Dedicated to the Memory
of Colin L. Mallows.

Theory of Multiple Psychometric Functions Based on Ratings, with Applications to Temporal-Order Judgments[1]

Saul Sternberg, Ronald L. Knoll, and Colin L. Mallows [2]

Abstract

By using a confidence-rating procedure and varying the stimulus feature being judged over a large range, it is possible to generate a family of psychometric functions (PMFs), each based on a different partition of the ratings. An earlier paper showed how the traditional single PMF based on binary-choice data from temporal-order judgments can be decomposed into sensory and decision components, when it is regarded as an estimate of a probability distribution. Here we extend this development to the confidence-rating procedure, and use it to elucidate the relations among the spreads and shapes of the resulting family of PMFs and their significance. For example, we determine conditions under which the functions can have the same spread and shape, differing only by translation on the stimulus axis. Application of the multiple-function approach to several models, whose tests depend on values of the PMF moments, shows it to have greater power than the single-function approach for understanding the perceptual process. In perceptual domains other than temporal order the most direct application of the proposed models and the multiple PMF method would be to the judgment of differences between pairs of stimuli, such as their pitch or brightness.

Currently (May, 2024) being revised.

---

1. An early draft of this paper, without Section 10, was issued as a Bell Laboratories Technical Memorandum and distributed informally in 1975, when the authors were colleagues at Bell Laboratories, with the title "Conditions for parallel psychometric functions based on rating-scale data: Applications to temporal-order judgments". Reports of research that was influenced by the 1975 version of this paper include those by Allan (1975b) and Ulrich (1987).
   The computations on which this report is based were conducted in R (R Core Team, 2019), and used the R-package cir (A. P. Oron, 2023).

2. Saul Sternberg is at the University of Pennsylvania (saul@psych.upenn.edu). Ronald L. Knoll (ronald.knoll@gmail.com) is retired. Colin L. Mallows died on November 4th, 2023, while this paper was under review. We are grateful to the late Lorraine G. Allan for stimulating our interest in these issues, to David E. Meyer for comments on an earlier draft, to Rolf Ulrich for encouragement, and to George Christiansen of the F. C. Haab Co. and Susan Reisbord, for facilitating manuscript revision during the 2023 heat wave.

Theory of Multiple Psychometric Functions Based on Ratings, with Applications to Temporal-Order Judgments

CONTENTS

1. Introduction

Often, in experiments in which confidence ratings are used in measuring discriminability or testing psychophysical models, relatively few distinct values of the stimulus variable are presented — sometimes only two — so that the data cannot provide information about the spread or shape of the psychometric function (PMF). On the other hand, when the PMF is of primary interest, the observer is often required to choose between only two response alternatives. Suppose that in the same experiment in which the stimulus varies from trial to trial over a large range, the observer is required to choose the response from an ordered set of $n > 2$ categories, such as binary decisions with confidence ratings. Then each of the n-1 partitions of the set of categories can be used to generate a distinct PMF.[3] The potential usefulness of some of the relations among the members of such a family of PMFs for answering theoretical questions has occasionally been recognized. For examples see Nachmias & Steinman (1963) and Eijkman, Thijssen, & Vendrik (1966) in vision, Thijssen & Vendrik (1968) in audition, and Ulrich (1987) in temporal-order perception.

In this paper we first review the representation of the PMF derived from binary-response temporal-order judgments in terms of the components of a general model discussed previously by Sternberg and Knoll (1973), and then generalize this treatment to the confidence-rating procedure. One set of results are specifications of the conditions under which the family of PMFs generated from the procedure can be expected all to have the same spread or shape — i.e., to be parallel, in the sense that they differ by only a translation on the stimulus (time-difference) axis. We also provide examples to show how relations among the shapes of the family of functions can provide tests of models that a single function would not permit and, in general, can help to decompose the observed data into separate contributions from sensory and decision processes.

One reason for our interest in these issues were the findings by L. G. Allan (1975a) of systematically non-parallel sets of PMFs from a procedure in which the observer was required on each trial to judge the order of a pair of stimuli as well as rating them as simultaneous versus successive, thereby generating four ordered response categories. In a second study (Allan, 1975b) the observer was required on each trial to judge order and rate the confidence in this judgment as high or low, again generating four ordered response categories. Data from four rating categories ($\mathbf{A}$ = 1, 2, 3, 4) can be used to generate a family of three PMFs: $F_1 = Pr\{\mathbf{A} > 1\}$, $F_2 = Pr\{\mathbf{A} > 2\}$, and $F_3 = Pr\{\mathbf{A} > 3\}$; in both studies Allan found that the middle function (F_2) was relatively symmetric while F_1 was positively skewed and F_3 negatively skewed. There was also a tendency in Allan's data for the variance of F_2 to be greater than that of F_1 or F_3. Ulrich (1987) used three

---

3. There is controversy as to whether multiple criteria on a perceptual dimension, required for multiple ratings, are less stable than a single criterion, resulting in a loss of sensitivity. Comparing ROC curves generated from binary decisions versus ratings, Egan, Schulman, & Greenberg (1959) showed no sensitivity loss in an auditory experiment, while Swets, Tanner, and Birdsall (1961) showed loss in a visual experiment. In their review, Green and Swets (1966, Sections 4.5, 11.2) conclude that there is minimal loss of sensitivity.

response alternatives ("S_x first", "simultaneous", and "S_y first") to generate two PMFs, similar to Allan's F_1 and F_3. See Section 10 for examples.

What, if anything, might justify or explain our intuition that the members of a PMF family should be parallel? And, if they are not parallel, can we learn something from the relations among their locations, spreads, and shapes?

2. Experimental Paradigm

Consider the following experiment on temporal-order perception: The stimuli are S_x presented at time t_x and S_y presented at time t_y. From trial to trial the time difference $t_y - t_x = d$ takes on various values that can be positive, zero, or negative. After each presentation the observer judges whether S_x appeared to occur before S_y (response $''t_x < t_y''$) or after S_y, and also provides a rating of confidence in the judgment. If we ignore the confidence ratings, the data can be used to estimate a traditional PMF,

$$F(d) = Pr\{''t_x < t_y'' \mid d\}, \tag{1}$$

in which the probability of the judgment that S_x preceded S_y typically increases monotonically with the stimulus variable d over a range from zero to one. Formally similar paradigms involving other sensory features can be treated similarly: Was S_y brighter or dimmer than S_x? Was S_y higher or lower in pitch than S_x?

3. Models for the Psychometric Function Generated by Binary Choice Data

In an earlier paper, Sternberg and Knoll (1973) showed how $F(d)$ could be described in terms of the components of a general independent-channels model. (We refer the reader to that paper for details.) In this model, which is a generalization of numerous models that have been proposed for temporal-order judgments, a "decision function" converts a difference in central "arrival times" of two sensory signals into an order judgment. Let the arrival times of stimuli S_x and S_y be represented by the random variables $\mathbf{U}_x$ and $\mathbf{U}_y$, respectively. The *arrival-time difference* $U_y - U_x$ depends, in turn, on the difference $d = t_y - t_x$ between stimulation times t_y and t_x and separate *arrival latencies* $\mathbf{R}_x$ and $\mathbf{R}_y$ according to

$$\mathbf{U}_y - \mathbf{U}_x = (t_y + \mathbf{R}_y) - (t_x + \mathbf{R}_x) = \mathbf{R}_y - \mathbf{R}_x + d \ . \tag{2}$$

The decision rule induces a decision function G on values of $\mathbf{W} = \mathbf{U}_y - \mathbf{U}_x$, associating an order-decision probability with each value of the arrival-time difference, such that for any value of d,

$$G(W) = Pr\{''t_x < t_y'' \mid \mathbf{U}_y - \mathbf{U}_x = W\} \ . \tag{3}$$

A simple decision rule, and one that is often assumed, is the deterministic decision rule: the observer reports S_x before S_y if and only if the arrival-time difference is non-negative (i.e., matches or exceeds a criterion of zero). Thus, $G(W) = 0$ when $W < 0$, and $G(W) = 1$, otherwise. This rule is readily generalized to an arbitrary criterion, β:

$$G(W) \equiv \begin{cases} 0, & W < \beta \\ 1, & W \geq \beta, \end{cases} \tag{4}$$

Much can be gained by representing the PMF, F(d) (which we assume to be a strictly monotonic increasing function) as the distribution function of a random variable **D**:

$$F(d) = Pr\{\mathbf{D} \leq d\}. \tag{5}$$

Sternberg and Knoll (1973, Section II) showed that for the decision rule expressed by Eq. 4,

$$\mathbf{D} = \mathbf{R}_x - \mathbf{R}_y + \beta. \tag{6}$$

They also showed that the decision function G need not be a step function; as long as it is a nondecreasing function it can be regarded as the distribution function of a random variable Δ that is stochastically independent of $\mathbf{R}_x - \mathbf{R}_y$, and Eg. 6 can be generalized[4] as

$$\mathbf{D} = \mathbf{R}_x - \mathbf{R}_y + \Delta. \tag{7}$$

The PMF can thus be expressed additively in terms of sensory ($\mathbf{R}_x - \mathbf{R}_y$ and decision (Δ) processes. That is, thinking of the PMF as the distribution function of a random variable, it can be expressed as the convolution of the distribution of arrival-time differences (or, more generally, of differences of the sensory feature being judged) and a stochastically independent distribution that represents the decision process. Given this representation, it follows that the first, second, and third moments of the PMF (and higher cumulants as well) can be written as sums of the corresponding cumulants of $\mathbf{R}_x - \mathbf{R}_y$ and Δ. This *moment-additivity* property means, for example, that if the decision process remains fixed, a change in the variance of $\mathbf{R}_x - \mathbf{R}_y$ is reflected as an equal change in the variance of the PMF.

Many plausible decision mechanisms generate nondecreasing functions G that are not step functions; one possibility, for example, is a rule like the deterministic one but with a criterion β that fluctuates from trial to trial. Let **B** represent the fluctuating criterion. Then, since $G(W) \equiv Pr\{''t_x < t_y'' \mid \mathbf{U}_y - \mathbf{U}_x = W\} = Pr\{\mathbf{B} \leq W\}$, the decision function G can be identified with the (cumulative) distribution of criterion values across trials. Thus G must be a nondecreasing function, Δ can be identified with **B**, and, as described by Eq. 7, **D** is the convolution of the distribution of arrival-latency differences with this criterion distribution.

---

4. In fact there is no need in this generalization for G to be a nondecreasing function. (It would seem that few plausible models would violate this condition; but see Section 8 for one such model.) If G rises from zero to one nonmonotonically it cannot be regarded as a distribution function of an actual random variable. Nevertheless F is given by convolution of G with the distribution function of $\mathbf{R}_x - \mathbf{R}_y$, and the formal calculation used to determine the cumulants of Δ still produce quantities that contribute additively to the corresponding moments of **D**. (Under these conditions F may or may not be monotonic, depending on details of $\mathbf{R}_x - \mathbf{R}_y$ and Δ.) Since such a Δ contributes in the same way to D as an actual random variable, it can be described as a "virtual random variable".

4. A Family of Psychometric Functions from Ratings

The arguments outlined above can be generalized to the *family* of PMFs generated by partitioning an ordered set of confidence ratings (or, more generally, an ordered set of response categories) at different levels. Suppose the observer uses ratings $\mathbf{A} = 1, 2, \ldots, n$, with $\mathbf{A} = 1$ representing high confidence that S_x did not occur before S_y (typically associated with large negative $\mathbf{U}_y - \mathbf{U}_x$ values) and $\mathbf{A} = n$ corresponding to high confidence that S_x did occur before S_y (typically associated with large positive $\mathbf{U}_y - \mathbf{U}_x$ values). Then the PMF of Eq. (1) can be replaced by a family of $n - 1$ such functions, $F_i(d)$, $i = 1, 2, \ldots, n - 1$, with

$$F_i(d) \equiv Pr\{\mathbf{A} > i \mid d\}. \tag{8}$$

The function $F_i(d)$ results from partitioning the ratings $\mathbf{A}$ into $0 < \mathbf{A} \le i$ and $i < \mathbf{A} \le n$, $i = 1, 2, \ldots, n - 1$. Again, $F_i(d)$ can be regarded as the distribution of a random variable $\mathbf{D}_i$. That is, by analogy with Eq. 5,

$$F_i(d) \equiv Pr\{\mathbf{D}_i \le d\}. \tag{9}$$

Given the rating procedure, where the i^{th} partition of the ratings is associated with a distinct decision process, represented by Δ_i, and a distinct PMF, F_i, Eq. (7) becomes

$$\mathbf{D}_i = (\mathbf{R}_x - \mathbf{R}_y) + \Delta_i. \tag{10}$$

Let $\mu_r(\Delta_i)$ be the r^{th} moment of Δ_i. Because of the invariance of $(\mathbf{R}_x - \mathbf{R}_y)$ across the differences among the $\{\Delta_i\}$ associated with different members $\{\mathbf{D}_i\}$ of the family of PMFs, together with moment additivity for stochastically independent random variables, moment differences $\mu_r(\Delta_i) - \mu_r(\Delta_j)$ among the decision processes will produce equal moment differences $\mu_r(\mathbf{D}_i) - \mu_r(\mathbf{D}_j)$ among the PMFs, and thus be observable.

Because, for all d, $Pr\{\mathbf{A} > i\} \ge Pr\{\mathbf{A} > i + 1\}$, it follows from the definition in Eq. 8 that the F_i are characterized by a dominance property:

$$F_i(d) \ge F_{i+1}(d), \ -\infty < d < \infty, \ i = 1, 2, \ldots, n - 2. \tag{11}$$

That is, the larger the rating index i the (lower, and) further to the right on the d-axis the PMF lies. To say more about relations among the PMFs requires a model.

In the following five sections we describe different models of the decision process as examples, and consider their implications.

5. Implications of a Model with Deterministic Decisions

We consider first the generalization to the confidence-rating procedure, diagrammed in Figure 1, of the deterministic decision rule that was described by Eq. 4.

Let β_i , i= 1, 2,..., n-1 be a set of ordered and fixed criteria on the continuum of arrival-time difference $U_y - U_x$, with $\beta_i \le \beta_{i+1}$ and $i = 1, 2, \ldots, n - 2$. To simplify statements, define $\beta_0 = -\infty$ and $\beta_n = \infty$. Then the conventionally assumed decision rule for the rating procedure (e.g., Green & Swets, 1966, Section 2.4) can be stated as follows:

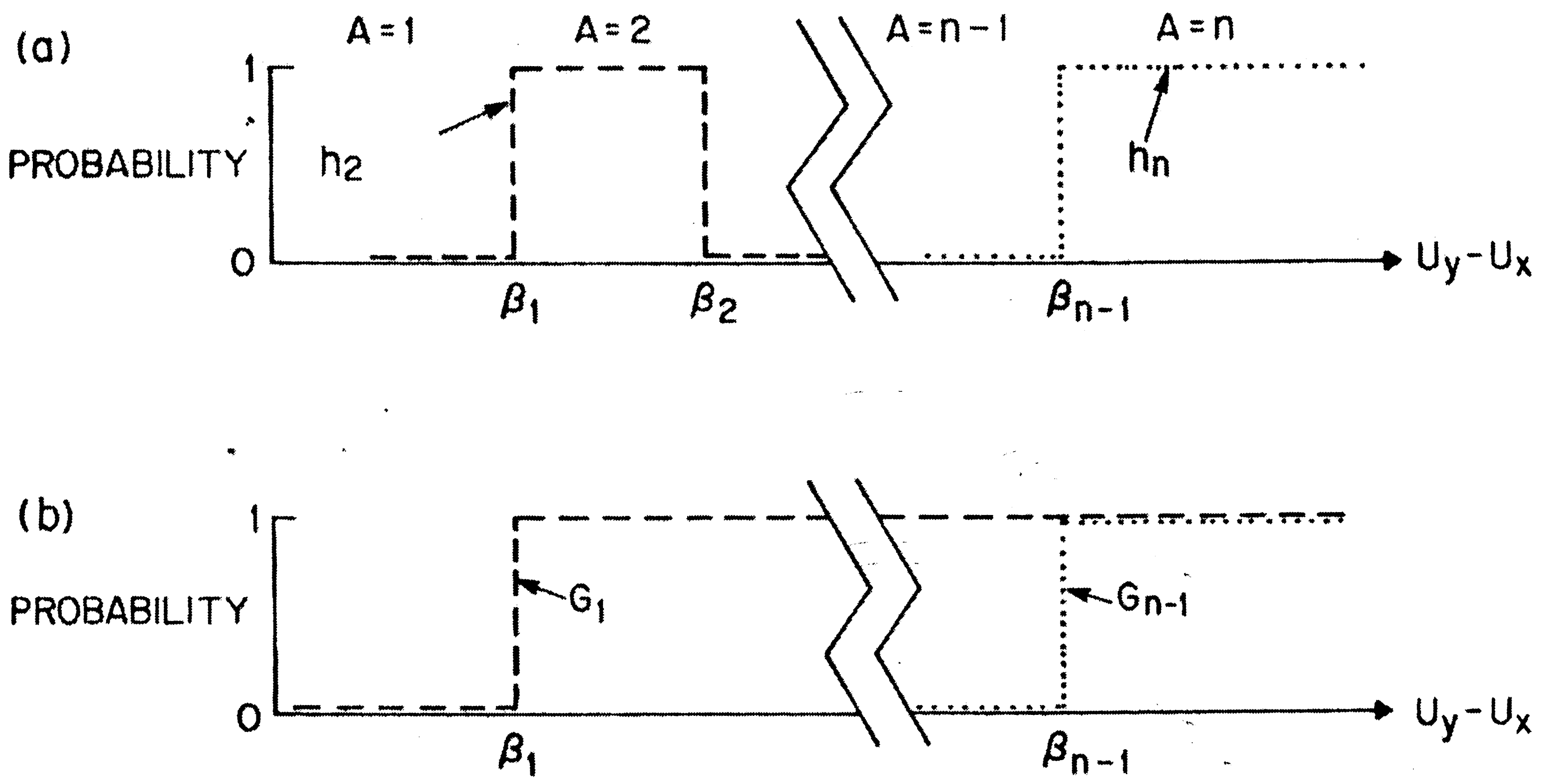

Figure 1.

(a) Relation between the arrival-time difference criteria β_i and ratings **A** in a generalization of the deterministic decision rule, and examples of rating-probability functions. As described in Eq. 14 the probability of rating $\mathbf{A} = i$ is represented by a function $h_i(U_y - U_x)$ that is unity for $\beta_{i-1} \leq U_y - U_x < \beta_i$ and zero elsewhere. For example, the rating $\mathbf{A} = 2$ is produced with probability 1.0 when the perceived arrival-time difference falls between criterion levels β_1 and β_2, and with probability = 0.0 when the percept is outside these limits; the rating $\mathbf{A} = n$ is produced when the percept falls above β_{n-1}.

(b) Examples of decision functions G_i for the same model. For example, when the percept exceeds β_1, the rating is greater than $\mathbf{A} = 1$; when the percept exceeds β_{n-1}, the rating is greater than $\mathbf{A} = n - 1$.

$$\mathbf{A} = i \;\; iff\, \beta_{i-1} \le \mathbf{U}_y - \mathbf{U}_x < \beta_i \;\;,\;\; i = 1, 2, \ldots, n \quad, \tag{12}$$

or

$$\mathbf{A} \le i \;\; iff \;\; \mathbf{U}_y - \mathbf{U}_x \le \beta_i \;\;,\;\; i = 1, 2, \ldots, n-1 \quad. \tag{13}$$

Thus, if we define a set of *rating-probability functions*, $h_i(W)$, each giving the probability of a particular rating as a function of the arrival-time difference, $W = U_y - U_x$,

$$h_i(W) = Pr\{A = i \mid \mathbf{U}_y - \mathbf{U}_x = W\},\;\; i = 1, 2, \ldots, n, \tag{14}$$

then the deterministic decision rule requires that

$$h_i(W) = \begin{cases} 1, & \beta_{i-1} \le W < \beta_i \\ 0, & elsewhere \end{cases} \quad, \;\; i = 1, 2, \ldots, n. \tag{15}$$

Now, by analogy with Eq. 3 we can define [5] a set of decision functions, one associated with each rating from A = 1 to A = n-1:

$$G_i(W) = Pr\{\mathbf{A} > i \mid \mathbf{U}_y - \mathbf{U}_x = W\},\;\; i = 1, 2, \ldots, n-1. \tag{16}$$

Note that

$$G_i(W) = \sum_{j=i}^{n} h_j(W), \quad -\infty < W < \infty,\;\; i = 1, 2, \ldots, n-1. \tag{17}$$

As in the case of the F_i, it follows from the definition of the G_i, that they, also, are characterized by a dominance property:

$$G_i(W) \ge G_{i+1}(W), \quad -\infty < W < \infty, \quad i = 1, 2, \ldots, n-1. \tag{18}$$

By analogy with Eq. 4, it follows from Eqs. 15 and 17 that for the deterministic decisions model of Eq. 12 the G_i are all step functions,

$$G_i(W) = \begin{cases} 0, W < \beta_i \\ 1, W \ge \beta_i \end{cases} \quad, \tag{19}$$

so that the random variables they represent are all constants. But from Eqs. 2, 8, and 15,

$$F_i(d) = Pr\{\mathbf{U}_y - \mathbf{U}_x \ge \beta_i\} = Pr\{\mathbf{R}_y - \mathbf{R}_x + d \ge \beta_i\} = Pr\{\mathbf{R}_x - \mathbf{R}_y + \beta_i \le d\}\;, \tag{20}$$

so that by analogy to Eq. 6, we can represent the $\mathbf{D}_i$ as follows:

$$\mathbf{D}_i = \mathbf{R}_x - \mathbf{R}_y + \beta_i \;, i = 1, 2, \ldots, n-1. \tag{21}$$

Thus for the deterministic decisions model the $\mathbf{D}_i$ represent random variables that differ from each other only because they involve different additive constants β_i; in terms of the $F_i(d)$ we have

$$F_i(d + \beta_i) = F_j(d + \beta_j), \tag{22}$$

showing that *for the deterministic decisions model the PMFs are parallel — i.e., differ only by translation on the d-axis.* That is, for $r > 1$, the $\mu_r(F_i)$ are the same for all i. Note

---

5. Without restrictions on the h_i, this definition, which leads to the relations between the G_i and the h_i expressed in Eq. 17, may produce one or more G_i that are nonmonotonic. As discussed in footnote 4, however, our analysis does not require monotonicity of the G_i.

that this result does not depend on the distributions of the arrival latencies $\mathbf{R}_x$ and $\mathbf{R}_y$.

Perhaps our intuition that the PMFs in a family should be parallel is based on an implicit belief in the deterministic decisions model.

6. Implications of a General Probabilistic Decisions Model

Generalizing further to nondeterministic decision rules (where the G_i defined in Eq. 16 are not all step functions) we have by analogy with Eq. 7,

$$\mathbf{D}_i = \mathbf{R}_x - \mathbf{R}_y + \Delta_i, \quad i = 1, \ldots, n-1, \tag{23}$$

with G_i defined[6] as the distribution function of Δ_i . Just as in the case of the general independent-channels model for the binary-choice experiment (Sternberg & Knoll, 1973, Section IIC), further specification of the G_i beyond the dominance property in Eq. 18 follows from particular models of the decision mechanism; three examples of such models are described in the sections below. But the formulation of the general model in Eq. 23 allows us to state the restriction on the decision functions G_i that is required if the PMFs F_i are to be parallel. Equation 23 makes it clear that in order for the distributions of $\mathbf{D}_i$ and $\mathbf{D}_j$ to differ by translation only ($\mathbf{D}_i \approx \mathbf{D}_j + K$), the distributions of Δ_i and Δ_j must differ by translation only ($\Delta_i \approx \Delta_j + K$).[7] That is, *for the general decisions model the F_i are parallel on the d-axis if and only if the G_i are parallel on the $U_y - U_x$ axis.*[8]

7. Implications of a Threshold Model

In one of the simplest nondeterministic decisions models, there is a threshold interval, centered around $U_y - U_x = 0$, within which different $\mathbf{U}_y - \mathbf{U}_x$ values cannot be discriminated from each other. (See Model 3 in Sternberg & Knoll, 1973, Section IIC.)[9]

---

6. Again, as discussed in footnotes 4 and 6, if G_i is nonmonotonic, it must be regarded as a distribution function of a "virtual" rather than "actual" random variable, but the arguments go through in the same way.

7. Here and elsewhere in this paper, "$\approx$" means "has the same distribution as".

8. For certain "pathological" distributions of $\mathbf{R}_x - \mathbf{R}_y$, parallel F_i may not require parallel G_i. Strictly speaking, then, such distributions must be excluded. This can be done by adding the condition that $\mathbf{R}_x - \mathbf{R}_y$ have a finite mean — a condition that will be satisfied in all cases of interest.

9. This model differs from the general threshold model considered by Ulrich (1987), in which the threshold can fluctuate and need not be centered around $U_y - U_x = 0$.

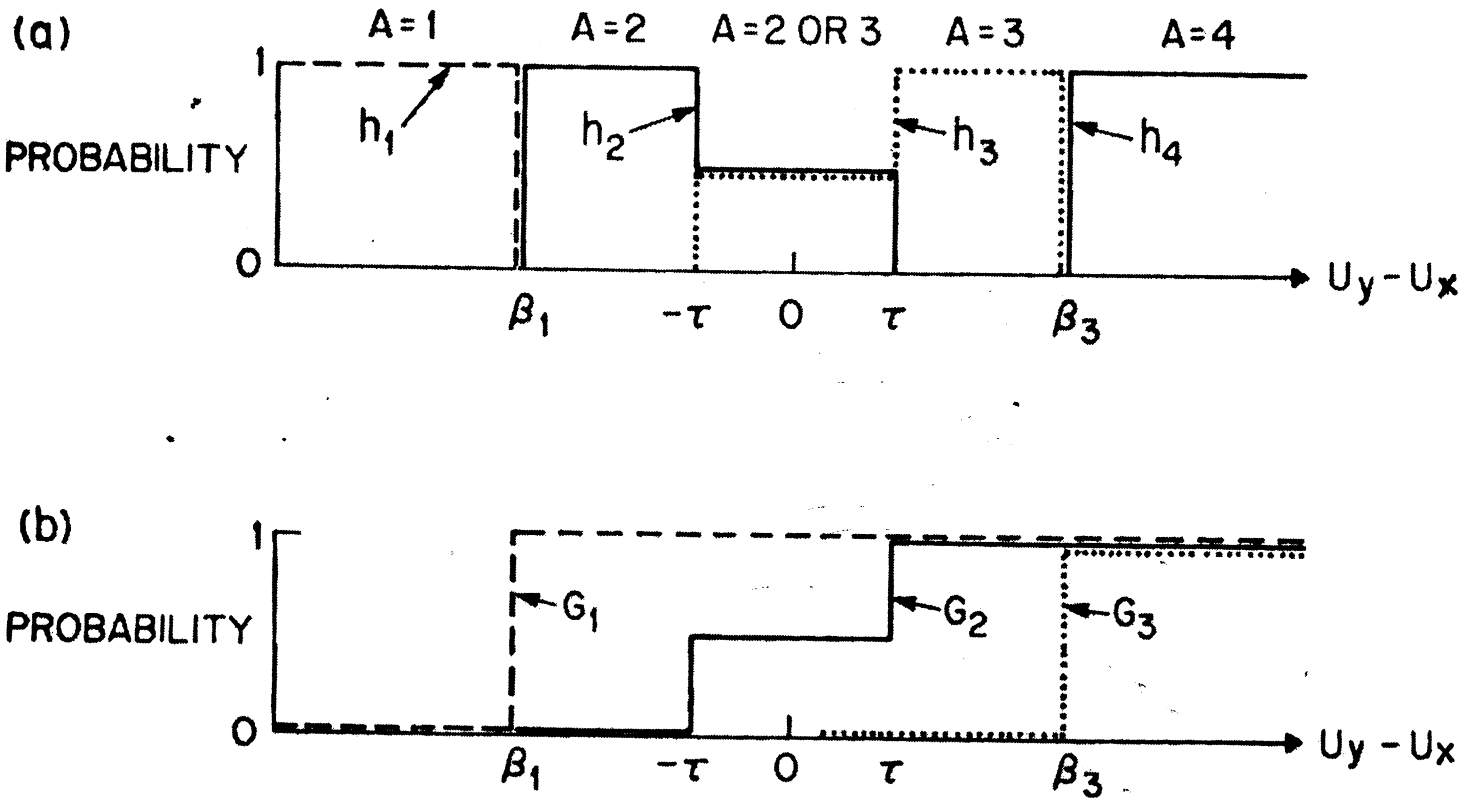

Figure 2.

(a) Rating-probability functions for a threshold model in which the threshold lies between $-\tau$ and $+\tau$.

(b) Decision functions for the same model.

Table 1.
Threshold Model:
Response Probabilities in $U_y - U_x$ intervals

Interval	h_1	h_2	h_3	h_4	G_1	G_2	G_3
$(-\infty, \beta_1)$	1	0	0	0	0	0	0
$(\beta_1, -\tau)$	0	1	0	0	1	0	0
$(-\tau, \tau)$	0	0.5	0.5	0	1	0.5	0
(τ, β_3)	0	0	1	0	1	1	0
$(\beta_3, +\infty)$	0	0	0	1	1	1	1

In the example shown in Figure 2 and described in Table 1, there are four ratings, and the threshold lies within the interval on the $U_y - U_x$ axis covered by the two middle ratings, $\mathbf{A} = 2$ and $\mathbf{A} = 3$; within the threshold region $(-\tau < U_y - U_x < \tau)$ these ratings are equiprobable, while outside the threshold region the deterministic decisions model applies.

Table 2.
Threshold Model: G_i Distributions

Value	Probability		
	G_1	G_2	G_3
β_1	1	0	0
$-\tau$	0	0.5	0
τ	0	0.5	0
β_3	0	0	1

Table 3.
Threshold Model: G_i Moments

Moment	Distribution		
	G_1	G_2	G_3
μ_1'	β_1	0	β_3
μ_2	0	τ^2	0
μ_3	0	0	0

Examination of Tables 2 and 3 and of plots of the G_i in Figure 2b shows that whereas Δ_1 and Δ_3 are constants, with zero variance (and zero third moment), Δ_2 has a two-point distribution with variance τ^2 (and also zero third moment). *Hence F_1, and F_3 must be parallel, while the middle PMF, F_2, must be flatter than the others,* with its variance larger by τ^2 than the variance of F_1 and F_3. Because the decision process contributes nothing to them, the third moments of F_1, F_2, and F_3 are due entirely to the contribution from $\mathbf{R}_x - \mathbf{R}_y$, and must therefore be equal; to the extent that it is plausible that the distribution of $\mathbf{R}_x - \mathbf{R}_y$ is symmetric, the third moments should equal zero.[10]

10. Because $\mu_2(F_2)$ is greater than $\mu_2(F_1)$ and $\mu_2(F_3)$, the standardized third moment, $\mu_3/\mu_2^{1.5}$, a measure of skewness, will be smaller for F_2 than for F_1 or F_3.

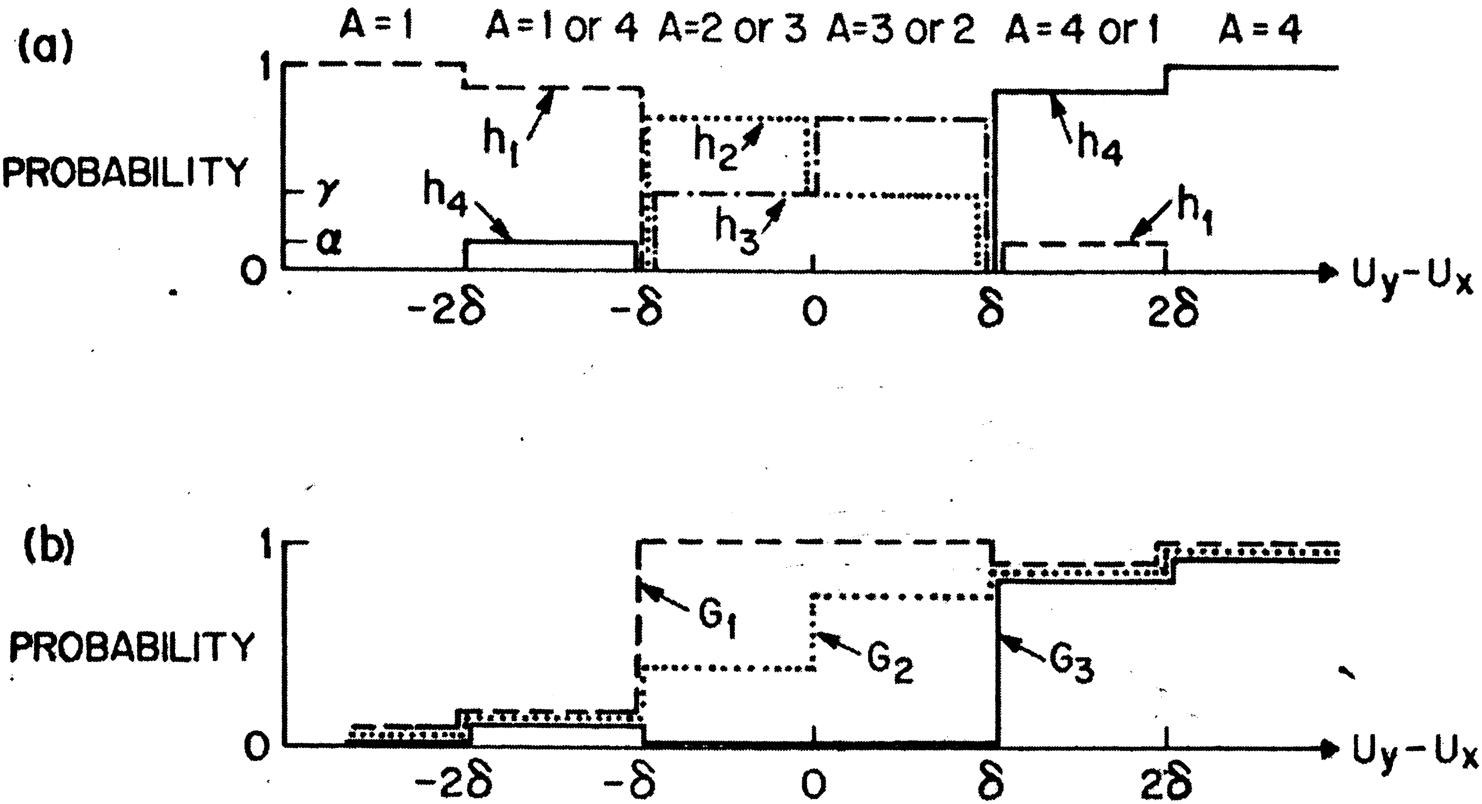

Figure 3.

(a) Rating-probability functions for the model described in Section 8. See text for explanation.

(b) Decision functions for the same model. Note that G_1 and G_3 are not monotonic.

8. Implications of a Model Where a Confident and Correct Report of Successiveness May be Associated with an Erroneous Report of Order

Here we consider implications of a "successiveness model", in which the mechanisms subserving the perception of order might be different and to some extent independent of those subserving the perception of successiveness. Given that the perception of the order of two events requires discrimination of their identities, whereas the perception of successiveness might not, the separation of these aspects of temporal-order judgments seems reasonable to consider. Such a model is analogous to the one considered by Wickelgren (1969) for comparison of pitches, in which the degree of similarity between the two pitches is discriminated by a different mechanism from the one that discriminates the direction of any difference.

This model might be suitable for one of the procedures used by Allan (1975a), in which observers judged the relative offset times of a tone and a light and provided one of four judgments: "successive and tone first" (A=1), "simultaneous and tone first" (A=2), "simultaneous and light first" (A=3), and "successive and light first" (A=4). When $\delta \le |U_y - U_x| \le 2\delta$ the observer can correctly and confidently judge S_x and S_y to be successive (**A** = 1 or **A** = 4), while misperceiving their order (**A** = 4) on a fraction $\alpha > 0$ of trials. When $0 \le |U_y - U_x| \le \delta$, the observer is sensitive to the sign of $U_y - U_x$ (**A** = 3 more likely than **A** = 2, when $U_y - U_x > 0$), but the ratings indicate misperception of the order (**A** = 2) on a fraction γ of trials, where $\gamma > \alpha$.

Table 4. Successiveness Model:
Response Probabilities in $U_y - U_x$ intervals

Interval	h_1	h_2	h_3	h_4	G_1	G_2	G_3
$(-\infty, -2\delta)$	1	0	0	0	0	0	0
$(-2\delta, -\delta)$	$1-\alpha$	0	0	α	α	α	α
$(-\delta, 0)$	0	$1-\gamma$	γ	0	1	γ	0
$(0, \delta)$	0	γ	$1-\gamma$	0	1	$1-\gamma$	0
$(\delta, 2\delta)$	α	0	0	$1-\alpha$	$1-\alpha$	$1-\alpha$	$1-\alpha$
$(2\delta, +\infty)$	0	0	0	1	1	1	1

A simple example of a set of rating-probability functions h_i that might arise from such a model in an experiment with four different ratings is shown in Figure 3 and listed with the corresponding G_i in Table 4. Whereas **A** = 1, for example, is most likely when $U_y - U_x < -\delta$, it also occurs with a low probability (α) when $\delta \le U_y - U_x < 2\delta$. Similarly, **A** = 2 is most likely when $-\delta \le U_y - U_x < 0$, but also occurs with a low probability (γ) when $0 \le U_y - U_x < \delta$. (We have used rating-probability functions h_i that are constant within intervals on the $U_y - U_x$ axis as well as intervals that are of equal width (δ) for illustrative purposes; in a more plausible model both these restrictions might be relaxed.) Unlike the G_i shown in Figures 1b and 2b, not all the G_i generated by the present model and shown in Figure 3b are nondecreasing functions; instead, G_1 and G_3 are nonmonotonic and as a result cannot correspond to actual random variables.[11] The

11. See footnotes 4 and 6.

distribution to which G_1 corresponds, for example, included in Table 5, would have negative probability $(-\alpha)$ at $U_y - U_x = \delta$. Nonetheless for present purposes we can apply the usual operations to derive the moments of the $\{G_i\}$, listed in Table 6, that combine additively with those of $\mathbf{R}_x - \mathbf{R}_y$ to produce the corresponding moments of the PMFs $\{F_i\}$, as implied by Eq. 23.

Table 5. Successiveness Model:
G_i Distributions

Value	Probability		
	G_1	G_2	G_3
-2δ	α	α	α
$-\delta$	$1-\alpha$	$\gamma-\alpha$	$-\alpha$
0	0	$1-2\gamma$	0
δ	$-\alpha$	$\gamma-\alpha$	$1-\alpha$
2δ	α	α	α

Table 6.
Successiveness Model: G_i Moments

Moment	Distribution		
	G_1	G_2	G_3
μ_1'	$-\delta$	0	δ
μ_2	$6\delta^2\alpha$	$6\delta^2\alpha + 2\delta^2\gamma$	$6\delta^2\alpha$
μ_3	$18\delta^3\alpha$	0	$-18\delta^3\alpha$

The results of these calculations, in Tables 5 and 6, show that G_2 has a greater variance (by $2\delta^2\gamma$) than G_1 or G_3, whose variances are equal, which means that F_2 has a greater variance (by $2\delta^2\gamma$) than F_1 or F_3, whose variances are equal. They also show that whereas G_2 is symmetric, G_1 is positively skewed (third moment = $18\delta^3\alpha$) and G_3, is negatively skewed by the same amount. If we assume that the $\mathbf{R}_x - \mathbf{R}_y$ distribution is symmetric, which is often plausible, this statement also applies to the PMFs, F_1, F_2, and F_3. Without this assumption, the model implies that $\mu_3(F_1) > \mu_3(F_2) > \mu_3(F_3)$, and that the magnitude of the two differences is $18\delta^3\alpha$.

Given the moments of the G_i, and the additive moment relations implied by Eq. 23, if the model is valid one can use appropriate combinations of the first three moments of F_1, F_2, and F_3 not only to provide estimates of the decision-function parameters α, γ, and δ, but also to provide estimates of the second and third moments of the sensory component $\mathbf{R}_x - \mathbf{R}_y$. This analysis furnishes a particularly vivid example of the increase gained in the power to decompose sensory and decision processes by using a family of PMFs.

9. Implications of a Model with Fluctuating Criteria

Consider a decision model involving criteria β_i on the $U_y - U_x$ axis, like the deterministic decisions model discussed in Section 5, but permit the criteria to fluctuate from trial to trial so that they become random variables $\mathbf{B}_i$. (For binary-choice data such a model with a single criterion was considered in Section 3.) We assume that on each trial the ordering of criteria, $\beta_i \le \beta_{i+1}$, assumed in Section 5 is preserved. This implies that for any $W = U_y - U_x$, $Pr\{\mathbf{B}_i \le W\} \ge Pr\{\mathbf{B}_{i+1} \le W\}$. Furthermore, since $Pr\{\mathbf{B}_i \le W\} = Pr\{\mathbf{A} > i \mid \mathbf{U}_y - \mathbf{U}_x = W\} = G_i(W)$ we see that not only can the decision function G_i be identified as the distribution function of the criterion $\mathbf{B}_i$ (which, incidentally, requires it to be a nondecreasing function) so that Δ_i and $\mathbf{B}_i$ are the same, but also that the dominance property (Eq. 18) required of the decision functions is guaranteed. The criterion distributions may overlap so long as the "amount" of overlap is not so great as to violate the dominance property. (If the distributions do overlap, however, the requirement of criterion ordering on every trial implies that the $\mathbf{B}_i$ cannot fluctuate independently.)

It should be noted that any model with nondecreasing G_i, can be regarded as equivalent to a model with multiple fluctuating criteria, given that it is reasonable to identify the G_i as criterion distributions. The model described in Section 8, however, is an example of one that cannot be equated in this way, because its G_1 and G_2 are not monotonic functions.

Having identified the G_i with the distributions of fluctuating criteria, we can immediately state the conditions for parallel PMFs (Section 6): *In a model with multiple fluctuating criteria the F_i are parallel if and only if the criterion distributions are identical except for location.* That is, if we define $\mathbf{B}_i^* = \mathbf{B}_i - E(\mathbf{B}_i)$ to be the distribution of the i^{th} criterion adjusted for zero mean, parallel F_i requires that

$$\mathbf{B}_i^* \approx \mathbf{B}_j^* \,,\ 1 \le i, j \le n-1 \quad . \tag{24}$$

How likely are the distributions of multiple criteria to differ only in mean? We are not aware of any discussion of this question, and here we mention only two of the considerations that might bear on it. The distributional identity requires, for example, that the criterion variance be the same for extreme criteria as for middle-range criteria. From a Weber-law viewpoint, on the other hand, one might expect the standard deviation of the criterion distribution to increase linearly with $|d|$, or with $|d - PSS|$ (where PSS is the point of subjective simultaneity). The result would be a generalized bidirectional Weber law: *the PMFs associated with more extreme ratings would be flatter.*

Perhaps a more compelling argument for constraining the criterion distributions arises from the inherent symmetry of the decision aspects of the experimental paradigm: Given a pair of stimuli, their assignment to S_x and S_y is arbitrary. A simple way of describing the consequence is that the series of density functions of the criteria $\mathbf{B}_1, \mathbf{B}_2, \ldots$ on the $U_y - U_x$ axis (i.e., viewed from the low $U_y - U_x$ end) should have the same sequence of shapes as the corresponding series on the $U_x - U_y$ axis (i. e., viewed from the opposite end). This implies that the density function of $\mathbf{B}_i^*$ should be the reflection of the density function of $\mathbf{B}_{n-i}^*$, or

$$\mathbf{B}_i^* \approx -\mathbf{B}_{n-i}^* \;,\ \ i = 1, 2, \ldots, n-1. \tag{25}$$

But in order that the F_i differ by translation only, Eq. 24 must also be satisfied. Equations 24 and 25 together imply that the criterion distributions are symmetric about their means: $\mathbf{B}_i^* \approx -\mathbf{B}_i^*$, i = 1, 2, . . . , n-1 . Conversely, given Eq. 25, *any asymmetry in the criterion distributions will produce shape differences among the F_i.*

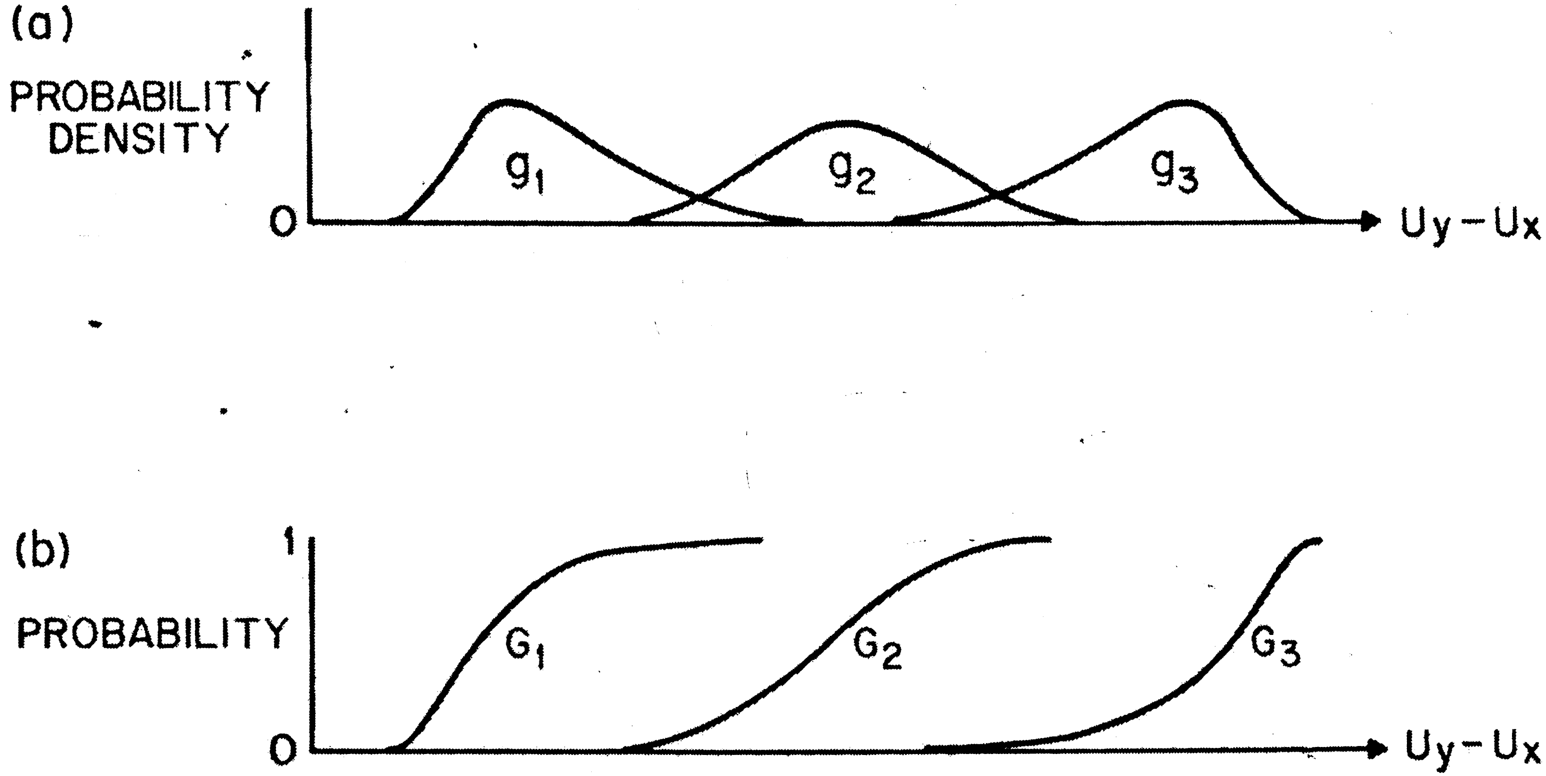

Figure 4.

Hypothetical criterion distributions with (a) density functions that may be asymmetric in shape, but satisfy the requirement (imposed by experimental symmetry) that the density g_1 of B_1^* is the reflection of the density g_3 of B_3^*, and (b) the nonparallel distribution functions to which they correspond.

Suppose, for example, that the distribution of the lowest criterion is skewed toward high $U_y - U_x$ values, as illustrated in Figure 4. Then the argument from symmetry of the experiment implies that the distribution of the highest criterion should be skewed toward low $U_y - U_x$ values (i.e., high $U_x - U_y$ values). The consequence of the arrangement in the figure is that the PMFs F_i would be more negatively skewed with larger i. In Section 5 we showed that observation of such shape differences among the F_i would require rejection of a model with fixed criteria; here we have shown that they are consistent with a model in which the criteria are permitted to fluctuate from trial to trial.

10. Families of Psychometric Functions from Three Experiments

To exemplify the inferential possibilities of the multiple-PMF method, we provide, as examples, families of PMFs from three experimental procedures that have been used in the study of temporal-order perception, procedures that used three or four ordered response alternatives. The PMFs we have selected from each report are qualitatively similar to the other PMFs in that report. However, they are not ideal for our purposes, as they include more instances of non-monotonicity than we would like, as several fail to span the full range of proportions from zero to one, and as they may contain lapses of attention (and associated guessing) for which we have not corrected. In the absence of better experiments, where necessary to deal with the first two inadequacies, we have extended the PMF range to span the full (0,1) interval, and have rendered the PMFs monotonic. The resulting adjusted PMFs are called "$adjF_i$", and are tabulated along with the PMFs, $\{F_i\}$, as measured.

In one of the procedures used by Allan (1975a) observers judged the offset times of a tone and a light, making a successiveness judgment ("simultaneous" or "successive") as well as an order judgment. The four combined judgments were, then, "successive and tone first" (A=1), "simultaneous and tone first" (A=2), "simultaneous and light first" (A=3), and "successive and light first" (A=4). (We are treating A=1 and A=4 as high confidence order judgments, and A=2 and A=3 as low confidence order judgments.) These permit defining three PMFs, $F_1 = Pr\{A > 1\}$, $F_2 = Pr\{A > 2\}$, and $F_3 = Pr\{A > 3\}$, each giving a rating proportion as a function of the time difference, d_i (tone offset time - light offset time). The values of F_1, F_2, and F_3 for Observer T. M. are shown in the first row of each of the three parts of Table 7, for each offset-time difference, d_i.

Table 7: Allan (1975a), Observer T.M.
(Panel A of Figure 5)

d (ms)	(-125)	-100	-75	-50	-25	0	+25	+50	+75	+100	(+125)
Trials		96	96	96	96	384	96	96	96	96	
F_1		0.088	0.208	0.413	0.635	0.860	0.915	0.935	0.905	0.947	
$adjF_1$	**0.000**	0.088	0.208	0.413	0.635	0.860	0.915	**0.918**	**0.929**	0.947	**1.000**
F_2		0.045	0.135	0.315	0.420	0.460	0.610	0.730	0.840	0.905	
$adjF_2$	**0.000**	0.045	0.135	0.315	0.420	0.460	0.610	0.730	0.840	0.905	**1.000**
F_3		0.003	0.073	0.138	0.135	0.080	0.135	0.315	0.665	0.892	
$adjF_3$	**0.000**	0.003	0.073	**0.095**	**0.118**	**0.126**	0.135	0.315	0.665	0.892	**1.000**

To get these data into the form of a distribution function, the proportions $\{p_i\}$ have to be extended to zero and one,[12] and rendered strictly monotonic.[13] The results of making these adjustments are shown in the second rows of each of the three parts of Table 7, as $adjF_1$, $adjF_2$, and $adjF_3$, and are plotted in Panel A of Figure 5. In this table, as well as tables 8 and 9, entries that have been created (to extend the PMFs to proportions zero and one) or adjusted (to achieve monotonicity) are printed in boldface.

In the procedure used in a second study by Allan (1975b), observers judged the order of the offset times of a light and a tone, and also made a two-level confidence judgment ("certain" or "uncertain"). The four responses were therefore "tone first" and "certain" (A=1), "tone first" and "uncertain" (A=2), "light first" and "uncertain" (A=3), and "light first" and "certain" (A=4). Again, these permit defining three PMFs, $F_1 = Pr\{A > 1\}$, $F_2 = Pr\{A > 2\}$, and $F_3 = Pr\{A > 3\}$, each as a function of the time difference, d. The values of F_1, F_2, and F_3 for Observer N. C. are shown in Table 8, and plotted in Panel B of Figure 5. In this case monotonizing was needed only for F_1.

12. If a PMF fails to cover the full range of proportions from 0.0 to 1.0, one explanation is that the range of d-values was too small. (In a better experiment, a sufficiently large range of d-values would be used to avoid this problem.) A plausible alternative reason is that the observer was inattentive on some trials, and made a response — a "guess" — that was independent of the stimulus, except perhaps when the discrimination was especially easy. (It seems possible that if attention is "elsewhere", and is returned to the task on presentation of the stimuli to be judged, but with a delay, the percept, degraded by the delay, may be useful for an easy discrimination, but not for a difficult one, as perhaps suggested by F_1 and F_3 in Table 7, and by F_1 and F_2 in Table 9.) Such lapses of attention cause estimation difficulty even when the form of the PMF is known and only a threshold needs to be estimated (Green, 1995). Here the form is unknown and is the object of study, making it more important to use suitably timed warnings, performance-based payoffs, adequate practice, and other methods to minimize their occurrence.

13. The observed PMF is assumed to be an estimate of a strictly monotonic PMF. Monotonizing the observed sequence of proportions F_i was done using the R function "cirPAVA", in the R-package "cir" (A. P. Oron, 2023). One undesirable effect this has is to obscure flat regions of the PMFs that may be due to lapses of attention.

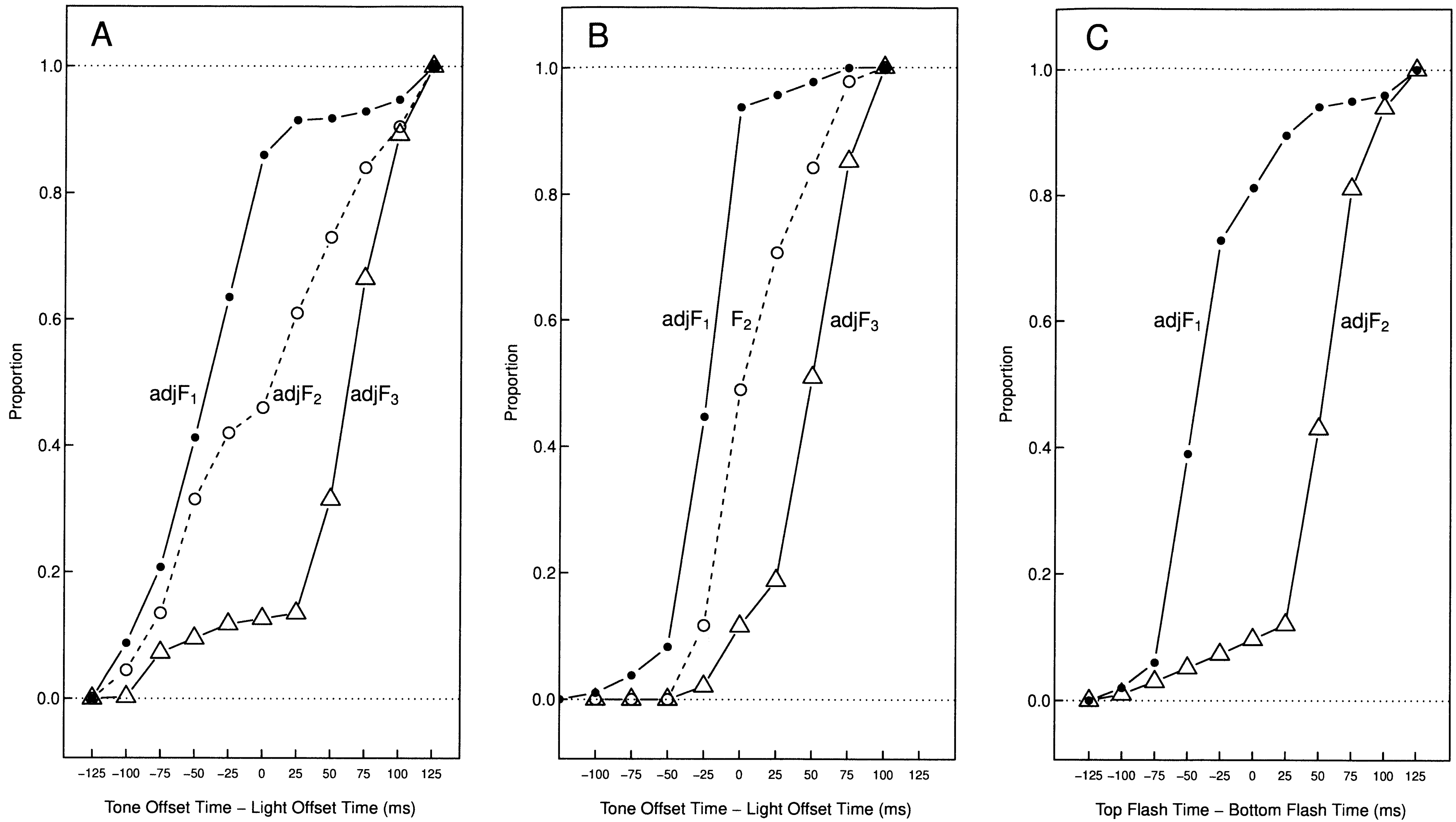

Figure 5.

Extended and monotonic versions of three families of psychometric functions:
Panel A: Observer TM in Allan (1975a). The plotted values are those called $adjF_1$,$adjF_2$, and $adjF_3$ in Table 7.
Panel B: Observer NC in Allan (1975b). The plotted values are those called $adjF_1$, F_2, and F_3 in Table 8.
Panel C: Observer GU, low-intensity condition in Ulrich (1987). The plotted values are those called $adjF_1$ and $adjF_3$ in Table 9.

Table 8: Allan (1975b), Observer N.C.
(Panel B of Figure 5)

d (ms)	(-125)	-100	-75	-50	-25	0	+25	+50	+75	+100
Trials		32	32	32	32	128	32	32	32	32
F_1		0.021	0.010	0.083	0.447	0.938	0.969	0.969	1.000	1.000
$adjF_1$	**0.000**	**0.010**	**0.038**	0.083	0.447	0.938	**0.959**	**0.979**	1.000	1.000
F_2		0.000	0.000	0.000	0.117	0.490	0.708	0.844	0.979	1.000
F_3		0.000	0.000	0.000	0.021	0.116	0.188	0.510	0.854	1.000

In Ulrich's (1987) "ternary response" procedure, observers were shown two brief flashes, one above the other, and judged whether the bottom flash was first, the flashes were "simultaneous", or the top flash was first. We used the data from Observer G. U. in the low-intensity condition. For comparison to Allan's observers, we noted that the frequency of "simultaneous" judgments by Ulrich's observer (37%) was similar to the sums of the frequencies of the two middle judgments by Allan's observers: 38% by T.M. in Allan (1975a) and 31% by N.C. in Allan (1975b). For this reason we think of Ulrich's "simultaneous" judgment as combining A=2 and A=3, with "bottom flash first" corresponding to A=1, and "top flash first" corresponding to A=4. The two PMFs are then $F_1 = Pr\{A > 1\}$, and $F_3 = Pr\{A > 3\}$, each as a function of the time difference, d. The results for Ulrich's Observer G. U. in the low-intensity condition are shown in Table 3, and plotted in Panel C of Figure 5. In this case, monotonizing was necessary for both PMFs.

Table 9: Ulrich (1987), Observer G.U., Low-Intensity Condition
(Panel C of Figure 5)

d (ms)	(-125)	-100	-75	-50	-25	0	+25	+50	+75	+100	(+125)
Trials		100	100	100	100	400	100	100	100	100	
F_1		0.020	0.060	0.390	0.730	0.950	0.940	0.940	0.910	0.960	
$adjF_1$	**0.000**	0.020	0.060	0.390	0.730	**0.812**	**0.894**	0.940	**0.950**	0.960	**1.000**
F_3		0.010	0.030	0.110	0.060	0.050	0.120	0.430	0.810	0.940	
$adjF_3$	**0.000**	0.010	0.030	**0.052**	**0.073**	**0.097**	0.120	0.430	0.810	0.940	**1.000**

As the predictions from some of the models described above are in terms of moments of PMFs, it is useful to have a method for estimating these moments. One way to do this is to use a modified form of the non-parametric Spearman-Karber method (Spearman, 1908, Epstein & Churchman, 1944; Church & Cobb, 1973; Sternberg, Knoll, & Zukofsky, 1982). According to this method, the estimated r^{th} raw moment is given by

$$\hat{\mu}'_r = \frac{1}{r+1} \sum_{i=1}^{k+1} (p_i - p_{i-1}) \left[\frac{s_i^{r+1} - s_{i-1}^{r+1}}{s_i - s_{i-1}} \right] , \tag{25}$$

where $\{s_i\}$ are the stimulus values (in this case, the $\{d_i\}$), and $\{p_i\}$ are the corresponding proportions. In a thorough evaluation of this method, Miller & Ulrich (2001) have shown that it is accurate in estimating the mean and variance of a PMF — sufficiently accurate to be superior to probit analysis in situations where probit analysis is appropriate — but that the estimate it provides of the standardized third central moment, $\hat{\mu}_3/\hat{\mu}_2^{3/2}$, a measure of skewness, has the correct sign but may be an underestimate.[14]

In an alternative "cdf-sample" method, we treated the PMF as a (cumulative) distribution function, and generated a sample associated with that distribution. We did this by interpolating closely-spaced points $d_1, d_2, \ldots, d_n$ and corresponding proportions $p_1, p_2, \ldots, p_n$ in the PMF, where $p_1 = 0$ and $p_n = 1$. In our implementation, $d_{i+1} - d_i = 1\ ms$. We then generated a subsample of d-values for each (d_i, d_{i+1}) interval, distributed uniformly in that interval, with the size of that subsample (approximately) proportional to the $(p_{i+1} - p_i)$ difference. To ensure sufficient accuracy for this (integer) approximation, we used a large multiplier of the difference. Thus, the size of the i^{th} subsample was the rounded value of $10^6 \times (p_{i+1} - p_i)$. The full sample was created by concatenating the subsamples; moments and other statistics were then determined from the full sample. Results of these computations, averaged over results of the two methods, are shown in Table 10.

Table 10: Moment Estimates from Two Methods

measure	$\hat{\mu}'_1/10$	$\hat{\mu}_2/10^3$	$\hat{\mu}_3/10^4$
Family 1			
$adjF_1$	-3.53	2.85	+16.31
$adjF_2$	+0.10	4.49	-2.51
$adjF_3$	+5.20	2.77	-22.12
Family 2			
$adjF_1$	-2.38	0.70	+0.22
F_2	+0.90	1.11	+1.84
F_3	+4.53	0.97	-2.02
Family 3			
$adjF_1$	-3.14	2.10	+13.91
$adjF_3$	+4.85	1.78	-11.76

The Spearman-Karber and cdf-sample methods gave results that differ by a mean of 0.03%.[15]

---

14. Using PMFs similar to those plotted in Figure 5, we found the bias to be negligible: The mean estimates were 99.998% and 100.04% of the true values, based on Spearman-Karber and cdf-sample methods, respectively.

15. For eight tests of the accuracy of such estimates, using PMFs similar to those plotted in Figure 5, but with known moments, the Spearman-Karber method recovered the first three moments with mean absolute errors of 0.0005%, 0.0016%, and 0.0034%, respectively. For the cdf-sample method the corresponding percentages are 0.033%, 0.043%, and 0.050%, respectively.

Model Tests. Three of the models we have discussed have testable quantitative properties:

According to the Deterministic Decisions Model (Section 5), the PMFs are parallel, which is clearly false, for all three Families.

According to the Threshold Model (Section 7), (a) F_2 has a larger variance than F_1 or F_3, for which Families 1 and 2 provide evidence; (b) F_1 and F_3 are parallel, falsified by all three families; and (c) the three PMFs have the same third moment, falsified by all three families.

According to the Successiveness Model (Section 8), and assuming that the distribution of $\mathbf{R}_x - \mathbf{R}_y$ is symmetric, which is plausible, (a) the variance of F_2 is greater than that of F_1 or F_3, consistent with Families 1 and 2; (b) F_1 is positively skewed and F_3 negatively skewed, consistent with all three Families. (c) F_2 is symmetric, supported by neither Family 1 nor Family 2. Without assuming the symmetry of $\mathbf{R}_x - \mathbf{R}_y$, (a) remains, (b) is replaced by $\mu_3(F_3) < \mu_3(F_2) < \mu_3(F_1)$, for which positive evidence is provided by families 1 and 3, but negative evidence by family 2, and (c) is deleted.

Given the three observed PMF families we considered, and the three models with testable quantitative properties, if more complete data were found with similar properties, the relations among the moments of family members suggest that the Successiveness Model would be the most promising.

11. Conclusions

Much can be gained at a small cost by enriching the response alternatives in a psychophysical experiment: Relations among the moments of the resulting family of PMFs can be highly informative. To exemplify the inferential possibilities, we described a set of models for judgments of temporal order, and determined their implications. We then used estimates of the moments of the members of three families of observed PMFs to test some of the models.

In answer to the questions with which we began this paper, our intuition that the members of a PMF family should be parallel would be explained if, for example, we believed that the deterministic decisions model (Section 5) were valid. And if the PMFs are not parallel, we have seen from PMF moment comparisons that much can be learned from differences among the spreads and shapes of family members, which enabled us to evaluate the other models we have described.

Some of our findings about the models include the following:

(1) For the deterministic decisions model the psychometric functions F_i are parallel — i.e., differ only by translation on the stimulus axis.

(2) For the general probabilistic decisions model, the PMFs F_i are parallel on the stimulus-axis if and only if the decision functions G_i are parallel on the $U_y - U_x$ axis.

(3) For the threshold model, with four ordered response categories, the middle psychometric function is flatter than the others, which are parallel.

(4) For a model with fluctuating criteria, the F_i are parallel if and only if the criterion

distributions are identical except for location.

(5) For the model of Section 8, where successiveness can be accurately discriminated while order may not be, and there are four ordered response categories, the PMFs F_1 and F_3 are skewed (positively and negatively, respectively), while F_2 has greater variance, but is symmetric.

Based on the observed PMF families we considered, and among the three models we tested quantitatively, it is this last model, the successiveness model, similar to a model proposed for pitch perception by Wickelgren (1969), that seems the most promising.

These findings depend on treating each PMF as the convolution of two stochastically independent distribution functions: the distribution of the sensory difference (here, the arrival time difference), and a decision process represented as a distribution function. Model evaluations made use of the relations among the first three moments of the PMFs within a family, together with the cumulant-additivity property for sums of stochastically independent random variables.

References

Allan, L. G. (1975a) The relationship between judgments of successiveness and judgments of order. *Perception & Psychophysics, 18,* 29-36.

Allan, L. G. (1975b) Temporal order psychometric functions based on confidence-rating data. *Perception & Psychophysics, 18,* 369-372.

Egan, J. P., Schulman, A. I, & Greenberg, G. Z. (1959). Operating characteristics determined by binary decisions and by ratings. *Journal of the Acoustical Society of America, 31,* 768-773.

Eijkman, E., Thijssen, J. M., & Vendrik, J. H. (1966). Weber's law, power law, and internal noise. *Journal of the Acoustical Society of America, 40,* 1164-1173.

Greeen, D. M. (1995). Maximum-likelihood procedures and the inattentive observer. *Journal of the Acoustical Society of America, 97,* 3749-3760.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Miller, J. & Ulrich, R. (2001). On the analysis of psychometric functions: The Spearman-Karber method. *Perception & Psychophysics, 63,* 1399-1420.

Nachmias, J. & Steinman, R. M. (1963). Study of absolute visual detection by the rating-scale method. *Journal of the Optical Society of America, 53,* 1206-1213.

Oron, A. P. (2023). R Package 'cir': Centered isotonic regression and dose-response utilities, Version 2.3.1.

Oron, A.P. and Flournoy, N. (2017). Centered isotonic regression: point and interval estimation for dose-response studies. *Statistics in Biopharmaceutical Research 9,* 258-267.

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Sternberg, S., & Knoll, R. L. (1973). The perception of temporal order: Fundamental issues and a general model. In S. Kornblum (Ed.), *Attention and performance IV.* New York: Academic Press, 1973. Pp. 629-685.

Sternberg, S., Knoll, R. L., & Zukofsky, P. (1982). Timing by skilled musicians. In D. Deutsch (Ed.), *The psychology of music* (pp. 181- 239). New York: Academic Press.

Swets, J. A., Tanner, W. P., Jr., & Birdsall, T. G. (1961). Decision processes in perception. *Psychological Review, 68,* 301 - 340.

Thijssen, J. M. & Vendrik, A. J. H. (1968). Internal noise and transducer function in sensory detection experiments: Evaluation of psychometric curves and of ROC curves. *Perception & Psychophysics, 3,* 387-400.

Ulrich, R. (1987). Threshold models of temporal-order judgments evaluated by a ternary response task. *Perception & Psychophysics, 42,* 224-239.

Wickelgren, W. A. (1969). Associative strength theory of recognition memory for pitch. *Journal of Mathematical Psychology, 6,* 13-61.

Appendix: R Functions for Spearman-Karber and CDF.to.Sample Methods

```
spearkarb <- function(s,p){
#s is vector of stimulus values
#p is corresponding vector of increasing probability values.
#first p must be zero; last p must be 1; otherwise warning.
#p values must increase monotonically, otherwise warning.
step <- median(diff(s))
if(p[1]>0){warning("Extrapolation to zero needed")}
if(p[length(p)]<1) {warning("Extrapolation to one needed")}
#extend stimvals and pvals to ensure extremes
S <- c(min(s)-step,s,max(s)+step)
P <- c(0,p,1)
dP <- diff(P)
if(min(dP)<0) {warning("Non-Monotonic Proportions")}
dS <- diff(S)
ratio <- dP/dS
d2S <- diff(S^2)
d3S <- diff(S^3)
d4S <- diff(S^4)
M1 <- sum(ratio*d2S)/2
M2 <- sum(ratio*d3S)/3
M3 <- sum(ratio*d4S)/4
mean <- M1
var <- M2 - M1^2
m2 <- var
m3 <- M3 - 3*M1*M2 + 2*M1^3
output <- c(mean,var,m3)
names(output) <- c("mean","var","m3")
return(output)
}

cdf.to.sample <- function(stimvals,propvals,multiplier=1000000,sep=1){
interpolated <- approx(stimvals,propvals,xout=seq(from=min(stimvals),
        to=max(stimvals),by=sep))
s.vals <- interpolated[[1]]
props <- interpolated[[2]]
len <- length(s.vals)
samp <- NA
for(kval in 1:(len-1)){
if (props[kval+1] - props[kval] > 0)
        {n.subsamp <- round(multiplier*(props[kval+1] - props[kval]))
        subsamp <- seq(from=s.vals[kval],to=s.vals[kval+1],
                length.out = n.subsamp)
                if( kval < (len-1) )
                {##remove last element if not last subsamp)
                subsamp <- subsamp[-n.subsamp]}
        samp <- c(samp,subsamp)
        }
}
##remove initial entry (NA) in samp
samp <- samp[!is.na(samp)]
return(samp)
}
```