.D.

This result helps explain scientists' apparent satisfaction with the Counting Rule. The majority of models that scientists have worked with are either identifiable or have been reparameterized so that they are. Moreover, these models typically satisfy the "smoothness" requirements of Proposition 21. Thus, within this class of models, the Counting Rule correctly assesses whether a model is WQT.

The problem with the Counting Rule is that it can give an incorrect assessment of testability for an important class of models, namely, nonidentifiable models.

Two Roles of Models: Estimating Parameters versus Testing Assumptions

Frequently, one sets up a model because one wants to estimate parameters that *measure* important aspects of psychological processes. For example, in signal detection theory, the parameter d' measures perceptual accuracy in a way that is unaffected by response bias.

In a different use of models, however, the parameters are of secondary importance and the *assumptions* of the model are central, because they represent a psychological theory. One's primary interest is in empirically determining whether this theory is correct. Whether the model's parameters are identifiable is not relevant, as long the model is testable. An example is provided by measurement axioms in

axiomatic measurement (Krantz, Luce, Suppes, and Tversky, 1971). Although one could, in principle, estimate the parameters or scale values (up to some degree of uniqueness), the principal interest in, say, additive conjoint measurement is in whether *additivity* holds. This can be assessed without estimating parameters. In the context of *finite state confusion models* (van Santen and Bamber, 1981) the assumption of strict sequentiality and restrictions on the number of internal states express interesting psychological hypotheses that also can be tested without estimating parameters. Another way of viewing the role of testable, non-identifiable, models is by realizing that a testable, non-identifiable, model may be a generalization of several identifiable models. In this case, rejection of the general model can be quite informative. For example, if we can reject the Threshold Model, then we can reject both the low- and high-threshold models as well.

Rank of the Jacobian Matrix and Degrees of Freedom: A General Goodness-of-fit Test

Bamber and van Santen (unpublished) have developed a general goodness-of-fit test for finite-dimensional models. This test is related to Hotelling's T^2 test and has an asymptotic χ^2-distribution. Significantly, the degrees of freedom are given by $n - r_{\max}$, where $r_{\max}$ is the maximum rank of the Jacobian matrix. Thus, the rank of the Jacobian matrix is of importance not only for assessing weak and strong quantitative testability, but also for goodness-of-fit testing. The test extends standard results (e.g., Cramér, 1946, Sect. 30.3; Birch, 1964) in two ways. First, it does not require that the Jacobian matrix have full-column rank (i.e., rank equal to the number of parameters). Second, it applies to non-multinomial random vectors. Here, we state the main results; the proofs may be found in the original paper.

Let $\mathbf{Y}_1, \ldots, \mathbf{Y}_k$ be k independently, identically distributed (i.i.d.) $n \times 1$ random vectors. Let

$$M_k = \frac{1}{k} \sum_{i=1}^{k} \mathbf{Y}_i$$

and

$$S_k = \frac{1}{k-1} \sum_{i=1}^{k} [\mathbf{Y}_i - M_k][\mathbf{Y}_i - M_k]'$$

denote the sample mean and sample covariance matrix.

The theorem requires some conditions that are different from those underlying the WQT Theorem. The most important of these is the following. Suppose $x_0 \in \mathbf{D} \subset R^m$ and $f: \mathbf{D} \to R^n$. Suppose that, for every neighborhood $\mathbf{N}$ of x_0, there exists a neighborhood $\mathbf{Q}$ of $f(x_0)$ such that

$$f[\mathbf{D}] \cap \mathbf{Q} \subset f[\mathbf{D} \cap \mathbf{N}].$$

Then f is said to be *self-contained at* x_0. Roughly speaking, this condition says that,

if $y \in f(\mathbf{D})$ is near $f(x_0)$, then at least one point in $f^{-1}(\{y\})$ is near x_0. Note that the Lissajous-figure model (Example 4) does not meet this condition.

The self-containment condition is a generalized version of Birch's (1964) Condition B. Birch employed that condition to correct a serious flaw in Cramér's (1946, Sect. 30.3) theorem on multinomial goodness-of-fit testing.

We can now test the hypothesis that $\mu[=E(\mathbf{Y}_i)] \in f[\mathbf{D}]$ as follows.

THEOREM 22. *Let* $(\mathbf{D}, f, \mathbf{O})$ *be a model where* $\mathbf{D} \subset R^m$ *and* $\mathbf{O} \subset R^n$. *Suppose the i.i.d. random vectors* $\mathbf{Y}_1,..., \mathbf{Y}_k$ *have nonsingular covariance matrix* Σ *and mean* μ. *Suppose that there exists some interior point* x_0 *of* $\mathbf{D}$ *such that* $\mu = f(x_0)$, f *is both continuously differentiable and self-contained at* x_0, *and* $\rho(\mathbf{J}_{xf})$ *has a local maximum at* x_0. *Then, as* $k \to \infty$,

$$\inf_{x \in \mathbf{D}} k \cdot [M_k - f(x)]' S_k^{-1} [M_k - f(x)]$$

is asymptotically χ^2 *distributed with* $n - \rho(\mathbf{J}_{x_0 f})$ *degrees of freedom.*

Remarks. First, the χ^2 distribution with zero degrees of freedom is the distribution of a random variable that equals zero with probability one. Second, if S_k is nonsingular, then so is Σ. Conversely, if Σ is nonsingular, then $\lim_{k \to \infty} P(S_k \text{ is nonsingular}) = 1$. Third, note that the goodness-of-fit statistic is minimum Hotelling's T^2. Fourth, the theorem assumes that f is continuously differentiable at x_0 and that $\rho(\mathbf{J}_{xf})$ has a local maximum at x_0. This implies that $\rho(\mathbf{J}_{xf})$ is locally constant about x_0 (Flett, 1966, p. 412, third paragraph). Fifth, Appendix B shows that, if $\mathbf{D}$ is open and connected and if f is analytic, then $\rho(\mathbf{J}_{xf})$ is at its global maximum at almost every point in $\mathbf{D}$.

We now state a theorem that specifically applies to multinomial situations. It is essentially a corollary to the previous theorem.

THEOREM 23. *Let* $(\mathbf{D}, f, \mathbf{O})$ *be a model where* $\mathbf{D} \subset R^m$ *and* $\mathbf{O} \subset R^n$. *Let* $f_j(x)$ *denote the jth component of the vector* $f(x)$. *Suppose that, for all* $x \in \mathbf{D}$, $\sum_{j=1}^n f_j(x) = 1$ *and* $f_j(x) \geq 0$ *for* $j = 1,..., n$. *Let* $A_1,..., A_n$ *be n random events, with non-zero probabilities* $P(A_1),..., P(A_n)$. *Suppose there is some interior point* x_0 *of* $\mathbf{D}$ *such that* $f(x_0) = [P(A_1),..., P(A_n)]'$, f *is both continuously differentiable and self-contained at* x_0, *and* $\rho(\mathbf{J}_{xf})$ *has a local maximum at* x_0. *Suppose that, on each of k independent trials, exactly one of the random events* $A_1,..., A_n$ *occurs. Let* p_{kj} $(j = 1,..., n)$ *denote the proportion of k trials resulting in event* A_j. *Then, as* $k \to \infty$,

$$\inf_{x \in \mathbf{D}} k \cdot \sum_{j=1}^{n} \frac{[p_{kj} - f(x)]^2}{p_{kj}}$$

is asymptotically χ^2 *distributed with* $n - 1 - \rho(\mathbf{J}_{x_0 f})$ *degrees of freedom.*

Remark. This goodness-of-fit statistic is a minimized, modified χ^2 statistic of the type Σ *(observed-expected)*^{2}*/observed.*

Related Work

Identifiability of Markov learning models. This issue has been studied intensively by many psychologists. Greeno, Millward, and Merryman (1971) and Larkin and Wickens (1980) presented procedures for reparameterizing a model in a canonical form having at most as many as and possibly fewer parameters than the original model.

Polson and Huizinga (1974) defined a model to be *empirically identifiable* if, given the results of an experiment, the maximum likelihood estimate of the model's parameters is unique. They suggested using empirical identifiability as a diagnostic sign; they claimed that, if a Markov learning model is not empirically identifiable, then it is most likely not identifiable.

Bayesian parameter estimation. If a model is not identifiable in the sense of having a one-to-one prediction function (Definition 19), then some or all points in the model's prediction range will be consistent with more than one point in the parameter domain. In other words, empirical observations will not completely constrain model parameters. Nevertheless, empirical observations may partially constrain parameters by limiting them to a subset of the parameter domain. Chechile and Meyer (1976; Chechile, 1977) have proposed Bayesian parameter estimation procedures for use with models that are not identifiable in the sense of Definition 19. These procedures take advantage of the partial constraint on parameters afforded by empirical observations. These procedures operate as follows: (a) A prior probability distribution over the points of the parameter domain is assumed. (b) Using the prior parameter distribution and the empirical observations, a posterior parameter distribution is calculated. (c) A measure of central tendency of the posterior parameter distribution is taken as the parameter estimate.

Because these estimation procedures can be applied to models that are not identifiable in the sense of Definition 19, Chechile (1977) has argued that the concept of model identifiability should be extended. He has proposed a Bayesian concept that defines identifiability in terms of posterior parameter distributions.

Although Chechile and Meyer's procedure can be applied to models having more parameters than observations, the goal of their procedures are quite different from the goal of this paper. Chechile and Meyer's procedures are applied to a model which is assumed to be correct and which may be either testable or untestable. The goal of their procedure is to estimate model parameters. In contrast, the goal of this paper is to present procedures for determining whether a model is testable.

VII. Summary

We discussed under which conditions a model is testable in the sense that it has a prediction range **P** with measure zero in the observation space **O**. We claim that this type of testability, which we call *quantitative testability*, formally expresses the

intuitive notion of testability that people have in mind when they say that a model has sufficiently few parameters to be testable.

With examples we show that, contrary to common belief, assessing quantitative testability is not a simple matter of comparing the number of "independent" parameters with the number of "independently" observable quantities (the Counting Rule). Specifically, models may be quantitatively testable even when they have more "independent" parameters than observable quantities. The parameters of such models, we point out, are not identifiable.

Reparameterization of a model, in an attempt to reduce the number of "independent" parameters, is not always a workable approach to assessing quantitative testability. There are two reasons for this. First, not finding the appropriate reparameterization does not imply that one does not exist. Second, and more important, the measure of **P** depends not only on the number of "independent" parameters, but also on properties of the prediction function f. Put in other words, the model's prediction function may introduce dependencies between the predicted observations that cause the measure of **P** to be zero, even when the number of "independent" parameters is equal to or larger than the number of independently observable quantities. These dependencies may be very difficult to find by trial and error. Our goal was to provide a mechanical procedure for discovering their existence.

The procedure consists of calculating the rank of a model's Jacobian matrix. Assuming that certain "smoothness" conditions are met, a model is quantitatively testable if and only if for all possible parameter values, this rank is smaller than the number of observable quantities. When the model's prediction function is analytic, the maximum rank is attained almost everywhere; thus, the rank needs to be determined at only one randomly chosen point. This procedure can be performed by standard computer programs.

We argue that, in the face of measurement error, one wants not only **P** but also its closure in **O** to have measure zero. We discussed conditions under which this occurs.

Finally, we point out that the rank of the Jacobian matrix is of importance not only for determining testability, but also for determining identifiability and for performing statistical tests of goodness-of-fit.

Appendix A

Example 4 (Lissajous figure). Let $(\mathbf{D}, f, \mathbf{O})$ be a model where $\mathbf{D} = R$, $\mathbf{O} = [-1, 1]^2 \subset R^2$, and $f(x_1) = (\sin x_1, \sin \vartheta x_1)$, where ϑ is an irrational constant. This model is WQT but not SQT.

Proof. WQT follows from Theorem 9 (Sect. IV). It will now be shown that $f[\mathbf{D}]$ is *dense* in **O**, from which (by the definition of denseness) it follows that $\lambda(\mathrm{Cl}(f[\mathbf{D}]) \cap \mathbf{O}) > 0$ and hence that the model is not SQT.

Choose any a such that $-1 \leqslant a \leqslant 1$. Choose α such that $\sin \alpha = a$. Let

$$\mathbf{D}^a = \{\alpha + 2\pi j \mid j \text{ an integer}\},$$

$$\mathbf{L}^{\vartheta} = \{\vartheta j + k \mid j \text{ and } k \text{ integers}\}.$$

Then

$$f[\mathbf{D}^a] = \{[a, \sin(\vartheta\alpha + 2\pi x)] \mid x \in \mathbf{L}^{\vartheta}\}.$$

Since ϑ is irrational, $\mathbf{L}^{\vartheta}$ is dense in the real line (Halmos, 1950, Sect. 16, Theorem C). Hence, $f[\mathbf{D}^a]$ is dense in $\{a\} \times [-1, 1]$ and, thus, $f[\mathbf{D}]$ is dense in $\mathbf{O}$. Q.E.D.

Appendix B

Proof That the Jacobian Matrix of an Analytic Function Is at Its Maximum Rank Almost Everywhere

The proof uses a number of well-known facts concerning analytic functions. Except where otherwise noted, proofs of these facts may be found in Sections 9.1–9.4 of Dieudonné (1960).

Lemma B.1. *Let f be a real-valued analytic function defined on a nonempty open interval I of R. Let $Z = f^{-1}(\{0\})$. Then either $Z = I$ or Z has measure zero.*

Proof. Either (a) all points in Z are isolated or (b) at least one point of Z is a cluster point of Z. If (a), then each point of Z may be covered by an open interval that contains no other points of Z. Then, by the Lindelöf theorem (Kelley, 1955, Chap. 1, Theorem 15), Z is countable and, therefore, has measure zero. If (b), let $c \in Z \subset I$ be a cluster point of Z. Then f's power-series expansion about c must be identically zero everywhere in some interval containing c (Rudin, 1953, Theorem 8.5). So, by the principle of analytic continuation, f must be identically zero everywhere in I. Q.E.D.

Lemma B.2. *Let f be a real-valued analytic function defined on a nonempty open box B_m (i.e., Cartesian product of nonempty open real intervals) contained in R^m. Let $Z = f^{-1}(\{0\})$. Then either $Z = B_m$ or Z has measure zero.*

Proof. Use induction on m. Lemma B.1 shows that the desired result is true for $m = 1$. Assume that it is true for $m = 1, \ldots, k$. Consider a $(k+1)$-dimensional nonempty open box $B_k \times I$ ($B_k \subset R^k$, $I \subset R$) with a real-valued analytic function f defined on it. Let $Z = f^{-1}(\{0\})$. Since f is continuous, Z is measurable. For each $y \in I$, let $Z_y = \{x \in B_k \mid f(x, y) = 0\}$. Suppose $Z \neq B_k \times I$. Then, for some $\alpha \in B_k$ and $\beta \in I$, $f(\alpha, \beta) \neq 0$. Consider the set $I_\alpha = \{y \in I \mid f(\alpha, y) \neq 0\}$. By Lemma B.1, I_α con-

sists of all of I except for a subset of measure zero. Now, for each $y \in I_\alpha$, $\alpha \notin Z_y$ and so, by the induction assumption, Z_y has measure zero. Hence, by Fubini's theorem (Hoffman, 1975, p. 354), Z has measure zero. Q.E.D.

LEMMA B.3. *Let f be a real-valued analytic function defined on* $\mathbf{S}$, *a nonempty open connected subset of R^m. Let $Z = f^{-1}(\{0\})$. Then either $Z = \mathbf{S}$ or Z has measure zero.*

Proof. Suppose $Z \neq \mathbf{S}$. For each $x \in \mathbf{S}$, let B_x be an open box such that $x \in B_x \subset \mathbf{S}$ and let $Z_x = B_x \cap Z$. The principle of analytic continuation implies that, if $Z_x = B_x$ for some $x \in \mathbf{S}$, then $Z = \mathbf{S}$. Hence, for every x, $Z_x \neq B_x$ and so, by Lemma B.2, every Z_x has measure zero. By the Lindelöf Theorem (Kelley, 1955, Chap. 1, Theorem 15), a countable number of B_x's cover $\mathbf{S}$ and, so, a countable number of Z_x's cover Z. Since each Z_x has measure zero, so does Z. Q.E.D.

Notation. Suppose $f: \mathbf{S} \to R^n$, where $\mathbf{S}$ is open in R^m. Let $\rho_{\max} = \max_{x \in \mathbf{S}} \rho(\mathbf{J}_{xf})$, and let $\mathbf{S}_{\max} = \{x \mid x \in \mathbf{S} \text{ and } \rho(\mathbf{J}_{xf}) = \rho_{\max}\}$.

PROPOSITION B.4. *Let $\mathbf{S}$ be a nonempty open connected subset of R^m. Let $f: \mathbf{S} \to R^n$ be an analytic function. Then $\mathbf{S}_{\max}$ is open and $\mathbf{S} - \mathbf{S}_{\max}$ has Lebesgue measure zero.*

Proof. Suppose $\rho_{\max} > 0$ as otherwise the theorem is trivially true. Then, $\mathbf{S} - \mathbf{S}_{\max}$ is the set of all $x \in \mathbf{S}$ such that every $\rho_{\max} \times \rho_{\max}$ submatrix of $\mathbf{J}_{xf}$ has a determinant of zero. Since f is an analytic function, so are these determinants. Thus, $\mathbf{S} - \mathbf{S}_{\max}$ is the intersection of the zeroes of a finite number of real-valued analytic functions. Since analytic functions are continuous, $\mathbf{S} - \mathbf{S}_{\max}$ is closed in $\mathbf{S}$ and, so, $\mathbf{S}_{\max}$ is open. Moreover, by Lemma B.3, $\mathbf{S} - \mathbf{S}_{\max}$ either (a) equals $\mathbf{S}$ or (b) has measure zero. Since (a) is precluded by the definition of $\mathbf{S}_{\max}$, (b) must hold. Q.E.D.

Remark. If $x \in \mathbf{S}_{\max}$, it is an interior point of $\mathbf{S}_{\max}$. So, if x_{round} is x rounded off by some sufficiently small amount, then $x_{\text{round}} \in \mathbf{S}_{\max}$.

THEOREM B.5. *Suppose $\mathbf{S}$ is a connected open subset of R^m and the analytic function f maps $\mathbf{S}$ into R^n. Suppose $\mathbf{D} \subset \mathbf{S}$ and $\lambda(\mathbf{D}) > 0$. Let X be a random point in $\mathbf{D}$ whose distribution P is generated by a probability density function on $\mathbf{D}$. Then*

$$P[\rho(\mathbf{J}_{Xf}) = \max_{x \in \mathbf{D}} \rho(\mathbf{J}_{xf})] = 1.$$

Proof. As previously, let r and $\rho_{\max}$ denote the maximum Jacobian matrix rank over $\mathbf{D}$ and over $\mathbf{S}$, respectively. Obviously, $r \leqslant \rho_{\max}$. It will now be shown that $r = \rho_{\max}$. Suppose $r < \rho_{\max}$. Then $\mathbf{D} \subset \mathbf{S} - \mathbf{S}_{\max}$. So, by Proposition B.4, $\lambda(\mathbf{D}) = 0$. But this contradicts the assumption that $\lambda(\mathbf{D}) > 0$. Hence, $r = \rho_{\max}$. Thus

$$P[\rho(\mathbf{J}_{Xf}) < r] = P[x \in \mathbf{D} \cap (\mathbf{S} - \mathbf{S}_{\max})].$$

By Proposition B.4, $\lambda(\mathbf{S} - \mathbf{S}_{\max}) = 0$. So, because X's distribution P is generated by a density function on $\mathbf{D}$,

$$P[x \in \mathbf{D} \cap (\mathbf{S} - \mathbf{S}_{\max})] = 0.$$

Hence

$$P[\rho(\mathbf{J}_{\chi f}) = r] = 1 - P[\rho(\mathbf{J}_{\chi f}) < r] = 1. \qquad \text{Q.E.D.}$$

ACKNOWLEDGMENTS

The first author's work on this paper was supported by Veterans Administration Medical Research funds. The second author was supported by United States Air Force, Life Sciences Directorate, Grant AFOSR 80-0279. This paper has benefitted from extensive comments by Consulting Editor William Batchelder. Conversations with Richard Schweickert and James Townsend have helped to clarify some of the ideas presented in this paper. Also, careful reading by James Harper, and comments by James Neely and Peter Schoenemann, are appreciated. Morris W. Hirsch and Ralph Carr provided us with proof outlines for Appendix B and Example 4, respectively. Portions of this paper were presented at the Sixteenth Annual Mathematical Psychology Meeting, Boulder, Colorado, August 11, 1983.

REFERENCES

BAMBER, D., & VAN SANTEN, J. P. H. (1980). *Testing discrete state models using conditional probability matrices.* Paper presented at the Thirteenth Annual Mathematical Psychology Meeting, Madison, WI.

BAMBER, D., & VAN SANTEN, J. P. H. (in preparation). Discrete state models.

BAMBER, D., & VAN SANTEN, J. P. H. (unpublished). An asymptotically distribution-free test of the fit of multivariate means to nonlinear models.

BIRCH, M. W. (1964). A new proof of the Pearson–Fisher theorem. *Annals of Mathematical Statistics*, **35**, 817–824.

CHECHILE, R. (1977). Likelihood and posterior identification: Implications for mathematical psychology. *British Journal of Mathematical and Statistical Psychology*, **30**, 177–184.

CHECHILE, R., AND MEYER, D. L. (1976). A Bayesian procedure for separately estimating storage and retrieval components of forgetting. *Journal of Mathematical Psychology*, **13**, 269–295.

CRAMÉR, H. (1946). *Mathematical methods of statistics.* Princeton, N.J.: Princeton Univ. Press.

DIEUDONNÉ, J. (1960). *Foundations of modern analysis* (Also published as *Treatise on analysis*, Vol. I). New York: Academic Press.

ESTES, W. K. (1959). Component and pattern models with Markovian interpretations. In R. R. Bush and W. K. Estes (Eds.), *Studies in mathematical learning theory.* Stanford, Calif.: Stanford Univ. Press.

FLETT, T. M. (1966). *Mathematical analysis.* New York: McGraw–Hill.

GELBAUM, B. R., & OLMSTED, J. M. H. (1964). *Counterexamples in analysis*, San Francisco: Holden–Day.

GREENO, J. G., MILLWARD, R. B., AND MERRYMAN, C. T. (1971). Matrix analysis of identifiability of some finite Markov models. *Psychometrika*, **36**, 389–408.

HALMOS, P. R. (1950). *Measure theory.* Princeton, N.J.: Van Nostrand.

HOFFMAN, K. (1975). *Analysis in Euclidean space.* Englewood Cliffs, N.J.: Prentice–Hall.

KELLEY, J. L. (1955). *General topology.* Princeton, N.J.: Van Nostrand.

KRANTZ, D. H. (1969). Threshold theories of signal detection. *Psychological Review*, **76**, 308–324.

KRANTZ, D. H., LUCE, R. D., SUPPES, P., & TVERSKY, A. (1971). *Foundations of measurement*, Vol. I. New York: Academic Press.

LARKIN, J. H., AND WICKENS, T. D. (1980). Population states and eigenstructure: A simplifying view of Markov learning models. *Journal of Mathematical Psychology*, **22**, 176–208.

LAWSON, C. L., AND HANSON, R. J. (1974). *Solving least squares problems.* Englewood Cliffs, N.J.: Prentice–Hall.

LUCE, R. D. (1959). *Individual choice behavior.* New York: Wiley.

LUCE, R. D. (1963a). Detection and recognition. In R. D. Luce, R. R. Bush, and E. Galanter (Eds.), *Handbook of mathematical psychology*, Vol. I. New York: Wiley.

LUCE, R. D. (1963b). A threshold theory for simple detection experiments. *Psychological Review*, **70**, 61–79.

MALCOLM, M. A. (1972). Algorithms to reveal properties of floating-point arithmetic. *Communications of the ACM*, **15**, 949–951.

NASH, J. C. (1979). *Compact numerical methods for computers: Linear algebra and function minimisation.* New York: Halsted Press.

POLSON, P. G., AND HUIZINGA, D. (1974). Statistical methods for absorbing Markov-chain models for learning: Estimation and identification. *Psychometrika*, **39**, 3–22.

RAO, C. R. (1973). *Linear statistical inference and its applications* (2nd ed.). New York: Wiley.

RUDIN, W. (1953). *Principles of mathematical analysis.* New York: McGraw–Hill.

SARD, A. (1942). The measure of the critical values of differentiable maps. *Bulletin of the American Mathematical Society*, **48**, 883–890.

SPERLING, G. (1983). A unified theory of attention and signal detection, In R. Parasuraman and D. R. Davies (Eds.), *Varieties of attention.* New York: Academic Press.

STEWART, G. W. (1973). *Introduction to matrix computations.* New York: Academic Press.

SWETS, J. A., TANNER, W. P., JR., AND BIRDSALL, T. G. (1961). Decision processes in perception. *Psychological Review*, **68**, 301–340.

VAN SANTEN, J. P. H., AND BAMBER, D. (1981). Finite and infinite state confusion models. *Journal of Mathematical Psychology*, **24**, 101–111.

RECEIVED: August 22, 1983