Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Mathematical Psychology 47 (2003) 184–204
Journal of Mathematical Psychology
http://www.elsevier.com/locate/jmp

Thurstonian-type representations for "same-different" discriminations: Deterministic decisions and independent images

Ehtibar N. Dzhafarov

Department of Psychological Sciences, Purdue University, 703 Third Street, West Lafayette, IN 47907-2004, USA

Received 11 July 2001; revised 15 January 2002

Abstract

A discrimination probability function $\psi(\mathbf{x},\mathbf{y})$ obtained in the "same-different" paradigm assigns to every ordered pair of stimuli $(\mathbf{x},\mathbf{y})$ the probability with which they are judged to be different. This function is said to possess the regular minimality property if, for any stimulus pair $(\mathbf{a},\mathbf{b})$,

$$\arg\min_{\mathbf{y}} \psi(\mathbf{a},\mathbf{y}) = \mathbf{b} \Leftrightarrow \arg\min_{\mathbf{x}} \psi(\mathbf{x},\mathbf{b}) = \mathbf{a}.$$

That is, $\mathbf{b}$ is the point of subjective equality for $\mathbf{a}$ if and only if $\mathbf{a}$ is the point of subjective equality for $\mathbf{b}$. If the value of $\psi(\mathbf{a},\mathbf{b})$ across all such pairs $(\mathbf{a},\mathbf{b})$ is not constant, the function is said to possess the nonconstant self-similarity property. A *Thurstonian-type representation* for $\psi(\mathbf{x},\mathbf{y})$ (with independent images and deterministic decisions) is a model in which the two stimuli are mapped into two independent random variables $P(\mathbf{x})$ and $Q(\mathbf{y})$ taking on their values in some "perceptual" space; and the decision whether the two stimuli are different is determined by the realizations of the two random variables in a given trial. Thurstonian-type representations can also be called "random utility" ones, provided one imposes no a priori restrictions on the structure of the perceptual space, the distributions of $P(\mathbf{x})$ and $Q(\mathbf{y})$, or the decision rules used. It is shown that (A) any $\psi(\mathbf{x},\mathbf{y})$ has a Thurstonian-type representation; but (B) if $\psi(\mathbf{x},\mathbf{y})$ possesses the regular minimality and nonconstant self-similarity properties, it cannot have a "well-behaved" Thurstonian-type representation, in which the probability with which $P(\mathbf{x})$, or $Q(\mathbf{y})$, falls within a given subset of the perceptual space has appropriately defined bounded directional derivatives with respect to $\mathbf{x}$ (respectively, $\mathbf{y}$). This regularity feature is likely to be found in most conceivable Thurstonian-type models constructed to fit empirical data.

1. Introduction

When two stimuli are presented to an observer for a comparison judgment, one of the most common ways of conceptualizing this situation consists in positing that these stimuli are mapped into two random variables ("images") taking on their values in some hypothetical internal ("perceptual") space, and that the judgment produced by the observer in a given trial is determined (uniquely or probabilistically) by the realizations of these two random images in this trial. The classical and best-known example of this approach is Thurstone's (1927a, b) seminal theory of preference judgments. This theory deals with the situation where stimuli, generally varying along many physical dimensions, are presented in pairs, $(\mathbf{x},\mathbf{y})$, and the observer is asked to compare them in terms of "greater-less" with respect to some semantically unidimensional subjective attribute $\mathscr{P}$ (say, "loudness", or "brightness"). Thurstone assumes that stimuli $\mathbf{x}$ and $\mathbf{y}$ are mapped into two random variables, $P(\mathbf{x})$ and $Q(\mathbf{y})$, with their values belonging to the set of reals (the unidimensional perceptual space representing the attribute $\mathscr{P}$); and $\mathbf{x}$ is judged to be greater than $\mathbf{y}$ in attribute $\mathscr{P}$ if and only if the realization p of $P(\mathbf{x})$ exceeds the realization q of $Q(\mathbf{y})$. Thurstone assumes that $P(\mathbf{x})$ and $Q(\mathbf{y})$ are normally distributed, but this constraint is inessential, alternative distributions having been considered by many, in a variety of contexts (see, e.g., Luce & Suppes, 1965; Robertson & Strauss, 1981; Yellott, 1997). Depending on the version ("case") of the theory, $P(\mathbf{x})$ and $Q(\mathbf{y})$ in it can be stochastically independent or stochastically interdependent (the latter possibility giving rise to an important conceptual issue, mentioned later). The observables that

E-mail address: ehtibar@purdue.edu.

0022-2496/03/$ - see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S0022-2496(02)00027-5

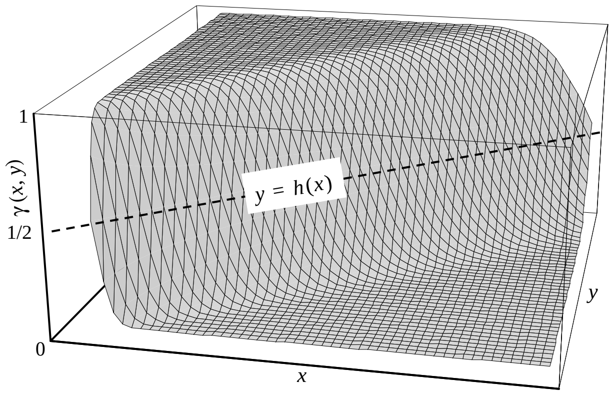

Fig. 1. Possible appearance of a "greater-less" discrimination probability function $\gamma(x,y)$ for unidimensional continuous stimuli. As discussed in Section 3, the intersection of the surface with the plane $\gamma = \frac{1}{2}$ forms the PSE (point of subjective equality) line $y = h(x)$ (in this case, straight).

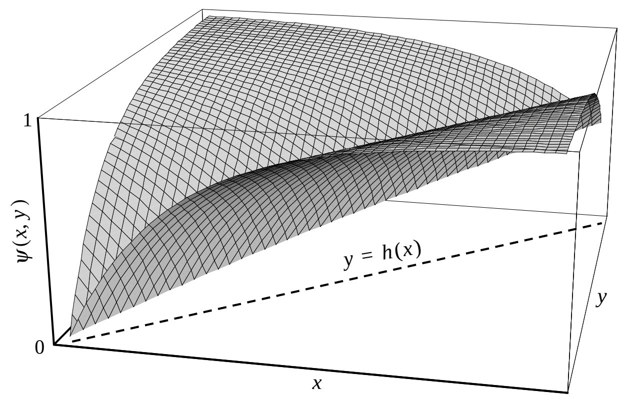

Fig. 2. Possible appearance of a "same-different" discrimination probability function $\psi(x,y)$ for unidimensional continuous stimuli. As discussed in Section 3, the PSE (point of subjective equality) line $y = h(x)$ is formed by the points (x,y) at which the functions $y \to \psi(x,y)$ and $x \to \psi(x,y)$ reach their minima. Note that the level of the minima is different at different points taken on $y = h(x)$. (The function $\psi(x,y)$ shown represents a special case of an alternative to Thurstonian-type models described in the companion paper, Dzhafarov, 2003.)

Thurstone's theory is aimed at predicting are the preference probabilities

$$\gamma(\mathbf{x},\mathbf{y}) = \Pr[\mathbf{y} \text{ is judged to be greater than } \mathbf{x} \text{ in attribute } \mathscr{P}].$$

Fig. 1 illustrates a possible appearance of such a function when the stimuli are unidimensional (and denoted by x, y, in accordance with the notation conventions adopted in this paper).

Historically, the paradigm of preference judgments (with respect to a designated unidimensional attribute) has been so dominant in studying perceptual discriminations that the two notions are often treated as being interchangeable. The present paper, however, focuses on another type of perceptual discriminations, the "*same-different*" comparison paradigm. Stimuli in this paradigm are also presented pairwise, but observers are asked to judge whether $\mathbf{x}$ and $\mathbf{y}$ are the same or different from each other, with no explicitly designated attributes along which the comparisons should be made. The discrimination probability function to be accounted for in this paradigm is

$$\psi(\mathbf{x},\mathbf{y}) = \Pr[\mathbf{y} \text{ is judged to be different from } \mathbf{x}].$$

This function, whose possible appearance is shown in Fig. 2 (for unidimensional stimuli), is very different from the "greater-less" probability function $\gamma(\mathbf{x},\mathbf{y})$, both in its mathematical properties (which should be obvious from comparing Fig. 2 with Fig. 1) and in the theoretical analysis it affords (which is shown in Section 3).

Luce and Galanter (1963), who call the "same-different" paradigm "unordered discriminations", show how it can be handled by a very minor modification of the original Thurstone's theory. They adopt Thurstone's stimulus-to-image mapping scheme: $\mathbf{x}$ and $\mathbf{y}$ are mapped into random variables $P(\mathbf{x})$ and $Q(\mathbf{y})$, taking on their values in the set of reals. But the decision rule now is posited to be different: $\mathbf{y}$ is judged to be different from $\mathbf{x}$ if and only if $P(\mathbf{x}) - Q(\mathbf{y}) > \varepsilon_1$ or $Q(\mathbf{y}) - P(\mathbf{x}) > \varepsilon_2$" (where $\varepsilon_1, \varepsilon_2$ are some positive constants). When complemented by specific assumptions regarding the distributions of $P(\mathbf{x})$ and $Q(\mathbf{y})$ (Luce & Galanter follow Thurstone in assuming their normality), the stochastic relationship between them (Luce & Galanter assume independence), and the dependence of their parameters (in this case, means and variances) on, respectively, $\mathbf{x}$ and $\mathbf{y}$, this model generates a specific function $\psi(\mathbf{x},\mathbf{y})$, testable vis-a-vis empirical observations.

Aside from simplicity considerations, the assumption that the hypothetical images $P(\mathbf{x})$ and $Q(\mathbf{y})$ are unidimensional (real-valued) is significantly less compelling in the context of "same-different" discriminations than it is in the context of "greater-less" ones. In the latter case the unidimensionality of the perceptual space reflects the sematic unidimensionality of the attribute $\mathscr{P}$ along which the comparisons are being made. In the case of "same-different" comparisons, even if stimuli are physically unidimensional, this justification is absent, and it seems more plausible to think of the perceptual representations as being multiattribute, if analyzable into attributes at all. In many models for "same-different" comparisons, therefore, $P(\mathbf{x})$ and $Q(\mathbf{y})$ are assumed to be distributed (usually, multivariate-normally) in Re^m, the m-dimensional space of real-component vectors (Ennis, 1992; Ennis, Palen, & Mullen, 1988; Suppes & Zinnes, 1963; Zinnes & MacKey, 1983; Thomas, 1996, 1999). This generalization immediately broadens the class of possible decision rules: one can posit, for example, that $\mathbf{y}$ is judged to be different from $\mathbf{x}$ if and only if an appropriately defined distance between $P(\mathbf{x})$ and $Q(\mathbf{y})$ exceeds a critical value; or one can assume that the space is partitioned into many subregions ("categories"), and $\mathbf{y}$ is judged to be different from $\mathbf{x}$ if and only if $P(\mathbf{x})$ and $Q(\mathbf{y})$ fall into two different subregions.

Once on this road, however, it seems natural to entertain the possibility that the random images $P(\mathbf{x})$ and $Q(\mathbf{y})$ may be more complex than being representable by points in Re^m. Thus, if one thinks of perceptual images as, say, pictorial templates or processes developing in time, they should be represented by functions or relations rather than real-component vectors. These and similar possibilities lead one to a sweeping generalizations of Luce and Galanter's (1963) modification of Thurstone's modeling scheme, the one in which $\mathbf{x}$ and $\mathbf{y}$ are mapped into random images $P(\mathbf{x})$ and $Q(\mathbf{y})$ whose possible values belong to a space $\mathfrak{R}$ *of arbitrary nature*, constrained only by the requirement that it support the probability measures associated with $P(\mathbf{x})$ and $Q(\mathbf{y})$. The choice of the response in a given trial, "same" or "different", depends on the realizations p of $P(\mathbf{x})$ and q of $Q(\mathbf{y})$.

If a discrimination probability function $\psi(\mathbf{x},\mathbf{y})$ is generated by such a model, with some choice of the space $\mathfrak{R}$, random images $P(\mathbf{x}), Q(\mathbf{y})$, and a decision rule mapping their realizations (p,q) into responses, then this function $\psi(\mathbf{x},\mathbf{y})$ is said to have a *Thurstonian-type representation* (defined in detail in Section 4). It should be clear from this description that the notion of a Thurstonian-type representation generally does not imply any restrictions on possible distributions of $P(\mathbf{x})$ and $Q(\mathbf{y})$ or on possible decision rules.

Remark 1.1. Probabilities and their modeling in this paper are always treated on a population rather than sample level. Thus, a discrimination probability function $\psi(\mathbf{x},\mathbf{y})$ is considered having (or not having) a Thurstonian-type representation if a Thurstonian-type model exists (respectively, does not exist) that generates this $\psi(\mathbf{x},\mathbf{y})$ precisely.

Using traditional terminology, a Thurstonian-type representation for $\psi(\mathbf{x},\mathbf{y})$ can be called a *random utility model*, provided the term is taken in its broadest meaning. In random utility models the internal space $\mathfrak{R}$ is usually assumed to be an interval of reals (see, e.g., Luce & Suppes, 1965; Regenwetter & Marley, 2001), while in the present paper the notion of a random image includes as special cases such entities as the "random relations" and "random functions" analyzed in Regenwetter and Marley (2001) as conceptual alternatives to real-valued "random utilities". Niederée and Heyer (1997) use the term "*generalized* random utility models" to incorporate such constructs. (The legitimacy of the notion of a random variable with values in an arbitrary measure space is discussed in Section 4.)

Overall, the terminology used in the literature in relation to random utilities and modifications of Thurstone's theory is diverse if not confusing. Thus, the term "Thurstone model" is used by Strauss (1979) to designate the classical Thurstone's model for preference judgments under the constraint that $P(\mathbf{x}), Q(\mathbf{y})$ are independent and their distributions (not necessarily normal) differ in the shift parameter only; if $P(\mathbf{x}), Q(\mathbf{y})$ are distributed differently (in the set of reals), the model is called "a generalized Thurstone model". To prevent confusions, the term "Thurstonian-type" as used in this paper should be taken strictly as defined (here and, more rigorously, in Section 4): *any model* (*for preference or "same-different" judgments*) *in which* $\mathbf{x},\mathbf{y}$ *are mapped into random images* $P(\mathbf{x}), Q(\mathbf{y})$ *in an arbitrary space* $\mathfrak{R}$, *with the subsequent mapping* (*deterministic or probabilistic*) *of their realizations into one of the two responses*. (Certain caveats, however, apply to the joint distribution of $P(\mathbf{x}), Q(\mathbf{y})$, as mentioned below.)

Remark 1.2. A.A.J. Marley (pers. comm., July 13, 2002) suggested that the term "*random-image representation*" might be preferable to the term "Thurstonian-type representation". While he is not entirely happy with either terminology, he prefers using a term that does not refer to Thurstone's original model, as he thinks the latter is too narrow compared to the constructs considered in this paper.

The present paper only deals with Thurstonian-type representations in which the random images $P(\mathbf{x}), Q(\mathbf{y})$ are *stochastically independent*, while the decision rule is *deterministic*: a given pair of realizations (p,q) of the random images $P(\mathbf{x}), Q(\mathbf{y})$ leads to one and only one of the two responses, "same" or "different". All models for "same-different" discriminations referenced above are Thurstonian-type models with stochastically independent random images and deterministic decision rules. In a companion paper (Dzhafarov, 2003) the analysis and its conclusions are extended to Thurstonian-type models in which the decision rules may be *probabilistic* (i.e., every realization of the two random images may lead to each of the two responses, with some probabilities), while the images $P(\mathbf{x})$ and $Q(\mathbf{y})$ may be *stochastically interdependent* (but *selectively attributed* to, respectively $\mathbf{x}$ and $\mathbf{y}$, in a well-defined sense).

Remark 1.3. The selective attribution of images to stimuli is viewed as an inherent feature of Thurstonian-type representations: $\mathbf{x}$ is mapped into *its* image $P(\mathbf{x})$, while $\mathbf{y}$ is mapped into *its* image $Q(\mathbf{y})$. The models in which $(\mathbf{x},\mathbf{y})$ as a pair is mapped into a single image $R(\mathbf{x},\mathbf{y})$ are not included in the class of Thurstonian-type models. In particular, I do not include in this class the models in which every pair of stimuli being presented evokes a single-random variable taking on its values on an axis of "subjective pairwise differences" (as, e.g., in the model by Takane & Sergent, 1983). When $P(\mathbf{x})$ and $Q(\mathbf{y})$ are stochastically independent, the meaning of their selective attribution to, respectively, $\mathbf{x}$ and $\mathbf{y}$ is plain, but in the case of stochastically interdependent images one faces a formidable conceptual problem (see Dzhafarov, 1999, 2001a, in press). Even if the perceptual image caused by $(\mathbf{x}, \mathbf{y})$ can be decomposed into P and Q such that the *marginal distribution* of P depends on $\mathbf{x}$ only while the *marginal distribution* of Q depends on $\mathbf{y}$ only, what constraints should be imposed on the dependence of the *joint* distribution of (P, Q) upon $(\mathbf{x}, \mathbf{y})$ to enable one to speak of P being the image of $\mathbf{x}$ alone and of Q as the image of $\mathbf{y}$ alone? A satisfactory answer to this question is less obvious than it might seem. In relation to Thurstonian-type models this problem is dealt with in Dzhafarov (2003), based on the general theory of selective influence developed in Dzhafarov (in press).

The focus of this paper is on the applicability of Thurstonian-type modeling (with independent images and deterministic decisions) to the "same-different" comparison paradigm with stimuli belonging to a *continuous stimulus space*. In other words, stimuli $\mathbf{x}, \mathbf{y}$ vary along one or several continuous physical dimensions, such as length, intensity, or color coordinates. Discrimination probability functions $\psi(\mathbf{x}, \mathbf{y})$ with continuously varying stimuli $(\mathbf{x}, \mathbf{y})$ are analyzed in Dzhafarov (in press c), where it is argued that these functions possess two basic properties, called *regular minimality* and *nonconstant self-similarity*. The main thesis of the present paper is that *although one can find a Thurstonian-type representation for any discrimination probability function* $\psi(\mathbf{x}, \mathbf{y})$, *a function* $\psi(\mathbf{x}, \mathbf{y})$ *that possesses the two basic properties just mentioned cannot have a Thurstonian-type representation that is "well-behaved"*. The well-behavedness of a Thurstonian-type representation is a constraint imposed on the dependence of the random images $P(\mathbf{x}), Q(\mathbf{y})$ upon, respectively, $\mathbf{x}$ and $\mathbf{y}$. The well-behavedness means that the probabilities with which either of $P(\mathbf{x}), Q(\mathbf{y})$ falls within an area of the perceptual space $\mathfrak{R}$ have appropriately defined and bounded directional derivatives with respect to, respectively, $\mathbf{x}$ or $\mathbf{y}$. The class of well-behaved Thurstonian-type models is likely to include most realistically conceivable Thurstonian-type models constructed to fit empirical data.

2. Notation conventions and plan of the paper

The reader may find it helpful to occasionally consult this section to keep track of the presentation logic and notational issues.

Boldface lowercase letters $(\mathbf{x}, \mathbf{y}, \mathbf{u}, \ldots)$ always denote real-valued vectors (usually points and directions in a stimulus space); their components, if shown, are superscripted, e.g., $\mathbf{x} = (x^1, \ldots, x^n)$, $\mathbf{u} = (u^1, \ldots, u^n)$.

Uppercase Gothic letters $(\mathfrak{R}, \mathfrak{U}, \mathfrak{B}$, etc.) denote sets, lowercase Gothic letters $(\mathfrak{p}, \mathfrak{q}, \mathfrak{a}, \ldots)$ denote subsets of the "perceptual space" $\mathfrak{R}$.

Uppercase Greek letters Σ and Ω denote sets of subsets of $\mathfrak{R}$.

Lowercase and uppercase italics designate real-valued quantities, except for letters p, q (and P, Q) that are reserved to denote elements of $\mathfrak{R}$ (respectively, random variables with values in $\mathfrak{R}$). In one place, however (in the proof of Theorem 5.1), where $\mathfrak{R} \subseteq \mathrm{Re}^m$, its elements are denoted by $\mathbf{p} = (p^1, \ldots, p^m)$ and $\mathbf{q} = (q^1, \ldots, q^m)$.

The development presented in this paper can be summarized as follows.

1. Section 3 compares the "greater-less" and "same-different" discrimination probability functions $\gamma(x, y)$ and $\psi(x, y)$, for continuous unidimensional stimuli. It provides a simple demonstration for the failure of well-behaved Thurstonian-type models to account for the two basic properties of $\psi(x, y)$, regular minimality and nonconstant self-similarity. This failure is contrasted with the fact that well-behaved Thurstonian-type models easily account for the analogous basic property of $\gamma(x, y)$ (called "regular mediality").
2. The class of Thurstonian-type models with stochastically independent images and deterministic decision rules is formally defined in Section 4, and it is proved subsequently (Theorem 5.1) that any discrimination probability function $\psi(\mathbf{x}, \mathbf{y})$ has a Thurstonian-type representation of this kind, provided no a priori restrictions are imposed on the possible distributions of random images or on the decision rule.
3. Following some technical preparation in Section 6, the intuition underlying the general notion of a well-behaved Thurstonian-type model is outlined in Section 7, followed by a rigorous definition.
4. It is proved then (Theorem 8.1) that the discrimination probability functions $\psi(\mathbf{x}, \mathbf{y})$ predicted by well-behaved Thurstonian-type models possess a certain smoothness property, called "near-smoothness" (essentially, a weak form of continuous differentiability).
5. A general definition of the regular minimality and nonconstant self-similarity properties of $\psi(\mathbf{x}, \mathbf{y})$ is given in Section 9. It is proved then (Theorem 10.1) that together these two properties exclude the possibility that $\psi(\mathbf{x}, \mathbf{y})$ possesses the near-smoothness property.
6. The obvious conclusion follows: $\psi(\mathbf{x}, \mathbf{y})$ subject to regular minimality and nonconstant self-similarity cannot have a well-behaved Thurstonian-type representation.
7. This conclusion is then shown to extend to a broader class of well-behaved Thurstonian models, obtained

by relaxing some of the constraints imposed in Section 7.

8. The development is aided by an appendix containing lemmas labeled A.1, A.2, etc.

To follow the mathematical derivations the reader should be familiar with basic concepts of abstract measure theory. With the help provided in the paper, however, knowledge of standard multivariate calculus should be sufficient to understand the general logic and the results.

3. Comparing two discrimination paradigms

To simplify the discussion, let the stimuli presented for a comparison judgment ("greater-less", with respect to some attribute $\mathscr{P}$, or "same-different") be physically unidimensional. In accordance with the general definition of a continuous stimulus space given in the next section, the stimulus values x, y in this case belong to an *open interval of reals* $\mathfrak{M}$ (finite or infinite).

As emphasized in Dzhafarov (2002d), the key fact about pairwise presentations (x, y) is that stimuli x and y belong to two distinct *observation areas* (spatial and/or temporal intervals): thus, x may be presented first, followed by y, or x may be presented to the left and y to the right of a fixation point. As a result, (x, y) and (y, x) are distinct pairs, and (x, x) is a *pair* of stimuli rather than a single stimulus. In the following the observation area to which a stimulus belongs is encoded by its ordinal position within a pair (x, y), and the respective observation areas are referred to as the "first" and the "second" areas. One consequence of treating (x, y) as an ordered pair is that $\psi(x, y)$ and $\psi(y, x)$ in the "same-different" paradigm are generally different, and so are $\gamma(x, y)$ and $1-\gamma(y, x)$ in the "greater-less" paradigm. (For a detailed discussion of distinct observation areas see Dzhafarov, 2002d)

A central notion in the analysis of perceptual discriminations is that of a *point of subjective equality* (PSE). The intuitive meaning of this notion is as follows. For a given stimulus x (belonging to the first observations area), a stimulus y (in the second observation area) is the PSE for x if the "subjective dissimilarity" between this y and x is smaller than the dissimilarity between x and any other stimulus belonging to the second observation area (generally, this point y need not be equal to x). Analogously, the PSE for y (belonging to the second observation area) is the stimulus x (in the first observation area) whose dissimilarity to y is smaller than that of any other stimulus taken in the first observation area. It seems natural to expect that a reasonable operationalization of this notion should ensure that a PSE for a given stimulus (in either observation area) is uniquely defined, and that the relation of "being a PSE of" is symmetric:

$$y \text{ is the PSE for } x \Leftrightarrow x \text{ is the PSE for } y,$$

with x and y in both cases belonging to the first and second observation areas, respectively. Dzhafarov (2002d) proposes to treat this regularity property (the existence, uniqueness, and symmetry of PSE) as a fundamental law for perceptual discriminations.

To discuss first the "greater-less" comparison paradigm, refer to Figs. 1 and 3. In accordance with the traditional psychophysical focus on "correlated" physical and subjective continua, let the attribute $\mathscr{P}$ be chosen so that $\gamma(x, y)$ is strictly increasing in y and strictly decreasing in x (recall that γ is the probability that y, in the second observation area, is judged to be greater than x, in the first observation area). As an example, x, y may represent physical intensity of successively presented tones, and $\mathscr{P}$ their loudness. Let $\gamma(x, y)$ be continuous in (x, y). Clearly, the PSE for x with respect to $\mathscr{P}$ should be defined as the value of y for which $\gamma(x, y)=\frac{1}{2}$, and the same equality determines the PSE (with respect to $\mathscr{P}$) for y: the value of x for which $\gamma(x, y)=\frac{1}{2}$. Denoting the PSE

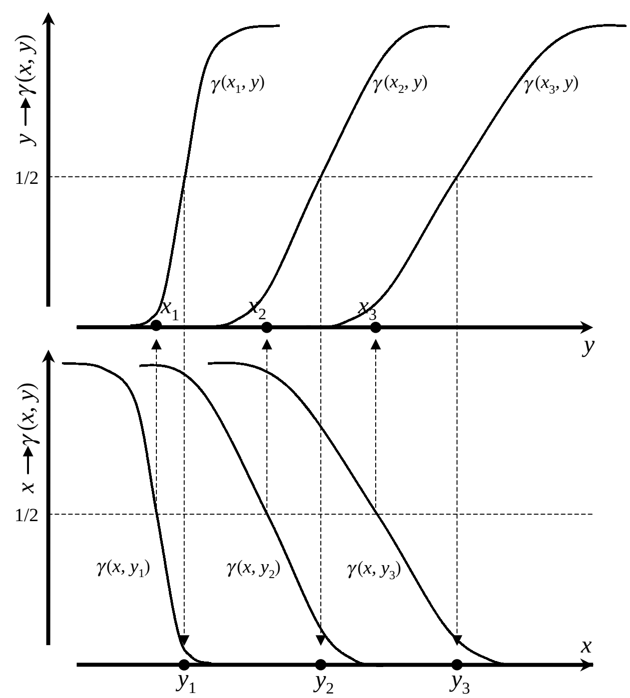

Fig. 3. Illustration of the regular mediality property. Shown are cross-sections $y \rightarrow \gamma(x, y)$ and $x \rightarrow \gamma(x, y)$ of a "greater-less" discrimination probability function $\gamma(x, y)$ for unidimensional continuous stimuli (see Fig. 1). The cross-sections $y \rightarrow \gamma(x, y)$ shown in the upper panel are made at arbitrarily chosen $x=x_1, x_2, x_3$. The cross-sections $x \rightarrow \gamma(x, y)$ shown in the lower panel are made at $y=y_1, y_2, y_3$ which are the medians of the three curves in the upper panel. The regular mediality property is illustrated by the fact that then x_1, x_2, x_3 are the medians of the three curves in the lower panel. In other words, y_i is the median of $\gamma(x_i, y)$ if and only if x_i is the median of $\gamma(x, y_i)$, $i=1,2,3$.

for x by $h(x)$, the PSE for y by $g(y)$, and assuming that both these functions are defined on the entire interval $\mathfrak{M}$ (i.e., a PSE exists for every stimulus, in either observation area), a moment's reflection reveals that both $h(x)$ and $g(y)$ are one-to-one, onto, and continuous transformations $\mathfrak{M} \to \mathfrak{M}$, and that $g \equiv h^{-1}$. By analogy with the regular minimality property introduced below, this fundamental regularity condition can be termed *regular mediality*.

Since regular mediality holds for any continuous $\gamma(x, y)$ such that the equation $\gamma(x, y) = \frac{1}{2}$ can always be solved for both x and y, this property is predicted by any reasonable Thurstonian-type model, including Thurstone's original theory. Let the perceptual space $\mathfrak{R}$ be the set of reals, and let the random images $P(x), Q(y)$ be independent and normally distributed. Let their respective means $\mu_P(x), \mu_Q(y)$ be two arbitrarily smooth (e.g, infinitely differentiable) increasing functions $\mathfrak{M} \xrightarrow{\text{onto}} \mathrm{Re}^+$. Then one can always choose standard deviation functions $\sigma_P(x), \sigma_Q(y)$, also arbitrarily smooth, so that the discrimination probability function

$$\gamma(x, y) = \Phi\left(\frac{\mu_Q(y) - \mu_P(x)}{\sqrt{\sigma_P^2(x) + \sigma_Q^2(y)}}\right),$$

where Φ is the standard normal integral, is increasing in y, decreasing in x, and continuous (in fact, arbitrarily smooth) in (x, y). This is achieved, for example, if one chooses $\sigma_P(x) \equiv \sigma_P$, $\sigma_Q(y) \equiv \sigma_Q$ (constants), or $\sigma_P(x) = k\sqrt{\mu_P(x)}$, $\sigma_Q(y) = k\sqrt{\mu_Q(y)}$, $k > 0$. Since $\mu_P(x)$ and $\mu_Q(y)$ have identical ranges (Re^+), the equation $\gamma(x, y) = \frac{1}{2}$ can be solved for y and x at, respectively, any x and y, resulting in

$$h(x) = \mu_Q^{-1}[\mu_P(x)], \quad g(y) = \mu_P^{-1}[\mu_Q(y)].$$

These functions satisfy the regular mediality condition because $g \equiv h^{-1}$. Anticipating the subsequent development, the point to be emphasized here is that the regular mediality property for $\gamma(x, y)$ is predicted by Thurstonian-type models that are arbitrarily "well-behaved" (intuitively, $P(x), Q(y)$ have "nice" distributions whose parameters smoothly depend on stimuli).

The situation changes dramatically as we turn to "same-different" comparisons. Refer to Figs. 2 and 4. In accordance with the so-called "First Assumption of multidimensional Fechnerian scaling" (discussed in greater generality in Section 9), $\psi(x, y)$ is continuous in (x, y), $y \to \psi(x, y)$ reaches a global minimum at some point $h(x)$, continuously changing with x, and $x \to \psi(x, y)$ reaches a global minimum at some point $g(y)$, continuously changing with y. Clearly, $h(x)$ should be taken as the PSE for x, and $g(y)$ as the PSE for y. The symmetry of the PSE relation in this case, $g \equiv h^{-1}$, is called the property of *regular minimality*. Unlike the regular mediality above, this property is not a mathematical necessity, but it is both intuitively plausible and corroborated by available empirical evidence (see Dzhafarov, 2002d). It is also known (Dzhafarov, 2002d) that the minimum level $\psi(x, h(x))$ of the function $y \to \psi(x, y)$ is generally different for different values of x, and, analogously, the minimum level $\psi(g(y), y)$ of the function $x \to \psi(x, y)$ is generally different for different values of y. This well-documented property is called *nonconstant self-similarity*. It has no analogue in the "greater-less" discrimination probability functions, where the level of $\gamma(x, y)$ at which the PSE are taken is $\frac{1}{2}$ by definition.

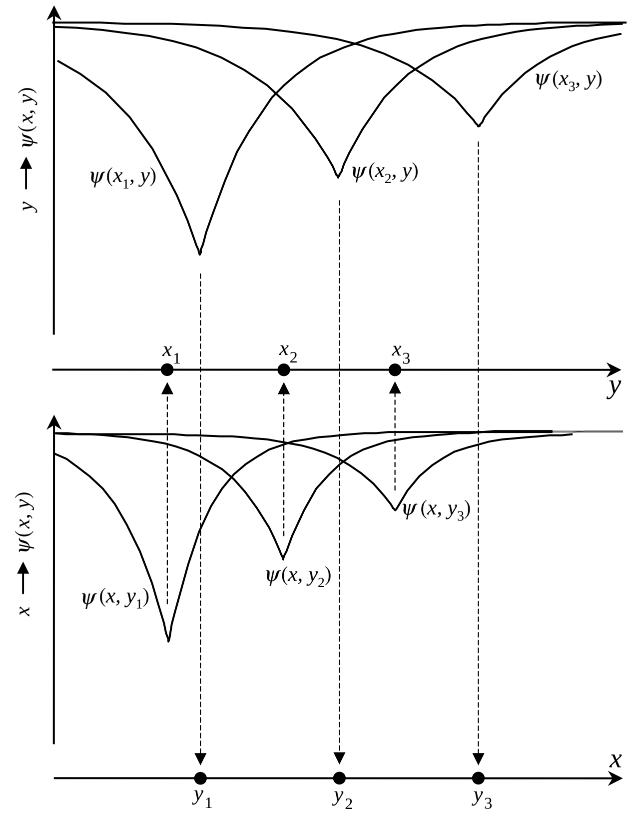

Fig. 4. Illustration of the regular minimality and nonconstant self-similarity properties. Shown are cross-sections $y \to \psi(x, y)$ and $x \to \psi(x, y)$ of a "same-different" discrimination probability function $\psi(x, y)$ for unidimensional continuous stimuli (see Fig. 2). The cross-sections $y \to \psi(x, y)$ shown in the upper panel are made at arbitrarily chosen $x = x_1, x_2, x_3$. The cross-sections $x \to \psi(x, y)$ shown in the lower panel are made at $y = y_1, y_2, y_3$ which are the points at which the three curves in the upper panel reach their minima. The regular minimality property is illustrated by the fact that then x_1, x_2, x_3 are the points at which the three curves in the lower panel reach their minima. In other words, $\arg\min_y \psi(x_i, y) = y_i$ if and only if $\arg\min_x \psi(x, y_i) = x_i$, $i = 1, 2, 3$. The nonconstant self-similarity property is illustrated by the fact that the minimum level $\psi(x_i, y_i)$ is different for different PSE pairs (x_i, y_i), $i = 1, 2, 3$.

Remark 3.1. The term "self-similarity" (or "self-dissimilarity") is due to the fact that "ideally" the minima of the functions $y \to \psi(x, y)$ and $x \to \psi(x, y)$ are reached at

$y = x$, in which case it is said that there is no *constant error*.

Can these two properties, regular minimality and nonconstant self-similarity, be predicted by a Thurstonian-type model with "nicely" distributed and "nicely" dependent on stimuli images, like the ones used above to model $\gamma(x, y)$? It turns out that any attempt to come up with such a model will fail. Take, for example, the same $P(x), Q(y)$ as used above, independent and normally distributed on the set of reals, with their means $\mu_P(x), \mu_Q(y)$ being smooth increasing functions $\mathfrak{M} \xrightarrow{\text{onto}} \mathrm{Re}^+$ and their standard deviations being chosen as either $\sigma_P(x) \equiv \sigma_P$, $\sigma_Q(y) \equiv \sigma_Q$ (constants) or $\sigma_P(x) = k\sqrt{\mu_P(x)}$, $\sigma_Q(x) = k\sqrt{\mu_Q(x)}$. Let the decision rule be the symmetric version of the one proposed by Luce and Galanter (1963): "say that y is different from x if and only if $|P(x) - Q(y)| > \varepsilon$". Then

$$\psi(x, y) = 1 - \Phi\left(\frac{\varepsilon - [\mu_Q(y) - \mu_P(x)]}{\sqrt{\sigma_P^2(x) + \sigma_Q^2(y)}}\right) + \Phi\left(\frac{-\varepsilon - [\mu_Q(y) - \mu_P(x)]}{\sqrt{\sigma_P^2(x) + \sigma_Q^2(y)}}\right).$$

Choosing first $\sigma_P(x) \equiv \sigma_P$, $\sigma_Q(y) \equiv \sigma_Q$, one can easily show that $y \to \psi(x, y)$ and $x \to \psi(x, y)$ reach their minima at, respectively,

$$y = \mu_Q^{-1}[\mu_P(x)] = h(x), \quad x = \mu_P^{-1}[\mu_Q(y)] = g(y),$$

which are the same functions as obtained above for $\gamma(x, y)$, satisfying thereby the requirement $g \equiv h^{-1}$. With constant standard deviations, however, the minimum level of $y \to \psi(x, y)$ and $x \to \psi(x, y)$ is constant,

$$\psi(x, h(x)) = \psi(g(y), y) = 2\Phi\left(\frac{-\varepsilon}{\sqrt{\sigma_P^2 + \sigma_Q^2}}\right).$$

The $\psi(x, y)$ generated by this model, therefore, while satisfying the regular minimality condition, also has the property of *constant* self-similarity.

Choosing $\sigma_P(x) = k\sqrt{\mu_P(x)}$, $\sigma_Q(x) = k\sqrt{\mu_Q(x)}$ does make the minimum levels of $y \to \psi(x, y)$ and $x \to \psi(x, y)$ vary with, respectively, x and y, satisfying thereby the nonconstant self-similarity requirement. By a simple though cumbersome derivation, however, one can show that in this case one loses the regular minimality property: in fact, one cannot find a single pair (x, y) at which the functions $y \to \psi(x, y)$ and $x \to \psi(x, y)$ reach their minima simultaneously. Indeed, denoting

$$A(x, y, \varepsilon) = \exp\left[-\frac{(\varepsilon + \mu_Q(y) - \mu_P(x))^2}{2k^2(\mu_Q(y) + \mu_P(x))}\right],$$

$$B(x, y, \varepsilon) = \exp\left[-\frac{(\varepsilon - \mu_Q(y) + \mu_P(x))^2}{2k^2(\mu_Q(y) + \mu_P(x))}\right],$$

the conditions $\frac{\partial\psi(x,y)}{\partial y} = 0$ and $\frac{\partial\psi(x,y)}{\partial x} = 0$ can be shown to be equivalent to, respectively,

$$[\mu_Q(y) + 3\mu_P(x) - \varepsilon]A(x, y, \varepsilon) - [\mu_Q(y) + 3\mu_P(x) + \varepsilon]B(x, y, \varepsilon) = 0$$

and

$$[3\mu_Q(y) + \mu_P(x) + \varepsilon]A(x, y, \varepsilon) - [3\mu_Q(y) + \mu_P(x) - \varepsilon]B(x, y, \varepsilon) = 0.$$

Their addition yields

$$[\mu_Q(y) + \mu_P(x)][A(x, y, \varepsilon) - B(x, y, \varepsilon)] = 0,$$

which can easily be proved impossible for positive $\mu_P(x)$, $\mu_Q(y)$, and ε.

The subsequent development shows that these Thurstonian-type models cannot be repaired by any modifications of the perceptual space $\mathfrak{R}$, decision rules, or distributions of $P(x)$ and $Q(y)$, insofar as the dependence of these distributions on stimuli x, y remains sufficiently "well-behaved". For the "same-different" discrimination probabilities $\psi(x, y)$ the conjunction of the properties of regular minimality and nonconstant self-similarity is incompatible with well-behaved Thurstonian-type representations. In a sense, the only reason such representations are possible for the "greater-less" discrimination probabilities $\gamma(x, y)$ is that there is no analog of nonconstant self-similarity for these functions: by definition, the regular mediality property is associated with the fixed probability level $\frac{1}{2}$.

A rigorous and general definition for the notion of "well-behavedness" is given in Section 7 (further generalized in Section 11). Put semi-formally, in the case of unidimensional continuous stimuli the "well-behavedness" of, say, the distribution of $P(x)$ at a point $x = x_0$ means that in a sufficiently small neighborhood $[x_0 - a, x_0 + a]$ the probabilities $\Pr[P(x) \in \mathfrak{p}]$ for all possible events $\mathfrak{p}$ have bounded right- and left-hand derivatives with respect to x.

4. Thurstonian-type representations

A Thurstonian-type representation (with independent random images and a deterministic decision rule) for a discrimination probability function $\psi(\mathbf{x}, \mathbf{y})$ is defined by the construct

$$\{\mathfrak{R}, A_{\mathbf{x}}, B_{\mathbf{y}}, \mathfrak{S}\}, \tag{1}$$

with the following meaning of the terms.

(i) $\mathbf{x}$ and $\mathbf{y}$ are stimuli whose values belong to an n-dimensional *continuous stimulus space* $\mathfrak{M} \subseteq \mathrm{Re}^n$ ($n \geqslant 1$), an *open connected area* with its dimensions representing physical attributes varying in the experiment. As discussed in the previous section, $\mathbf{x}$ and $\mathbf{y}$ belong to distinct observation areas (spatial and/or temporal intervals), encoded by their positions within the pair $(\mathbf{x}, \mathbf{y})$. The representation $\mathfrak{M}$ for a given set of physical stimuli is not unique: any homeomorphic (one-to-one, onto, continuous together with its inverse) mapping $\mathfrak{M} \to \mathfrak{M}' \subseteq \mathrm{Re}^n$ creates an *equivalent* representation for the stimulus space. This means, in particular, that Re^n in the inclusion $\mathfrak{M} \subseteq \mathrm{Re}^n$ can always be replaced by any subset of Re^n homeomorphically related to Re^n (e.g., an open n-dimensional unit cube).

Remark 4.1. It is allowed, if convenient, to apply different homeomorphic transformations to stimuli in the first and second observation areas, including the possibility that they map onto different sets, $\mathfrak{M}', \mathfrak{M}''$. It may have been better, therefore, to speak of two generally different stimulus spaces for $\mathbf{x}$ and for $\mathbf{y}$ in $(\mathbf{x}, \mathbf{y})$. The notion of a single stimulus space is retained in the present paper for expository simplicity only.

(ii) $\mathfrak{R}$ is a hypothetical *perceptual space*, whose points $p, q, \ldots$ are referred to as (perceptual) *images*. No restrictions are imposed on its possible structure: it can be a finite set of states, an area of Re^m, a space of continuous processes, a space of compact areas belonging to Re^m, etc. The term "space" (rather than "set") is used here due to the structure imposed on $\mathfrak{R}$ by the probability measures considered next.

(iii) $A_{\mathbf{x}}(\mathfrak{p})$ and $B_{\mathbf{y}}(\mathfrak{q})$, $\mathfrak{p}, \mathfrak{q} \subseteq \mathfrak{R}$, are *probability measures* defined on respective sigma-algebras Σ_A, Σ_B, the same for all values of $\mathbf{x}, \mathbf{y}$. (A sigma-algebra is a set of subsets of $\mathfrak{R}$ that includes $\mathfrak{R}$ and is closed under countable applications of standard set-theoretic operations.) $A_{\mathbf{x}}(\mathfrak{p})$ and $B_{\mathbf{y}}(\mathfrak{q})$ are associated with random images $P(\mathbf{x})$ and $Q(\mathbf{y})$ of stimuli $\mathbf{x}$ and $\mathbf{y}$,

$$A_{\mathbf{x}}(\mathfrak{p}) = \Pr[P(\mathbf{x}) \in \mathfrak{p}], \quad \mathfrak{p} \in \Sigma_A,$$

$$B_{\mathbf{y}}(\mathfrak{q}) = \Pr[Q(\mathbf{y}) \in \mathfrak{q}], \quad \mathfrak{q} \in \Sigma_B.$$

The measures $A_{\mathbf{x}}$ and $B_{\mathbf{y}}$, as well as their domains Σ_A and Σ_B, are generally different, to reflect the fact that $\mathbf{x}$ and $\mathbf{y}$ belong to distinct observation areas.

Remark 4.2. In the simplest case, when $\mathfrak{R}$ is a finite set of states $\{1, \ldots, k\}$, the distributions of $P(\mathbf{x})$ and $Q(\mathbf{y})$ are defined by $\alpha(i, \mathbf{x}) = \Pr[P(\mathbf{x}) = i]$ and $\beta(i, \mathbf{y}) = \Pr[Q(\mathbf{y}) = i]$, $i = 1, \ldots, k$. In this case, $A_{\mathbf{x}}(\mathfrak{p}) = \sum_{i \in \mathfrak{p}} \alpha(i, \mathbf{x})$, $B_{\mathbf{y}}(\mathfrak{q}) = \sum_{i \in \mathfrak{q}} \beta(i, \mathbf{y})$, and $\Sigma_A = \Sigma_B$ is the set of all 2^k subsets of $\{1, \ldots, k\}$. The reader who wishes to overlook measure-theoretic technicalities may think of this example throughout the paper, ignoring all references to sigma-algebras and measurability.

(iv) $\mathfrak{S} \subseteq \mathfrak{R} \times \mathfrak{R}$ is the area that determines (and is determined by) a *decision rule*: $\mathbf{x}$ and $\mathbf{y}$ are judged to be different in a given trial if and only if $(p, q) \in \mathfrak{S}$, where p and q are the values of, respectively, $P(\mathbf{x})$ and $Q(\mathbf{y})$ in this trial. For brevity, I refer to $\mathfrak{S}$ as the *decision area*, instead of the more explicit "area corresponding to the decision that $\mathbf{x}$ and $\mathbf{y}$ are different". (Nothing would have changed in the treatment to follow if we considered instead the complementary area, mapped into the response "same".) The measures $A_{\mathbf{x}}$ and $B_{\mathbf{y}}$ induce on $\mathfrak{R} \times \mathfrak{R}$ the probabilistic product measure $AB_{\mathbf{xy}} = A_{\mathbf{x}} \times B_{\mathbf{y}}$, defined on the sigma-algebra Σ_{AB} generated by $\Sigma_A \times \Sigma_B$,

$$AB_{\mathbf{xy}}(\mathfrak{s}) = \Pr[(P(\mathbf{x}), Q(\mathbf{y})) \in \mathfrak{s}], \quad \mathfrak{s} \in \Sigma_{AB}.$$

The decision area $\mathfrak{S}$ is assumed to be AB-measurable (i.e., $\in \Sigma_{AB}$).

Remark 4.3. One can say "AB-measurable" rather than "$AB_{\mathbf{xy}}$-measurable" because the sigma-algebra Σ_{AB} is the same for all $(\mathbf{x}, \mathbf{y})$. Analogously, the terms "A -measurable" and "B-measurable" are used below instead of "$A_{\mathbf{x}}$-measurable" and "$B_{\mathbf{y}}$-measurable".

Remark 4.4. For $\mathfrak{R} = \{1, \ldots, k\}$ introduced in Remark 4.2, Σ_{AB} is the set of all 2^{k^2} subsets of the set $\{1, \ldots, k\} \times \{1, \ldots, k\}$, and $AB_{\mathbf{xy}}(\mathfrak{s}) = \sum_{(i,j) \in \mathfrak{s}} \alpha(i, \mathbf{x}) \beta(j, \mathbf{y})$.

(v) The general relation between the unobservables $\{\mathfrak{R}, A_{\mathbf{x}}, B_{\mathbf{y}}, \mathfrak{S}\}$ of the model and the observable $\psi(\mathbf{x}, \mathbf{y})$ is given by

$$\psi(\mathbf{x}, \mathbf{y}) = \Pr[(P(\mathbf{x}), Q(\mathbf{y})) \in \mathfrak{S}] = AB_{\mathbf{xy}}(\mathfrak{S}).$$

We have (see, e.g., Hewitt & Stromberg, 1965, p. 384)

$$AB_{\mathbf{xy}}(\mathfrak{S}) = \int_{q \in \mathfrak{R}} A_{\mathbf{x}}[\mathfrak{a}(q)]\, dB_{\mathbf{y}}(q) = \int_{p \in \mathfrak{R}} B_{\mathbf{y}}[\mathfrak{b}(p)]\, dA_{\mathbf{x}}(p), \tag{2}$$

where

$$\mathfrak{a}(q) = \{p \in \Re: (p,q) \in \mathfrak{S}\},$$
$$\mathfrak{b}(p) = \{q \in \Re: (p,q) \in \mathfrak{S}\} \quad (3)$$

are, respectively, q- and p-sections of the area $\mathfrak{S}$. We know (e.g., Hewitt & Stromberg, 1965, p. 380) that all q-sections of $\mathfrak{S}$ are A-measurable (i.e., $\in \Sigma_A$) and all p-sections of $\mathfrak{S}$ are B-measurable ($\in \Sigma_B$). We also know (Hewitt & Stromberg, 1965, p. 381) that the functions $q \to A_{\mathbf{x}}[\mathfrak{a}(q)]$ and $p \to B_{\mathbf{y}}[\mathfrak{b}(p)]$ are, respectively, B-measurable and A-measurable. (The B-measurability of $q \to A_{\mathbf{x}}[\mathfrak{a}(q)]$ means that the set of all q for which $A_{\mathbf{x}}[\mathfrak{a}(q)] \leqslant a$ is B-measurable for any a; the A-measurability of $q \to A_{\mathbf{x}}[\mathfrak{a}(q)]$ is understood analogously.)

Remark 4.5. For $\Re = \{1, \dots, k\}$ considered in Remarks 4.2 and 4.4, we have $dA_{\mathbf{x}}(i) = \alpha(i, \mathbf{x})$, $dB_{\mathbf{y}}(i) = \beta(i, \mathbf{y})$, and (2) has the form

$$AB_{\mathbf{xy}}(\mathfrak{S}) = \sum_{i=1}^{k} \sum_{j \in \mathfrak{a}(q)} \alpha(j, \mathbf{x}) \beta(i, \mathbf{y})$$
$$= \sum_{i=1}^{k} \sum_{j \in \mathfrak{b}(p)} \beta(j, \mathbf{y}) \alpha(i, \mathbf{x}).$$

In particular, if the decision rule is "**x** and **y** are different if and only if $P(\mathbf{x}) \neq Q(\mathbf{y})$", then the meaning of both $j \in \mathfrak{a}(q)$ in the middle expression and $j \in \mathfrak{b}(p)$ in the right-hand expression is $j \neq i$.

As a simple example of the decision area $\mathfrak{S}$ when the perceptual space $\Re$ is a subset of Re^m, let $D(p,q)$ be some continuous metric imposed on Re^m, and assume that **y** is judged to be different from **x** if and only if

$$D[P(\mathbf{x}), Q(\mathbf{y})] > \varepsilon,$$

where ε is a fixed positive constant. Then

$$\mathfrak{S} = \{(p,q) \in \Re^2: D(p,q) > \varepsilon\}.$$

Another simple example, for an arbitrary perceptual space $\Re$, is provided by the "category-based" discriminations. Let $\Re$ be partitioned into several measurable areas, $\{\mathfrak{r}_1, \dots, \mathfrak{r}_k\}$, representing different categories, and let **y** be judged to be different from **x** if and only if their random images $P(\mathbf{x})$ and $Q(\mathbf{y})$ fall within two different category areas. Then

$$\mathfrak{S} = \{(p,q) \in \Re^2: (p,q) \notin \mathfrak{r}_1^2 \cup \dots \cup \mathfrak{r}_k^2\}.$$

The Thurstonian-type models with this decision rule are in most respects indistinguishable from the finite-space models, with $\Re = \{1, \dots, k\}$ and the decision rule mentioned in Remark 4.5.

One may consider much more complex decision rules, but as long as they remain deterministic (i.e., a given pair (p,q) always evokes a given response) they can always be presented as areas $\mathfrak{S} \subseteq \Re^2$. In particular, the formulation (algorithm, logic) of a rule may very well change from one pair of (p,q)-values to another ("image-dependent" decision rules). The analysis to follow applies to all such situations.

Remark 4.6. The random images $P(\mathbf{x})$, $Q(\mathbf{y})$ in this paper are also referred to as "random variables". It is difficult to avoid using this term in a probabilistic context. The standard "Kolmogorovian" definition of a random variable, however, is a real-valued measurable function on a sample space. Clearly our meaning is much broader: the sample space we deal with is $\Re$, and the values of, say, $P(\mathbf{x})$ are simply elements of $\Re$. The term "distribution of $P(\mathbf{x})$" is understood as synonymous with the measure function $A_{\mathbf{x}}(\mathfrak{p})$, $\mathfrak{p} \in \Sigma_A$. Formally, one can define $P(\mathbf{x})$ as the identity function from $\Re$ onto $\Re$, associated with the probability measure $A_{\mathbf{x}}$. Strictly speaking, however, the notion of a random image $P(\mathbf{x})$ is redundant, and is used here only because $\Pr[P(\mathbf{x}) \in \mathfrak{p}]$ (the probability of $P(\mathbf{x})$ falling within $\mathfrak{p}$) is more intuitive than $A_{\mathbf{x}}(\mathfrak{p})$ (the probability of $\mathfrak{p}$ "occurring"). Clearly, if $\Re$ is a subset of Re^m, $P(\mathbf{x})$ is a random m-vector in the conventional sense.

5. The universal Thurstonian-type representability

The theorem given in this section says that in the absence of additional restrictions the idea of the Thurstonian-type representability for $\psi(\mathbf{x}, \mathbf{y})$ (with stochastically independent perceptual images and deterministic decisions) is not a falsifiable assumption. Rather it is a theoretical language providing an alternative description for $\psi(\mathbf{x}, \mathbf{y})$: *any* function ψ mapping $\mathfrak{M} \times \mathfrak{M}$ into probabilities can be generated by *some* Thurstonian-type model. The proof is constructive: it provides a simple procedure by which, given a function $\psi(\mathbf{x}, \mathbf{y})$, one can construct a space $\Re$, probability measures $A_{\mathbf{x}}, B_{\mathbf{y}}$ on it, and a decision area $\mathfrak{S}$, such that the model $\{\Re, A_{\mathbf{x}}, B_{\mathbf{y}}, \mathfrak{S}\}$ generates $\psi(\mathbf{x}, \mathbf{y})$ precisely.

It is worthwhile to precede the general proof by demonstrating and discussing this procedure on the following example. Let $\mathfrak{M} = (0,1)$, that is, stimuli x, y vary between 0 and 1, and let $\psi(x,y)$ be an arbitrary discrimination probability function. Assume that the random image $P(x)$ of stimulus x is (p^1, p^2), where $p^1 = x$ while p^2 is uniformly distributed between 0 and 1. Analogously, assume that the random image $Q(y)$ is (q^1, q^2), where $q^1 = y$ and q^2 is uniformly distributed between 0 and 1 (independently of p^2). The perceptual space $\Re$ here is the unit square $(0,1)^2$. Consider the following decision rule: y is judged to be different from x if and only if

$$0<p^2<\sqrt{\psi(p^1,q^1)} \quad \text{and} \quad 0<q^2<\sqrt{\psi(p^1,q^1)}. \qquad (4)$$

A moment's contemplation will reveal that the probability of this happening is $\psi(x,y)$.

Some readers may feel uncomfortable about the fact that the first component of the image $P(x)=(p^1,p^2)$ in this construction is a deterministic quantity ($p^1=x$). Clearly, the perceiver could achieve perfect discriminability by simply ignoring the second component of $P(x)$. It would be a mistake, however, to dismiss the model above on the grounds that its decision rule is "irrational" or "suboptimal". We expressly lack any basis for applying these concepts, as they only make sense with respect to a particular set of operations that the perceiver is posited to have the option of applying to the elements of $\mathfrak{R}$. Thus, it does not follow from the fact that (p^1,p^2) uniquely determines the value of $\exp(p^1)+\exp(p^2)$ that the perceiver is able to use this number for making a judgment. By the same logic, it does not follow from the fact that a perceptual image can be represented by a numerical vector that the perceiver is able to use this vector's components in all conceivable computations or judgments involving them. In general, given a Thurstonian-type representation, the only operation that the perceiver can and should be posited to perform is the computation, for every pair of images (p,q), of whether this pair belongs or does not belong to one specific decision area $\mathfrak{S}$. In the construction above, for example, the observer is assumed to compute the truth value of the statement (4), but this does not imply that this observer can extract and utilize p^2, q^2, or $\psi(p^1,q^1)$ in any other computation.

Note that one can easily modify this construction to make both components of the random images stochastic. Assume, for example, that $P(x)=(p^1,p^2)$ and $Q(y)=(q^1,q^2)$ are defined by

$$p^1=\frac{v-x}{2},\quad p^2=\frac{v+x}{2},$$
$$q^1=\frac{w-y}{2},\quad q^2=\frac{w+y}{2},$$

where v,w are independent random variables uniformly distributed between 0 and 1. Modify the decision rule accordingly: y is judged to be different from x if and only if

$$0<p^1+p^2<\sqrt{\psi(p^2-p^1,q^2-q^1)},$$
$$0<q^1+q^2<\sqrt{\psi(p^2-p^1,q^2-q^1)}.$$

It is easy to see that this model is equivalent to the previous one, because of which it too generates $\psi(x,y)$ precisely.

Theorem 5.1. *Any function* $\psi:(\mathbf{x},\mathbf{y})\to[0,1]$ *has a Thurstonian-type representation* $\{\mathfrak{R},A_\mathbf{x},B_\mathbf{y},\mathfrak{S}\}$ (*with stochastically independent images and a deterministic decision rule*).

Proof. With no loss of generality, let $\mathbf{x},\mathbf{y}\in\mathfrak{M}\subseteq(0,1)^n$ (this can always be achieved by a suitable homeomorphic transformation of $\mathfrak{M}$). Define the perceptual space $\mathfrak{R}$ as

$$\mathfrak{R}=\mathfrak{M}\times(0,1)\subseteq(0,1)^{n+1},$$

with the conventional Lebesgue sigma-algebra Σ on it. Denote the elements of $\mathfrak{R}$ by $\mathbf{p}=(\mathbf{p}',p^{n+1})$, with $\mathbf{p}'=(p^1,\dots,p^n)$. Note that $\mathbf{p}'$ may or may not belong to the set $\mathfrak{M}\subseteq(0,1)^n$.

For any $\mathbf{p}'\in\mathfrak{M},\mathbf{q}'\in\mathfrak{M}$, define $\mathfrak{s}(\mathbf{p}',\mathbf{q}')$ as the square region

$$\mathfrak{s}(\mathbf{p}',\mathbf{q}')=(0,\sqrt{\psi(\mathbf{p}',\mathbf{q}')})\times(0,\sqrt{\psi(\mathbf{p}',\mathbf{q}')}),$$

and choose the decision area $\mathfrak{S}\subseteq\mathfrak{R}^2\subseteq(0,1)^{2n+2}$ to be

$$\mathfrak{S}=\{(\mathbf{p},\mathbf{q}):\ \mathbf{p}'\in\mathfrak{M},\mathbf{q}'\in\mathfrak{M},(p^{n+1},q^{n+1})\in\mathfrak{s}(\mathbf{p}',\mathbf{q}')\}.$$

(In this construction $\mathfrak{s}(\mathbf{p}',\mathbf{q}')$ could be replaced with any other subset of $(0,1)^2$ whose Lebesgue measure is $\psi(\mathbf{p}',\mathbf{q}')$.)

Put $\Sigma_A=\Sigma_B=\Sigma$, the set of all Lebesgue-measurable subsets. For $\mathfrak{r}\in\Sigma$, define

$$A_\mathbf{x}(\mathfrak{r})=\int_{\mathbf{p}\in\mathfrak{r}}\alpha(\mathbf{p},\mathbf{x})\,d\mathbf{p},\quad B_\mathbf{y}(\mathfrak{r})=\int_{\mathbf{q}\in\mathfrak{r}}\alpha(\mathbf{q},\mathbf{y})\,d\mathbf{q},$$

with the (generalized) density functions α and β defined as

$$\alpha(\mathbf{p},\mathbf{x})=\delta(\mathbf{p}'-\mathbf{x}),\quad \beta(\mathbf{q},\mathbf{y})=\delta(\mathbf{q}'-\mathbf{y}),$$

where δ is the (multivariate) Dirac delta function. $A_\mathbf{x}(\mathfrak{r})$ is a legitimate probability measure on Σ, because $\alpha(\mathbf{p},\mathbf{x})\geqslant 0$, and

$$\int_{\mathbf{p}\in\mathfrak{R}}\alpha(\mathbf{p},\mathbf{x})\,d\mathbf{p}$$
$$=\int_{\mathbf{p}'\in\mathfrak{M}}\int_0^1\delta(\mathbf{p}'-\mathbf{x})\,dp^{n+1}\,d\mathbf{p}'$$
$$=\int_{\mathbf{p}'\in\mathfrak{M}}\delta(\mathbf{p}'-\mathbf{x})\left(\int_0^1 dp^{n+1}\right)d\mathbf{p}'=1;$$

analogously for $B_{\mathbf{y}}(\mathfrak{r})$. Then

$$
\begin{aligned}
&AB_{\mathbf{xy}}(\mathfrak{S})\\
&= \int_{(\mathbf{p},\mathbf{q})\in\mathfrak{S}} \alpha(\mathbf{p},\mathbf{x})\beta(\mathbf{q},\mathbf{y})\, d\mathbf{p}\, d\mathbf{q}\\
&= \int_{\mathbf{p}'\in\mathfrak{M}} \int_{\mathbf{q}'\in\mathfrak{M}} \int_{(p^{n+1},q^{n+1})\in\mathfrak{s}(\mathbf{p}',\mathbf{q}')} \delta(\mathbf{p}'-\mathbf{x})\\
&\quad \times \delta(\mathbf{q}'-\mathbf{y})\, dq^{n+1}\, dp^{n+1}\, d\mathbf{q}'\, d\mathbf{p}'\\
&= \int_{\mathbf{p}'\in\mathfrak{M}} \int_{\mathbf{q}'\in\mathfrak{M}} \delta(\mathbf{p}'-\mathbf{x})\delta(\mathbf{q}'-\mathbf{y})\\
&\quad \times \left(\int_{(p^{n+1},q^{n+1})\in\mathfrak{s}(\mathbf{p}',\mathbf{q}')} dq^{n+1}\, dp^{n+1}\right) d\mathbf{q}'\, d\mathbf{p}'\\
&= \int_{\mathbf{p}'\in\mathfrak{M}} \int_{\mathbf{q}'\in\mathfrak{M}} \delta(\mathbf{p}'-\mathbf{x})\delta(\mathbf{q}'-\mathbf{y})\psi(\mathbf{p}',\mathbf{q}')\, d\mathbf{q}'\, d\mathbf{p}'\\
&= \psi(\mathbf{x},\mathbf{y}).
\end{aligned}
$$

This completes the construction of a Thurstonian-type representation for $\psi(\mathbf{x},\mathbf{y})$. □

Remark 5.1. To generalize the concluding part of the discussion that precedes the theorem, once a Thurstonian-type representation for $\psi(\mathbf{x},\mathbf{y})$ is constructed, one can construct another representation by (a) replacing $\mathfrak{R}$ with $\mathfrak{R}^* = F(\mathfrak{R})$, where F is any one-to-one transformation, (b) defining the sigma-algebra Σ^* on $\mathfrak{R}^*$ by the rule "$\mathfrak{r}\in\Sigma^*$ if and only if $F^{-1}(\mathfrak{r})\in\Sigma$" (which makes F a measurable mapping), (c) putting $A^*_{\mathbf{x}}(\mathfrak{r}) = A_{\mathbf{x}}[F^{-1}(\mathfrak{r})]$, $B^*_{\mathbf{y}}(\mathfrak{r}) = B_{\mathbf{y}}[F^{-1}(\mathfrak{r})]$, and (d) defining $\mathfrak{S}^*\subseteq\mathfrak{R}^*\times\mathfrak{R}^*$ by the rule "$(p,q)\in\mathfrak{S}^*$ if and only if $(F^{-1}(p),F^{-1}(q))\in\mathfrak{S}$". Then, obviously, $A^*B^*_{\mathbf{xy}}(\mathfrak{S}^*) = AB_{\mathbf{xy}}(\mathfrak{S}) = \psi(\mathbf{x},\mathbf{y})$.

It must be emphasized that the sole purpose of this theorem is to show that *some* Thurstonian-type representation exists for every imaginable discrimination probability function. It is hardly worth mentioning that the construction described in the theorem (or any of its modifications mentioned in Remark 5.1) is of *no* interest to a model-builder: the probability measures $A_{\mathbf{x}}$ and $B_{\mathbf{y}}$ used in the proof are singular, concentrated on sets of measure zero, and the decision rule is void of any interpretability. An adherent of the idea that stimuli are mapped into random variables is likely to think of more regular, "better behaved" random variables, like multivariate or univariate normal distributions, combined with some easily interpretable decision rule, such as the distance-based or category-based rules mentioned at the end of Section 4 (Dai, Versfeld, & Green, 1996; Ennis, 1992; Ennis et al., 1988; Luce & Galanter, 1963; Sorkin, 1962; Suppes & Zinnes, 1963; Thomas, 1996, 1999; Zinnes & MacKey, 1983).

Unfortunately for this line of modeling, it is doomed to fail. As shown in this paper, even if decision rules are not constrained to be reasonable or interpretable, no realistic discrimination probability function can be accounted for by sufficiently "nice" distributions with parameters sufficiently "nicely" dependent on stimuli.

Remark 5.2. In relation to the notion of well-behaved Thurstonian-type representations (to be defined in Section 7), it is useful to observe just in what respect the distribution constructed in Theorem 5.1 fails to be "nice". For every random variable in $\mathfrak{R}$ one can choose a measurable subset $\mathfrak{r}$ of $\mathfrak{R}$ and ask with what probability $A_{\mathbf{x}}(\mathfrak{r})$ the random variable falls in this subset. With $\mathfrak{R}$ and $A_{\mathbf{x}}(\mathfrak{r})$ defined as in Theorem 5.1, choose

$$\mathfrak{r}_0 = \{\mathbf{p} = (\mathbf{p}', p^{n+1}):\ \mathbf{p}' = (p_0^1,\ldots,p_0^n),\ p^{n+1}\in(0,1)\}.$$

Then $A_{\mathbf{x}}(\mathfrak{r}_0) = 0$ for any $\mathbf{x}\neq(p_0^1,\ldots,p_0^n)$, but the value of $A_{\mathbf{x}}(\mathfrak{r})$ "jumps" to 1 as soon as $\mathbf{x} = (p_0^1,\ldots,p_0^n)$. This violates the main intuition behind the notion of well-behavedness, which is that for any $\mathfrak{r}$ the value of $A_{\mathbf{x}}(\mathfrak{r})$ must change with $\mathbf{x}$ continuously.

6. Patches of discrimination probabilities

Here, I introduce a local parametrization of stimuli that greatly simplifies the development by replacing in all subsequent arguments the n-component stimuli $\mathbf{x} = (x^1,\ldots,x^n)$, $\mathbf{y} = (y^1,\ldots,y^n)$ with certain unidimensional representations thereof, denoted x, y.

This local parametrization is achieved by choosing a particular stimulus $\mathbf{s}\in\mathfrak{M}\subseteq\mathrm{Re}^n$, a nonzero direction of change $\mathbf{u}\in\mathrm{Re}^n$ (a stimulus–direction combination $(\mathbf{s},\mathbf{u})$ is traditionally called a *line element*), and considering values of $\mathbf{x}$ varying within an arbitrarily small vicinity of $\mathbf{s}$ along the direction $\mathbf{u}$,

$$\mathbf{x} = \mathbf{s} + \mathbf{u}x,\quad x\in[-a,a],\ a>0.$$

Recall that $\mathfrak{M}$ is open, because of which one can always find a sufficiently small a (generally depending on $\mathbf{s},\mathbf{u}$) such that $\mathbf{x} = \mathbf{s} + \mathbf{u}x$ belongs to $\mathfrak{M}$ for all $x\in[-a,a]$. Clearly, for a fixed line element $(\mathbf{s},\mathbf{u})$ the value of $\mathbf{x}$ is uniquely encoded by x. Since we are interested in an arbitrarily small vicinity of $\mathbf{s}$, the precise value of a is never the issue: it can be made as small as one wishes.

Let now $\mathbf{h}$ be a fixed homeomorphism $\mathfrak{M}\to\mathfrak{M}$. Then, for any $y\in[-a,a]$, the value $\mathbf{y} = \mathbf{h}(\mathbf{s}+\mathbf{u}y)$ is uniquely encoded by y. As y varies within $[-a,a]$, the values of $\mathbf{y}$ form a segment of a continuous curve lying within $\mathfrak{M}$. The set

$$\{(\mathbf{x},\mathbf{y}):\ \mathbf{x} = \mathbf{s}+\mathbf{u}x,\ \mathbf{y} = \mathbf{h}(\mathbf{s}+\mathbf{u}y),\ (x,y)\in[-a,a]^2\}$$

then belongs to $\mathfrak{M}^2$, that is, consists of stimulus pairs. Once the line element $(\mathbf{s},\mathbf{u})$ and the homeomorphism $\mathbf{h}$ are fixed, all stimulus pairs within this set are uniquely parametrized by $(x,y)\in[-a,a]^2$. In particular, any two *corresponding* stimuli $\mathbf{x}$ and $\mathbf{y} = \mathbf{h}(\mathbf{x})$ are encoded by equal local coordinates $-a \leqslant x = y \leqslant a$, and the "central" pair $(\mathbf{s}, \mathbf{h}(\mathbf{s}))$ is represented by $x = y = 0$.

Remark 6.1. Anticipating the material of Section 9 (although sufficiently prompted by the discussion in Section 3), the homeomorphism $\mathbf{h}$ that we are interested in is the one that relates $\mathbf{x}$ to its PSE $\mathbf{y} = \mathbf{h}(\mathbf{x})$. At such pairs of stimuli the functions $\mathbf{x} \to \psi(\mathbf{x}, \mathbf{y})$ and $\mathbf{y} \to \psi(\mathbf{x}, \mathbf{y})$ reach their minima,

$$\mathbf{h}(\mathbf{x}) = \arg\min_{\mathbf{y}} \psi(\mathbf{x}, \mathbf{y}), \quad \mathbf{h}^{-1}(\mathbf{y}) = \arg\min_{\mathbf{x}} \psi(\mathbf{x}, \mathbf{y}).$$

The local parametrization being constructed, therefore, focuses one's attention on the stimulus values varying in arbitrarily small vicinities of the minima of $\psi(\mathbf{x}, \mathbf{y})$. For now, however (until Section 10), this interpretation of $\mathbf{h}$ is not essential, and $\mathbf{h}$ can be taken as an arbitrary fixed homeomorphism.

The function

$$\psi_{(\mathbf{s},\mathbf{u})}(x, y) = \psi(\mathbf{s} + \mathbf{u}x, \mathbf{h}(\mathbf{s} + \mathbf{u}y)) \tag{5}$$

defined on $(x, y) \in [-a, a]^2$ is called a *patch* of the discrimination probability function $\psi(\mathbf{x}, \mathbf{y})$ at the line element $(\mathbf{s}, \mathbf{u})$. For a fixed $(\mathbf{s}, \mathbf{u})$, it is convenient to drop the index in (5) and to simply speak of a patch $\psi(x, y)$, omitting or mentioning $(\mathbf{s}, \mathbf{u})$ and the patch domain $[-a, a]^2$ as needed. The homeomorphism $\mathbf{h}$ is assumed to be fixed throughout and need not be mentioned at all.

For any patch $\psi(x, y)$ of $\psi(\mathbf{x}, \mathbf{y})$ taken at some line element $(\mathbf{s}, \mathbf{u})$, one can rewrite $A_{\mathbf{x}}$, $B_{\mathbf{y}}$, $AB_{\mathbf{xy}}$ as $A_x(p)$, $B_y(q)$, AB_{xy}, to present the basic relationship (2) as

$$\psi(x, y) = AB_{xy}(\mathfrak{S}) = \int_{q \in \mathfrak{R}} A_x[\mathfrak{a}(q)]\, dB_y(q) = \int_{p \in \mathfrak{R}} B_y[\mathfrak{b}(p)]\, dA_x(p), \tag{6}$$

where $(p, q) \in \mathfrak{R}^2$, $(x, y) \in [-a, a]^2$. Recall, from Section 4, that the functions $q \to A_x[\mathfrak{a}(q)]$ and $p \to B_y[\mathfrak{b}(p)]$ are, respectively, B-measurable and A-measurable.

It is important to observe that representation (2) may fail to hold for $\psi(\mathbf{x}, \mathbf{y})$ even if (6) holds for all possible patches $\psi(x, y)$. For one thing, the stimulus space $\mathfrak{M}$ and the homeomorphism $\mathbf{h}$ may very well be such that some stimulus pairs $(\mathbf{x}, \mathbf{y})$ will not be covered by any of the patches. Also, it is possible that the perceptual space $\mathfrak{R}$, probability measures $A_x(p)$, $B_y(q)$, and a decision area $\mathfrak{S}$ satisfying (6) can be found for any patch $\psi(x, y)$, taken separately, but cannot be combined into a single Thurstonian-type representation across all the patches (this will happen whenever, say, $A_x(p)$ is different on the common part of two overlapping patches). This is the reason why Theorem 5.1, asserting the existence of a Thurstonian-type representation, has to be proved for the entire discrimination probability function $\psi(\mathbf{x}, \mathbf{y})$ rather than for its arbitrary patch $\psi(x, y)$.

At the same time, representation (2) holds *only if* (6) holds for all possible patches $\psi(x, y)$ of $\psi(\mathbf{x}, \mathbf{y})$. Equivalently put, if (6) does not hold for a single patch $\psi(x, y)$, then (2) does not hold for $\psi(\mathbf{x}, \mathbf{y})$. Therefore to prove that $\psi(\mathbf{x}, \mathbf{y})$ subject to the regular minimality and nonconstant self-similarity constraints (discussed in Section 3 and defined in Section 9) cannot have a "well-behaved" Thurstonian-type representation (defined in the next section), it is sufficient to show that there is at least one patch $\psi(x, y)$ of this $\psi(\mathbf{x}, \mathbf{y})$ that does not have a "well-behaved" Thurstonian-type representation. In fact, the development to follow shows this to be true for any "typical" patch $\psi(x, y)$ (as defined in Section 9).

7. Well-behavedness

Having fixed a line element $(\mathbf{s}, \mathbf{u})$, consider a probabilistic measure

$$A_x(\mathfrak{p}) = \Pr[P(x) \in \mathfrak{p}],$$

with x varying, in accordance with the "patch-wise" view just introduced, within an interval $[-a, a]$ that can be made arbitrarily small.

A natural intuition for the "well-behavedness" of $A_x(\mathfrak{p})$ is that, for any fixed event (= measurable set) $\mathfrak{p}$, its probability $A_x(\mathfrak{p})$ must change "sufficiently smoothly" in response to transitions $x \to x + dx$ within a very small $[-a, a]$. In the definition below this intuition is translated into the existence of unilateral derivatives $\frac{\partial}{\partial x+} A_x(\mathfrak{p})$ and $\frac{\partial}{\partial x-} A_x(\mathfrak{p})$. In practice, we almost always translate the idea of "sufficient smoothness" into piecewise continuous differentiability (i.e., continuous differentiability with a countable number of isolated exceptions). Thus, a freehand drawing of a continuous curve, however "jittery", is always parametrically representable by a pair of piecewise continuously differentiable functions. It seems that all continuous functions of applied mathematics are piecewise continuously differentiable. The requirement of unilateral differentiability adopted in this paper is even less stringent. In accordance with Lemma A.1, for any given $\mathfrak{p}$, the set of points $x \in [-a, a]$ where $\frac{\partial}{\partial x+} A_x(\mathfrak{p}) \neq \frac{\partial}{\partial x-} A_x(\mathfrak{p})$ is at most denumerable (i.e., empty, finite, or countably infinite), but it need not be a set of isolated points. Outside this countable set, $A_x(\mathfrak{p})$ is differentiable on $[-a, a]$ in the conventional sense, but its derivative is not required to be continuous.

The unilateral differentiability of $A_x(\mathfrak{p})$ prevents the latter from "jumping" in response to $x \to x + dx$. One's intuition of well-behavedness, however, requires, in addition, that $A_x(\mathfrak{p})$ may not come arbitrarily close to "jumping": $A_x(\mathfrak{p})$ is *not* well-behaved if, for any

$[-a, a]$, however small,

$$\left|\frac{A_x(\mathfrak{p}) - A_{x'}(\mathfrak{p})}{x - x'}\right|$$

can be made arbitrarily large by an appropriate choice of $\mathfrak{p}$ and of two distinct x, x' within $[-a, a]$. This means that a well-behaved $A_x(\mathfrak{p})$ should satisfy the *Lipschitz condition*:

$$\left|\frac{A_{x'}(\mathfrak{p}) - A_x(\mathfrak{p})}{x' - x}\right| < \text{const}$$

for some choice of a and all $x, x' \in [-a, a]$, $\mathfrak{p} \in \Sigma_A$. By a standard calculus argument, if $A_x(\mathfrak{p})$ is unilaterally differentiable, the Lipschitz condition is equivalent to the boundedness of the unilateral derivatives:

$$\left|\frac{\partial}{\partial x \pm} A_x(\mathfrak{p})\right| < \text{const}$$

for all $x \in [-a, a]$ and $\mathfrak{p} \in \Sigma_A$.

It is important to note that, by Lemma A.2, the Lipschitz condition implies that the unilateral derivatives $\frac{\partial}{\partial x \pm} A_x(\mathfrak{p})$ exist (and are then necessarily bounded) at all points of $[-a, a]$, except, perhaps, on a set of measure zero. The intuition of "sufficient smoothness" therefore can also be presented as an additional regularization of the Lipschitz condition, the requirement that the exceptional sets of measure zero be empty for all $\mathfrak{p} \in \Sigma_A$.

Definition 7.1. A probabilistic measure $A_x(\mathfrak{p})$, $\mathfrak{p} \in \Sigma_A$, $x \in [-a, a]$, is *well-behaved* if the left- and right-hand derivatives

$$D_A^{\pm}(\mathfrak{p}, x) = \frac{\partial A_x(\mathfrak{p})}{\partial x \pm}$$

exist and are bounded on $\Sigma_A \times [-a, a]$, that is,

$$|D_A^{\pm}(\mathfrak{p}, x)| \leqslant c < \infty.$$

The well-behavedness of $B_y(\mathfrak{q})$, $\mathfrak{q} \in \Sigma_B$, $y \in [-a, a]$, is defined analogously, with the derivatives denoted as $D_B^{\pm}(\mathfrak{q}, y)$.

A Thurstonian-type representation $\{\mathfrak{R}, A_x, B_y, \mathfrak{S}\}$ for a patch $\psi(x, y)$, $(x, y) \in [-a, a]^2$, is *well-behaved* if $A_x(\mathfrak{p})$ and $B_y(\mathfrak{q})$ are well-behaved.

Remark 7.1. As shown in Section 11, this definition can be significantly relaxed without affecting the results. In view of this generalization, the notion just defined can be called the "well-behavedness in the narrow (or absolute) sense".

Remark 7.2. The existence of left-hand (right-hand) derivatives at $-a$ (respectively, a) is not a problem, as a can always be made smaller than any previously chosen value.

Remark 7.3. Clearly, $A_x(\mathfrak{p})$ and $B_y(\mathfrak{q})$ are continuous in x and y (in fact, uniformly continuous, since $[-a,\ a]$ is closed).

Remark 7.4. The local parametrization (x, y) as introduced in Section 6 depends, for any given correspondence function $\mathbf{h}$, on the global parametrization $\mathfrak{M}$ of the stimulus space. As the well-behavedness involves differentiation, it will not be preserved under all allowable (homeomorphic) transformation of $\mathfrak{M}$. The requirement that a Thurstonian-type representation be well-behaved, therefore, should be taken as applying to *some* parametrization $\mathfrak{M}$ rather than to *all* of them. Clearly, if a Thurstonian-type model fails to account for a single patch $\psi(x, y)$ within a particular choice of $\mathfrak{M}$, it fails to account for the discrimination probability function $\psi(\mathbf{x}, \mathbf{y})$ defined on any allowable choice of $\mathfrak{M}$.

Most "textbook" distributions (normal, Weibull, gamma, etc.), univariate or multivariate, with their parameters depending on x in a piecewise continuously differentiable fashion (most likely candidates for Thurstonian-type models designed to fit empirical data) can be shown to be well-behaved in the sense of Definition 7.1. In most such cases (including the simple examples given in Section 3) one can show that the density of the random variable in question is right- and left-differentiable with respect to x at any point of $\mathfrak{R} \subseteq \mathrm{Re}^m$ $(m \geqslant 1)$. It is easy to prove (see Lemma A.4) that if a measure $A_x(\mathfrak{p})$ has a density $\alpha(p, x)$ with respect to some sigma-finite measure $M(\mathfrak{p})$, and if $\frac{\partial}{\partial x \pm} \alpha(p, x)$ exist and are dominated by a function integrable on the entire space $\mathfrak{R}$, then $A_x(\mathfrak{p})$ is well-behaved in the sense of Definition 7.1. (In view of the results arrived at below, this explains the failure of the examples given in Section 3.) The definition, however, also covers distributions whose densities generally are not unilaterally differentiable in x, due to having discontinuities of the first kind (finite "jumps") whose positions on $\mathfrak{R}$ change with x (as, e.g., in a uniform distribution with stimulus-dependent endpoints, or a shifted exponential distribution with stimulus-dependent shift).

Plainly, if parameters of an otherwise "nice" distribution change as a function of x discontinuously, then the value of $A_x(\mathfrak{p})$ within some set $\mathfrak{p}$ may "jump", which would put it outside the class of well-behaved probability measures. Singular distributions whose support changes with x (like the ones used in the proof of Theorem 5.1) are, of course, outside the class of well-behaved ones (see Remark 5.2).

As pointed out to me by A. Eremenko (personal communication, 2001), there exist absolutely continuous probability measures whose parameters change smoothly but that are not well-behaved in the sense of Definition 7.1. These can be found among distributions whose densities have discontinuities of the second kind (infinite "jumps") and the position of these singularities in the space $\mathfrak{R}$ changes as a function of x. As an example, taking $\mathfrak{R} = \mathrm{Re}$, the density

$$\alpha(p,x) = \begin{cases} \frac{1}{2\sqrt{p-x}} & \text{if } x < p \leqslant x+1, \\ 0 & \text{if otherwise}, \end{cases} \quad x \in [-a,a]$$

has a second-kind discontinuity at $p = x$. Choosing, say, $\mathfrak{p} = (-\frac{a}{2}, \frac{a}{2})$, one observes that

$$\frac{\partial}{\partial x-} A(\mathfrak{p},x)\bigg|_{x=-\frac{a}{2}} = \infty, \quad \frac{\partial}{\partial x-} A(\mathfrak{p},x)\bigg|_{x=\frac{a}{2}} = -\infty,$$

which contradicts Definition 7.1.

8. Near-smoothness theorem

The theorem presented in this section shows that a patch $\psi(x,y)$ generated by a well-behaved Thurstonian-type model possesses a certain smoothness property, a weak analog of componentwise continuous differentiability. For the lack of a better term, I call this property *near-smoothness*.

Definition 8.1. A patch $\psi(x,y)$, $-a \leqslant x, y \leqslant a$, is called *near-smooth* if it is both right- and left-differentiable in both x and y, with $\frac{\partial}{\partial x\pm}\psi(x,y)$ being continuous in y and $\frac{\partial}{\partial y\pm}\psi(x,y)$ being continuous in x.

All smooth (continuously differentiable) functions are near-smooth. Simple examples of non-smooth but near-smooth functions are $|x|+|y|$, $|xy|$, $1-\exp[-(|x|-|y|)^2]$, etc. $(-a \leqslant x, y \leqslant a)$. A simple example of a unilaterally differentiable but not near-smooth function is $|x-y|$. Clearly, a near-smooth $\psi(x,y)$ is componentwise continuous.

Theorem 8.1. *A patch $\psi(x,y)$ that has a well-behaved Thurstonian-type representation is near-smooth.*

Proof. We prove that the derivatives $\frac{\partial}{\partial y\pm}\psi(x,y)$ exist and are continuous in x. The proof that $\frac{\partial}{\partial x\pm}\psi(x,y)$ exist and are continuous in y is obtained by symmetrical argument.

Existence: By (6),

$$\psi(x,y) = AB_{xy}(\mathfrak{S}) = \int_{p\in\mathfrak{R}} B_y[\mathfrak{b}(p)]\, dA_x(p).$$

To prove that $\frac{\partial}{\partial y\pm}\psi(x,y)$ exist, observe that $\mathfrak{b}(p) \in \Sigma_B$, and, by Definition 7.1, $\frac{\partial}{\partial y\pm}B_y(\mathfrak{b}) = D_B^{\pm}(\mathfrak{b},y)$ are dominated on $\Sigma_B \times [-a,a]$ by a constant c. Hence

$$|D_B^{\pm}[\mathfrak{b}(p),y]| \leqslant c$$

on $\mathfrak{R} \times [-a,a]$. Since

$$\int_{p\in\mathfrak{R}} c\, dA_x(p) = c < \infty,$$

we apply Lemma A.3 to obtain

$$\frac{\partial}{\partial y\pm}\psi(x,y) = \int_{p\in\mathfrak{R}} D_B^{\pm}[\mathfrak{b}(p),y]\, dA_x(p).$$

Continuity: To prove that these derivatives are continuous in x, observe first that being the limits of the A-measurable functions

$$\frac{B(\mathfrak{b}(p), y\pm\varepsilon) - B(\mathfrak{b}(p),y)}{\varepsilon}, \quad \varepsilon \to 0+,$$

the functions $p \to D_B^{\pm}[\mathfrak{b}(p),y]$ are A-measurable. Fix y, and rewrite, for simplicity, $D_B^{\pm}[\mathfrak{b}(p),y]$ as $b(p)$. Since $b(p)$ is A-measurable and bounded, there is (see, e.g., Hewitt & Stromberg, 1965, pp. 172–173) a sequence of A-measurable simple functions

$$\varphi_i(p) = \sum_{j=1}^{n_i} b_{ij}\chi_{\mathfrak{p}_{ij}}(p),$$

with $\chi_{\mathfrak{p}_{ij}}$ being characteristic functions of pairwise disjoint A-measurable sets $\mathfrak{p}_{ij}$, such that $\varphi_i(p)$ converges to $b(p)$ uniformly. This means that there is a function $n(\varepsilon)$ such that

$$b(p) - \varepsilon \leqslant \varphi_i(p) \leqslant b(p) + \varepsilon$$

for all $i > n(\varepsilon)$. But then, for any $x \in [-a,a]$,

$$\int_{p\in\mathfrak{R}} [b(p)-\varepsilon]\, dA_x(p) \leqslant \int_{p\in\mathfrak{R}} \varphi_i(p)\, dA_x(p) \leqslant \int_{p\in\mathfrak{R}} [b(p)-\varepsilon]\, dA_x(p),$$

which is equivalent to

$$\int_{p\in\mathfrak{R}} b(p)\, dA_x(p) - \varepsilon \leqslant \int_{p\in\mathfrak{R}} \varphi_i(p)\, dA_x(p) \leqslant \int_{p\in\mathfrak{R}} b(p)\, dA_x(p) + \varepsilon.$$

By the construction logic of Lebesgue integrals,

$$\int_{p\in\mathfrak{R}} b(p)\, dA_x(p) = \lim_{i\to\infty} \int_{p\in\mathfrak{R}} \varphi_i(p)\, dA_x(p),$$

and the inequalities above indicate that this convergence is uniform on $x \in [-a,a]$. Now,

$$\int_{p\in\mathfrak{R}} \varphi_i(p)\, dA_x(p) = \sum_{j=1}^{n_i} \int_{p\in\mathfrak{R}} b_{ij}\chi_{\mathfrak{p}_{ij}}(p)\, dA_x(p) = \sum_{j=1}^{n_i} b_{ij} A_x(\mathfrak{p}_{ij}),$$

which is continuous in x, because so is $A_x(\mathfrak{p})$ for any A-measurable $\mathfrak{p}$. The limit of uniformly converging continuous functions being continuous, we have proved that $\frac{\partial}{\partial y\pm}\psi(x,y)$ are continuous in x. □

Remark 8.1. The significance of this theorem is in its obvious consequence: if a patch $\psi(x,y)$ of a discrimination probability function $\psi(\mathbf{x},\mathbf{y})$ at some $(\mathbf{s},\mathbf{u})$ is found not to be near-smooth, then one knows that $\psi(\mathbf{x},\mathbf{y})$ cannot be represented by any well-behaved Thurstonian-type model.

As shown below, there are compelling reasons to believe that "typical" patches of discrimination probability functions cannot be near-smooth.

9. Regular minimality and nonconstant self-similarity

In relation to unidimensional stimulus continua the two concepts in the title of this section are discussed in Section 3. In a general form they have been studied in Dzhafarov (2002d; for a brief summary see also Dzhafarov, 2001b) in the context of multidimensional Fechnerian scaling (Dzhafarov, 2002a,b,c; Dzhafarov & Colonius, 1999, 2001). This theory is not invoked in the present paper, except for some basic considerations related to (a weakened version of) the so-called *First Assumption* of multidimensional Fechnerian scaling. The reader is referred to Dzhafarov (2002d) for details of the theory and for a review of empirical evidence. Here, I will merely assert that the properties of regular minimality and nonconstant self-similarity seem to be corroborated by all available empirical data on "same-different" discrimination probabilities (Dzhafarov, 2002d; Indow, 1998; Indow, Robertson, von Grunau, & Fielder, 1992; Krumhansl, 1978; Rothkopf, 1957; Tversky, 1977; Zimmer & Colonius, 2000).

9.1. Regular minimality

Recall that stimuli $\mathbf{x},\mathbf{y}$ in $\psi(\mathbf{x},\mathbf{y})$ are assumed to belong to two distinct observation areas (spatial and/or temporal intervals), and the values of $\mathbf{x},\mathbf{y}$ vary within an open connected area $\mathfrak{M}\subseteq \mathrm{Re}^n$ $(n\geqslant 1)$.

It is part of the first assumption of multidimensional Fechnerian scaling that:

(i) for every $\mathbf{x}$, the function $\mathbf{y}\to\psi(\mathbf{x},\mathbf{y})$ achieves its global minimum at some value $\mathbf{y}=\mathbf{h}(\mathbf{x})$, $\mathbf{h}$ being continuous;
(ii) for every $\mathbf{y}$, the function $\mathbf{x}\to\psi(\mathbf{x},\mathbf{y})$ achieves its global minimum at some value $\mathbf{x}=\mathbf{g}(\mathbf{y})$, $\mathbf{g}$ being continuous;
(iii) $\mathbf{g}\equiv\mathbf{h}^{-1}$.

We say that $\psi(\mathbf{x},\mathbf{y})$ possesses the *regular minimality property* (or has *regular minima*) if this tripartite assumption is satisfied. Due to the identity $\mathbf{g}\equiv\mathbf{h}^{-1}$, both $\mathbf{g}$ and $\mathbf{h}$ are homeomorphisms $\mathfrak{M}\to\mathfrak{M}$. In the simplest case, of course, $\mathbf{h}$ and $\mathbf{g}$ are identity mappings (i.e., $\psi(\mathbf{x},\mathbf{y})$ achieves its minima at $\mathbf{x}=\mathbf{y}$).

Remark 9.1.1. In multidimensional Fechnerian scaling the functions $\mathbf{h},\mathbf{g}$ are assumed to be continuously differentiable (hence diffeomorphisms). We do not need this constraint in the present context.

As in the case of unidimensional stimuli, $\mathbf{y}=\mathbf{h}(\mathbf{x})$ in (i) can be called the PSE (in the second observation area) for $\mathbf{x}$ (belonging to the first observations area); analogously, $\mathbf{x}=\mathbf{g}(\mathbf{y})$ is the PSE (in the first observation area) for $\mathbf{y}$ (belonging to the second observations area). The symmetry of this relationship can also be presented as

$$\arg\min_{\mathbf{y}}\psi(\mathbf{x},\mathbf{y})=\mathbf{h}(\mathbf{x}) \Leftrightarrow \arg\min_{\mathbf{x}}\psi(\mathbf{x},\mathbf{y})=\mathbf{h}^{-1}(\mathbf{y}), \tag{7}$$

for all $\mathbf{x},\mathbf{y}$ in $\mathfrak{M}$.

Recall now the construction of local coordinates (x,y) at a line element $(\mathbf{s},\mathbf{u})$ introduced in Section 6, and take the correspondence homeomorphism $\mathbf{h}$ in that section to coincide with the PSE function $\mathbf{h}$ in (7). Since PSE pairs $(\mathbf{x},\mathbf{h}(\mathbf{x}))$ are encoded by equal local coordinates, $x=y$, relationship (7) implies that at any line element $(\mathbf{s},\mathbf{u})$,

$$\arg\min_{y}\psi(x,y)=x, \arg\min_{x}\psi(x,y)=y,$$

or, equivalently,

$$\psi(x,x)<\begin{cases}\psi(x,y)\\ \psi(y,x)\end{cases}, \tag{8}$$

for all $-a\leqslant x\neq y\leqslant a$. This can be called the regular minimality condition for patches $\psi(x,y)$.

9.2. Nonconstant self-similarity

One says that $\psi(\mathbf{x},\mathbf{y})$ possesses the *constant self-similarity property* if $\psi(\mathbf{x},\mathbf{h}(\mathbf{x}))$ does not depend on $\mathbf{x}$,

$$\psi(\mathbf{x},\mathbf{h}(\mathbf{x}))\equiv \text{const}.$$

It is a well-documented fact that "same-different" discrimination probabilities do not generally have this property (Dzhafarov, 2002d; Indow, 1998; Indow et al., 1992; Krumhansl, 1978; Rothkopf, 1957; Zimmer & Colonius, 2000). Our second assumption about $\psi(\mathbf{x},\mathbf{y})$ is therefore that it is subject to *nonconstant self-similarity*:

$$\psi(\mathbf{x},\mathbf{h}(\mathbf{x}))\not\equiv \text{const}. \tag{9}$$

The nonconstancy of $\psi(\mathbf{x},\mathbf{h}(\mathbf{x}))$ implies that at least at some line element $(\mathbf{s},\mathbf{u})$ the value of $\psi(\mathbf{s}+\mathbf{u}x,\mathbf{h}(\mathbf{s}+\mathbf{u}x))$ changes with x in the vicinity of $x=0$. Switching to the patch-wise view, this means that at least at some line elements $(\mathbf{s},\mathbf{u})$,

$$\psi(x,x)\not\equiv\text{const} \tag{10}$$

on any interval $-a\leqslant x\leqslant a$. A moment's reflection shows that if this was not true, that is, if for any $(\mathbf{s},\mathbf{u})$ one could find an interval $-a\leqslant x\leqslant a$ on which $\psi(x,x)$ was constant, then $\psi(\mathbf{x},\mathbf{h}(\mathbf{x}))$ would be forced to be constant throughout.

I refer to patches $\psi(x,y)$ for which (10) holds true as *typical patches* and to the line elements $(\mathbf{s},\mathbf{u})$ at which the patches are typical as *typical line elements*. Note that (8) holds for all patches and at all line elements, including the typical ones. Thus, a typical patch is subject to both regular minimality and nonconstant self-similarity.

10. "Non-near-smoothness" theorem

Recall Definition 8.1 of a near-smooth patch $\psi(x,y)$ and the obvious fact that near-smoothness implies componentwise continuity. For the theorem below we also need the (perhaps slightly less obvious) fact that if $\psi(x,y)$ is near-smooth, then the function $\psi(x,x)$ is continuous in x (Lemma A.5).

Theorem 10.1. *Let a discrimination probability function $\psi(\mathbf{x},\mathbf{y})$ be subject to both the regular minimality and nonconstant self-similarity constraints. Then a typical patch $\psi(x,y)$ of this function cannot be near-smooth.*

Remark 10.1. What the theorem says is that if $\psi(x,y)$ satisfies both (8) and (10), then it cannot be right- and left-differentiable in x and y, with $\frac{\partial}{\partial x\pm}\psi(x,y)$ continuous in y and $\frac{\partial}{\partial y\pm}\psi(x,y)$ continuous in x.

Remark 10.2. Following standard conventions, $\frac{\partial}{\partial x\pm}\psi(u,y)$ in the proof below stands for $\partial/\partial x\pm\psi(x,y)|_{x=u}$, $\frac{\partial}{\partial x\pm}\psi(u,u)$ for $\partial/\partial x\pm\psi(x,u)|_{x=u}$, etc.

Proof. Assume the contrary, and let a patch $\psi(x,y)$ be both typical and near-smooth on some $[-a,a]$. Let

$$\mathfrak{U}=\left\{u\in(-a,a)\colon \frac{\partial}{\partial x+}\psi(u,u)\neq\frac{\partial}{\partial x-}\psi(u,u)\right\}.$$

We show first that this set is at most denumerable (i.e., empty, finite, or countably infinite). Since $\frac{\partial}{\partial x+}\psi(u,y)$, $\frac{\partial}{\partial x-}\psi(u,y)$ are continuous in y, for any $u\in\mathfrak{U}$ there is an interval $\mathfrak{I}_u=(u-\delta_u,u+\delta_u)\subseteq(-a,a)$ such that $\frac{\partial}{\partial x+}\psi(u,y)\neq\frac{\partial}{\partial x-}\psi(u,y)$ for all $y\in\mathfrak{I}_u$. Consider the set $\mathfrak{Q}$ of all rationals in $(-a,a)$. For every $r\in\mathfrak{Q}$, let

$$\mathfrak{B}_r=\{u\in\mathfrak{U}\colon r\in\mathfrak{I}_u\}.$$

Then

$$\mathfrak{U}=\bigcup_{r\in\mathfrak{Q}}\mathfrak{B}_r,$$

because every $\mathfrak{B}_r$ is a subset of $\mathfrak{U}$, and every $u\in\mathfrak{U}$ belongs to some $\mathfrak{B}_r$, for one can find a rational point within any interval $\mathfrak{I}_u=(u-\delta_u,u+\delta_u)$. Since $\mathfrak{B}_r$ is a subset of all x such that $\frac{\partial}{\partial x+}\psi(x,r)\neq\frac{\partial}{\partial x-}\psi(x,r)$, we know from Lemma A.1 that $\mathfrak{B}_r$ is at most denumerable. Then, $\mathfrak{Q}$ being denumerable, $\mathfrak{U}=\bigcup_{r\in\mathfrak{Q}}\mathfrak{B}_r$ is also at most denumerable. Analogously we show that the set

$$\mathfrak{U}'=\left\{u\in(-a,a)\colon \frac{\partial}{\partial y+}\psi(u,u)\neq\frac{\partial}{\partial y-}\psi(u,u)\right\}$$

is at most denumerable.

Choose now any $u\notin\mathfrak{U}\cup\mathfrak{U}'$. Since $\psi(x,y)$ is differentiable in both x and y at $x=y=u$, and since, due to the near-smoothness of $\psi(x,y)$, $\frac{\partial}{\partial y}\psi(x,u)$ is continuous in x on $x\in[-a,a]$, we invoke Lemma A.6 to obtain

$$\frac{d}{du}\psi(u,u)=\frac{\partial}{\partial x}\psi(u,u)+\frac{\partial}{\partial y}\psi(u,u).$$

But, in view of (8), the two partials equal zero as they are the derivatives of functions $x\to\psi(x,u)$ and $y\to\psi(u,y)$ taken at their minima. Hence

$$\frac{d}{du}\psi(u,u)=0.$$

By Lemma A.5 $\psi(u,u)$ is continuous, and as its derivative is zero everywhere except on an at most denumerable set, $\psi(u,u)\equiv\text{const}$ by Lemma A.7. This contradicts the assumption that $\psi(x,y)$, being a typical patch, satisfies (10). □

The obvious consequence of this theorem is that if one views regular minimality and nonconstant self-similarity as basic properties of discrimination probability functions, any model predicting that these functions have near-smooth patches should be rejected. On recalling Theorem 8.1 and the subsequent Remark 8.1, we come to the following conclusion (recall the definitions of a typical patch and a typical line element given at the end of Section 9).

Main conclusion. *A typical patch $\psi(x,y)$, satisfying the regular minimality and nonconstant self-similarity conditions* (8) *and* (10), *does not have a well-behaved Thurstonian-type representation. As a result, no discrimination probability function $\psi(\mathbf{x},\mathbf{y})$ with regular minima and nonconstant self-similarity allows for a Thurstonian-type representation well-behaved at any of the typical line elements $(\mathbf{s},\mathbf{u})$.*

To rephrase, assume that a discrimination probability function $\psi(\mathbf{x},\mathbf{y})$ is generated by a well-behaved Thurstonian-type model. Then

(a) if the minima of $\mathbf{x} \to \psi(\mathbf{x}, \mathbf{y})$ or $\mathbf{y} \to \psi(\mathbf{x}, \mathbf{y})$ are generally different for different $\mathbf{x}$ (respectively, $\mathbf{y}$), then $\psi(\mathbf{x}, \mathbf{y})$ cannot satisfy the regular minimality principle: there will have to exist pairs $(\mathbf{x}_0, \mathbf{y}_0)$ such that even though $\mathbf{x}_0 \to \psi(\mathbf{x}_0, \mathbf{y})$ achieves its minimum at $\mathbf{y}_0$, the function $\mathbf{y}_0 \to \psi(\mathbf{x}_0, \mathbf{y})$ does not achieve its minimum at $\mathbf{x}_0$ (or vice versa);

(b) if $\psi(\mathbf{x}, \mathbf{y})$ satisfies the regular minimality principle, then the minima of all $\mathbf{x} \to \psi(\mathbf{x}, \mathbf{y})$ and $\mathbf{y} \to \psi(\mathbf{x}, \mathbf{y})$ will all have to be on one and the same level.

11. "Relativization" of well-behavedness

Definition 7.1 is both intuitive and sufficiently broad to incorporate realistically conceivable Thurstonian-type models. As it turns out, however, its scope can be significantly broadened while preserving the essential logic of the proof of Theorem 8.1.

The well-behavedness in Definition 7.1 can be called "absolute", in the following sense.

1. The well-behavedness of the two measures $A_x(\mathfrak{p})$ and $B_y(\mathfrak{q})$ in Definition 7.1 does not depend on the structure of the decision area $\mathfrak{S}$. If $A_x(\mathfrak{p})$ and $B_y(\mathfrak{q})$ are well-behaved, then they are well-behaved for any AB-measurable area $\mathfrak{S}$.
2. The well-behavedness of either of the two measures $A_x(\mathfrak{p})$ and $B_y(\mathfrak{q})$ in Definition 7.1 does not depend on the other. If $A_x(\mathfrak{p})$ is well-behaved, then it is well-behaved in combination with any probability measure $B_y(\mathfrak{q})$, and vice versa.

Consider the issue of the decision area first. For a well-behaved measure $A_x(\mathfrak{p})$ in Definition 7.1, the right- and left-hand derivatives $D_A^{\pm}(\mathfrak{p}, x)$ exist on any A-measurable set $\mathfrak{p}$, and they are dominated by a constant c across all such sets (constituting Σ_A). In view of the proof of Theorem 8.1, however, this requirement is excessively stringent. A sigma-algebra Σ_A is typically a very large set of subsets of $\mathfrak{R}$, some of which may have complex and peculiar structures. In the proof of Theorem 8.1, however, the existence and boundedness of the derivatives $D_A^{\pm}(\mathfrak{p}, x)$ is only applied to those subsets $\mathfrak{p}$ that are q-sections of the decision area $\mathfrak{S}$ (as defined in (3), Section 4). For most reasonably constructed decision rules these sections have a relatively simple, often very simple, structure.

Thus, assuming for simplicity $\mathfrak{R} = \mathrm{Re}$, the sigma-algebra of Lebesgue-measurable subsets of Re is a huge class whose cardinality equals that of all subsets of Re. It includes subsets of complex structure, those that cannot be presented as unions of denumerably many intervals. At the same time, in a distance-based decision rule,

$$\mathfrak{S} = \{(p, q) \in \mathrm{Re}^2\colon |p - q| > \varepsilon\},$$

all q-sections have the simple structure

$$\mathfrak{p} = (-\infty, q - \varepsilon) \cup (q + \varepsilon, \infty).$$

As a result, if a probability measure $A_x(\mathfrak{p})$ is such that its derivatives $D_A^{\pm}(\mathfrak{p}, x)$ exist and are bounded on intervals of this type, the proof of Theorem 8.1 will be valid, even if outside the class of such intervals the derivatives fail to exist or be bounded. With a category-based decision rule,

$$\mathfrak{S} = \mathrm{Re}^2 - \{(p, q) \in \mathrm{Re}^2\colon a_i \leqslant p, q \leqslant a_{i+1},\ i = 0, \ldots, k\}$$

(with the a_i-points partitioning Re into successive adjacent intervals, $a_0 = -\infty$, $a_{k+1} = \infty$), the situation is even simpler. The q-sections here consist of only $k + 1$ distinct unions $(-\infty, a_i) \cup (a_{i+1}, \infty)$, and the properties of $D_A^{\pm}(\mathfrak{p}, x)$ need not be posited outside this finite class of subsets.

These considerations lead us to the following generalization of Definition 7.1.

Definition 11.1. Let $\Omega_A = \{\mathfrak{a}(q)\}_{q \in \mathfrak{R}}$, $\Omega_B = \{\mathfrak{b}(p)\}_{p \in \mathfrak{R}}$ be the sets of, respectively, q-sections and p-sections of a decision area $\mathfrak{S}$. A probabilistic measure $A_x(\mathfrak{p})$, $\mathfrak{p} \in \Sigma_A$, $x \in [-a, a]$, is *well-behaved with respect to* $\mathfrak{S}$ if the left- and right-hand derivatives

$$D_A^{\pm}(\mathfrak{p}, x) = \frac{\partial A_x(\mathfrak{p})}{\partial x \pm}$$

exist and are bounded on $\Omega_A \times [-a, a]$.

Well-behavedness of $B_y(\mathfrak{q})$, $\mathfrak{q} \in \Sigma_B$, $y \in [-a, a]$, with respect to $\mathfrak{S}$ is defined analogously, with its unilateral derivatives existing and being bounded on $\Omega_B \times [-a, a]$.

A Thurstonian-type representation $\{\mathfrak{R}, A_x, B_y, \mathfrak{S}\}$ for a patch $\psi(x, y)$, $(x, y) \in [-a, a]^2$, is *well-behaved with respect to* $\mathfrak{S}$ if $A_x(\mathfrak{p})$ and $B_y(\mathfrak{q})$ are well-behaved with respect to $\mathfrak{S}$.

Consider now the issue of relating the measures A_x and B_y to each other. Inspection of the proof of Theorem 8.1 shows that the boundedness of, say, $D_B^{\pm}(\mathfrak{q}, y)$ is utilized twice: the first time to conclude that

$$\frac{\partial}{\partial y \pm} \psi(x, y) = \int_{p \in \mathfrak{R}} D_B^{\pm}[\mathfrak{b}(p), y]\, dA_x(p),$$

and the second time to be able to form a sequence of simple functions $\varphi_i(p)$ uniformly converging to $p \to D_B^{\pm}[\mathfrak{b}(p), y]$, at a fixed value of y. The latter step, however, only requires that $p \to D_B^{\pm}(\mathfrak{b}, y)$ be bounded on Ω_A for every given value of $y \in [-a, a]$, which is a weaker requirement than being bounded on $\Omega_A \times [-a, a]$. The validity of the first step, on the other hand, does not require any boundedness at all, provided $D_B^{\pm}[\mathfrak{b}(p), y]$ is dominated by some functions $g(p)$ integrable with respect to measure A_x. These considerations lead us to a further generalization.

Definition 11.2. Given a Thurstonian-type representation $\{\mathfrak{R}, A_x, B_y, \mathfrak{S}\}$ for a patch $\psi(x,y)$, $(x,y)\in[-a,a]^2$, let Ω_A, Ω_B be as in Definition 11.1. The probability measure $A_x(\mathfrak{p})$, $\mathfrak{p}\in\Sigma_A$, $x\in[-a,a]$, is *well-behaved with respect to* $(\mathfrak{S}, B_y)$ if

(i) for all $(\mathfrak{p},x)\in\Omega_A\times[-a,a]$, there exist
$$D_A^{\pm}(\mathfrak{p},x)=\frac{\partial A_x(\mathfrak{p})}{\partial x\pm};$$

(ii) for some function $c(x)\geqslant 0$ and for all $\mathfrak{p}\in\Omega_A$,
$$|D_A^{\pm}(\mathfrak{p},x)|\leqslant c(x);$$

(iii) for some function $g(\mathfrak{p})$, $\mathfrak{p}\in\Omega_A$, and for all $(\mathfrak{p},x)\in\Omega_A\times[-a,a]$
$$|D_A^{\pm}(\mathfrak{p},x)|\leqslant g(\mathfrak{p}),\qquad \int_{q\in\mathfrak{R}} g(\mathfrak{a}(q))\,dB_y(q)<\infty;$$

The well-behavedness of $B_y(\mathfrak{q})$, $q\in\Sigma_B$, $y\in[-a,a]$, with respect to $(\mathfrak{S}, A_x)$ is defined in a symmetrical fashion.

A Thurstonian-type representation $\{\mathfrak{R}, A_x, B_y, \mathfrak{S}\}$ for a patch $\psi(x,y)$, $(x,y)\in[-a,a]^2$, is *well-behaved in the relative* (*or broad*) *sense* if $A_x(\mathfrak{p})$ is well-behaved with respect to $(\mathfrak{S}, B_y)$ and $B_y(\mathfrak{q})$ is well-behaved with respect to $(\mathfrak{S}, A_x)$.

Plainly, absolute well-behavedness implies well-behavedness with respect to a decision area, which in turn implies well-behavedness in the relative sense.

The analysis of Theorem 8.1 makes it clear that the main conclusion of this paper, formulated at the end of Section 10, holds good if the well-behavedness in it is understood in the relative sense.

12. Conclusion

This is what we know about "same-different" discrimination probabilities and their Thurstonian-type representations.

1. Discrimination probability functions $\psi(\mathbf{x},\mathbf{y})$ possess two fundamental properties: regular minimality and nonconstant self-similarity. As argued in Dzhafarov (2002d), regular minimality is a fundamental constraint whose violation seems quite implausible and which is corroborated by available empirical evidence; nonconstant self-similarity is a well-documented empirical fact.
2. If one imposes no constraints on the random variables representing stimuli, admitting, in particular, singular probability measures (concentrated on sets of measure zero), *any* discrimination probability function $\psi(\mathbf{x},\mathbf{y})$ has a Thurstonian-type representation with independent perceptual images and a deterministic decision rule. This applies, of course, to $\psi(\mathbf{x},\mathbf{y})$ with regular minima and nonconstant self-similarity.
3. All existing and, one could claim, realistically conceivable Thurstonian-type models designed to fit empirical data utilize random images whose change in response to changing stimuli exhibits certain regularity features. One such feature, used in this paper to construct the definition of well-behavedness, is that the probabilities with which a random image falls in various areas of the perceptual space have right-hand and left-hand bounded derivatives with respect to (a certain parametric representation of) stimuli.
4. It turns out that $\psi(\mathbf{x},\mathbf{y})$ with regular minima and nonconstant self-similarity cannot be generated by Thurstonian-type representations that are well-behaved in this sense.

One might wonder whether the latter result might be due to the use of stochastically independent images, or deterministic decision rules (or both) rather than due to the well-behavedness of the probability measures as such. This issue is taken up in the companion paper (Dzhafarov, 2003), and it turns out that the introduction of stochastic decisions and interdependent (but selectively attributed to stimuli) random images does not alter the result in question.

It appears that insofar as one wishes to apply Thurstonian-type representations to "same-different" probabilities defined on continuous stimulus spaces, one will have to abandon the time-honored modeling practices of psychophysics (involving normal or other "nice" distributions) in favor of working with singular distributions, or distributions whose densities have prominent singularities (see the end of Section 7). Alternatively, one can abandon the idea of stochastic images representing individual stimuli altogether, assuming instead that every pair of stimuli presented for a comparison is mapped into a single random variable (or process) interpretable as a subjective image of the (*dis*)*similarity* between the two stimuli (as, e.g., in the model by Takane & Sergent, 1983). Finally, one can look for an appropriate rendering of the idea of a *deterministic dissimilarity function*, imposed directly on a stimulus space, such that the dissimilarity of $\mathbf{y}$ from $\mathbf{x}$ is mapped into $\psi(\mathbf{x},\mathbf{y})$ by a fixed monotonic transformation (see Dzhafarov, 2002b). An example demonstrating the latter approach is constructed in the companion paper (Dzhafarov, 2003).

Strictly speaking, all these conclusions and comments apply to the "same-different" comparison paradigm only. As shown in Section 3, well-behaved Thurstonian-type models seem adequate for dealing with "greater-less", preference judgments. In a psychophysical context, however, where one is likely to view random entities $P(\mathbf{x}), Q(\mathbf{y})$ as theoretical descriptors of *perceptual images*, one may be concerned with the fact that for one and the same stimulus space these descriptors are adequate in conjunction with one but not with another judgment scheme.

Of the mathematical problems left open by the development presented in this paper one can note two. The first one relates to Theorem 5.1. The theorem says that every conceivable discrimination probability function $\psi(\mathbf{x},\mathbf{y})$ admits a Thurstonian-type representation, and it proves this statement by using singular distributions in a Euclidean unit cube. It remains unknown whether this statement can also be proved by using nonsingular distributions, whose supports are sets of positive measure (not necessarily in a Euclidean space with Lebesgue measure). At present we only know that the class of Thurstonian-type representations from which one can always choose one for every conceivable $\psi(\mathbf{x},\mathbf{y})$ cannot be confined to well-behaved representations. This leads us to the second problem. The main conclusion of this paper is that well-behaved Thurstonian-type representations cannot account for discrimination probability functions subject to regular minimality and nonconstant self-similarity. Well-behavedness, however (even in the broadest, relative sense defined in Section 11), is only a sufficient condition for this failure. A complete characterization of the class of Thurstonian-type representations that are incompatible with the conjunction of regular minimality and nonconstant self-similarity remains unknown.

Acknowledgments

This research has been supported by the NSF Grant SES-0001925. The author is deeply grateful to Alex Eremenko, Hans Colonius, A.A.J. Marley, and R. Duncan Luce for discussions, criticism, and help. The paper especially benefited from A.A.J. Marley's detailed editing which resulted in several imprecisions corrected and formulations improved.

Appendix A. Auxiliary derivations and facts

Lemma A.1. *If $F(x)$ is both right- and left-differentiable on $[a,b]$, then it is differentiable on $[a,b]$, except on at most a denumerable set of points.*

Proof. See, e.g., Hewitt & Stromberg (1965, p. 262), or Bruckner (1978, p. 63). □

Lemma A.2. *If*

$$\left|\frac{F(x)-F(x')}{x-x'}\right| \leqslant c$$

for $x,x' \in [-a,a]$, then $\frac{d}{dx+}F(x)$ and $\frac{d}{dx-}F(x)$ exist almost everywhere on $(-a,a)$ (and are bounded by c).

Proof. See Bruckner (1978, pp. 53, 65). □

Lemma A.3. *Let*

$$F(y)=\int_{p\in\mathfrak{R}} f(p,y)\,dM(p), \quad y\in[-a,a],$$

where M is a sigma-finite measure. Let $y \to f(p,y)$ be both right- and left-differentiable on $[-a,a]$, with

$$\left|\frac{\partial}{\partial y\pm} f(p,y)\right| \leqslant g(p), \qquad \int_{p\in\mathfrak{R}} g(p)\,dM(p)<\infty.$$

Then $F(y)$ is both right- and left-differentiable (hence continuous) on $[-a,a]$, with

$$\frac{\partial}{\partial y\pm}F(y)=\int_{p\in\mathfrak{R}}\frac{d}{dx\pm}f(p,y)\,dM(p).$$

Remark A.1. This is a standard theorem of abstract analysis, given here only because it is usually formulated for common rather than unilateral derivatives.

Proof. In the case of right-differentiability, for $[y,y+\delta]\subset[-a,a]$,

$$\left|\frac{f(p,y+\delta)-f(p,y)}{\delta}\right| \leqslant \sup_{u\in[-a,a]}\left\{\frac{\partial}{\partial y+}f(p,u)\right\} \leqslant g(p)$$

whence the proof obtains by applying the Lebesgue Dominated Convergence theorem (e.g., Hewitt & Stromberg, 1965, pp. 172–173) to

$$\frac{f(p,y+\delta)-f(p,y)}{\delta} \to \frac{\partial}{\partial x+}f(p,y), \quad \text{as} \quad \delta\to 0+.$$

The proof for left-differentiability is analogous. □

Lemma A.4. *Let*

$$A_x(\mathfrak{p})=\int_{p\in\mathfrak{p}} \alpha(p,x)\,dM(p), \quad \mathfrak{p}\in\Sigma, x\in[-a,a],$$

where M is a sigma-finite measure on Σ. Let $\frac{\partial}{\partial x\pm}\alpha(p,x)$ exist at every $p\in\mathfrak{R}$, and let, for some $g(p)$,

$$\left|\frac{\partial}{\partial x\pm}\alpha(p,x)\right| \leqslant g(p), \qquad \int_{p\in\mathfrak{R}} g(p)\,dM(p)=c<\infty.$$

Then $\frac{\partial}{\partial x\pm}A_x(\mathfrak{p})$ exist and are bounded on $\Sigma\times[-a,a]$.

Proof. Apply Lemma A.3 to obtain

$$\frac{\partial}{\partial x\pm} A_x(\mathfrak{p}) = \int_{p\in\mathfrak{p}} \frac{\partial}{\partial x\pm}\alpha(p,x)\,dM(p),$$

and observe that

$$\begin{aligned}\int_{p\in\mathfrak{p}} \frac{\partial}{\partial x\pm}\alpha(p,x)\,dM(p) &\leqslant \int_{p\in\mathfrak{p}} \left|\frac{\partial}{\partial x\pm}\alpha(p,x)\right| dM(p)\\ &\leqslant \int_{p\in\mathfrak{R}} \left|\frac{\partial}{\partial x\pm}\alpha(p,x)\right| dM(p)\\ &\leqslant \int_{p\in\mathfrak{R}} g(p)\,dM(p) = c < \infty. \qquad \square\end{aligned}$$

Lemma A.5. *If $\psi(x,y)$ is near-smooth on $[-a,a]^2$, then $x \to \psi(x,x)$ is continuous on $[-a,a]$.*

Proof. For $\delta \to 0+$,

$$\begin{aligned}\psi(u+\delta,u+\delta) - \psi(u,u) &= [\psi(u+\delta,u+\delta) - \psi(u+\delta,u)]\\ &\quad + [\psi(u+\delta,u) - \psi(u,u)]\\ &= \frac{\partial}{\partial y+}\psi(u+\delta,u)\delta + \frac{\partial}{\partial x+}\psi(u,u)\delta + o\{\delta\} \to 0,\end{aligned}$$

because $\frac{\partial}{\partial y+}\psi(u+\delta,u) \to \frac{\partial}{\partial y+}\psi(u,u)$. Analogously for $\delta \to 0-$. $\square$

Lemma A.6. *If $\psi(x,y)$ is differentiable in both x and y at some $x = y = u$, and if $\frac{\partial}{\partial y}\psi(x,u)$ is continuous in x, then*

$$\frac{d}{du}\psi(u,u) = \frac{\partial}{\partial x}\psi(u,u) + \frac{\partial}{\partial y}\psi(u,u).$$

Remark A.2. This is essentially a theorem of elementary calculus, except that it is usually given under the assumption of continuous differentiability in (x,y).

Proof. For $\delta \to 0$,

$$\begin{aligned}\frac{\psi(u+\delta,u+\delta) - \psi(u,u)}{\delta} &= \frac{\psi(u+\delta,u+\delta) - \psi(u+\delta,u)}{\delta}\\ &\quad + \frac{\psi(u+\delta,u) - \psi(u,u)}{\delta}\\ &= \frac{\partial}{\partial y}\psi(u+\delta,u)\\ &\quad + \frac{\partial}{\partial x}\psi(u,u) + o\{1\},\end{aligned}$$

and $\frac{\partial}{\partial y}\psi(u+\delta,u) \to \frac{\partial}{\partial y}\psi(u,u)$. $\square$

Lemma A.7. *If $F(x)$ is continuous and $\frac{d}{dx}F(x) = 0$ everywhere except on an at most denumerable set of points, then $F(x) \equiv \text{const}$.*

Proof. See, e.g., Saks (1937, p. 275), or, for a more general treatment, Bruckner (1978, p. 203). $\square$

References

Bruckner, A. M. (1978). *Differentiation of real functions.* Berlin: Springer-Verlag.

Dai, H., Versfeld, N. J., & Green, D. (1996). The optimum decision rules in the same-different paradigm. *Perception and Psychophysics*, *58*, 1–9.

Dzhafarov, E. N. (1999). Conditionally selective dependence of random variables on external factors. *Journal of Mathematical Psychology*, *43*, 123–157.

Dzhafarov, E. N. (2001a). Unconditionally selective dependence of random variables on external factors. *Journal of Mathematical Psychology*, *45*, 421–451.

Dzhafarov, E. N. (2001b). Fechnerian scaling and Thurstonian modeling. In: E. Sommerfeld, R. Kompass, & T. Lachmann (Eds.), *Fechner Day 2001* (pp. 42–47). Lengerich: Pabst Science Publishers.

Dzhafarov, E. N. (2002a). Multidimensional Fechnerian scaling: Regular variation version. *Journal of Mathematical Psychology*, *46*, 226–244.

Dzhafarov, E. N. (2002b). Multidimensional Fechnerian scaling: Probability-distance hypothesis. *Journal of Mathematical Psychology*, *46*, 352–374.

Dzhafarov, E. N. (2002c). Multidimensional Fechnerian scaling: Perceptual separability. *Journal of Mathematical Psychology*, *46*, 564–582.

Dzhafarov, E. N. (2002d). Multidimensional Fechnerian scaling: Pairwise comparisons, regular minimality, and nonconstant self-similarity. *Journal of Mathematical Psychology*, *46*, 583–608.

Dzhafarov, E. N. (in press). Selective influence through conditional independence. *Psychometrika*.

Dzhafarov, E. N. (2003). Thurstonian-type representations for "same-different" discriminations: Probabilistic decisions and interdependent images. *Journal of Mathematical Psychology*, *47*, 229–243.

Dzhafarov, E. N., & Colonius, H. (1999). Fechnerian metrics in unidimensional and multidimensional stimulus spaces. *Psychonomic Bulletin and Review*, *6*, 239–268.

Dzhafarov, E. N., & Colonius, H. (2001). Multidimensional Fechnerian scaling: Basics. *Journal of Mathematical Psychology*, *45*, 670–719.

Ennis, D. M. (1992). Modeling similarity and identification when there momentary fluctuations in psychological magnitudes. In: F.G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 279–298). Hillsdale, NJ: Erlbaum.

Ennis, D. M., Palen, J. J., & Mullen, K. (1988). A multidimensional stochastic theory of similarity. *Journal of Mathematical Psychology*, *32*, 449–465.

Hewitt, E., & Stromberg, K. (1965). *Real and abstract analysis*. New York: Springer.

Indow, T. (1998). Parallel shift of judgment-characteristic curves according to the context in cutaneous and color discrimination. In: C.E. Dowling, F.S. Roberts, & P. Theuns (Eds.), *Recent progress in mathematical psychology* (pp. 47–63). Mahwah, NJ: Erlbaum.

Indow, T., Robertson, A. R., von Grunau, M., & Fielder, G. H. (1992). Discrimination ellipsoids of aperture and simulated surface colors by matching and paired comparison. *Color Research and Applications*, *17*, 6–23.

Krumhansl, C. L. (1978). Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density. *Psychological Review*, *85*, 445–463.

Luce, R. D., & Galanter, E. (1963). Discrimination. In: R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*, Vol. 1 (pp. 103–189). New York: Wiley.

Luce, R. D., & Suppes, P. (1965). Preference, utility, and subjective probability. In: R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*, Vol. 3 (pp. 249–410). New York: Wiley.

Niederée, R., & Heyer, D. (1997). Generalized random utility models and the representational theory of measurement: a conceptual link. In A.A.J. Marley (Ed.), *Choice, decision and measurement: Essays in honor of R. Duncan Luce* (pp. 155–189). Mahwah, NJ: Erlbaum.

Regenwetter, M., & Marley, A. A. J. (2001). Random relations, random utilities, and random functions. *Journal of Mathematical Psychology*, *45*, 864–912.

Robertson, C. A., & Strauss, D. J. (1981). A characterization theorem for random utility variables. *Journal of Mathematical Psychology*, *23*, 184–189.

Rothkopf, E. Z. (1957). A measure of stimulus similarity and errors in some paired-associate learning tasks. *Journal of Experimental Psychology*, *53*, 94–101.

Saks, S. (1937). *Theory of the integral*. New York: Stechert.

Sorkin, R. D. (1962). Extension of the theory of signal detectability to matching paradigms in psychoacoustics. *Journal of the Acoustical Society of America*, *34*, 1745–1751.

Suppes, P., & Zinnes, J.L. (1963). Basic measurement theory. In: R.D. Luce, R.R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*, Vol. 1 (pp. 3–76). New York: Wiley.

Strauss, D. J. (1979). Some results on random utility models. *Journal of Mathematical Psychology*, *20*, 35–52.

Takane, Y., & Sergent, J. (1983). Multidimensional scaling models for reaction times and same-different judgments. *Psychometrika*, *48*, 393–423.

Thomas, R. D. (1996). Separability and independence of dimensions within the same-different judgment task. *Journal of Mathematical Psychology*, *40*, 318–341.

Thomas, R. D. (1999). Assessing sensitivity in a multidimensional space: Some problems and a definition of a general d'. *Psychonomic Bulletin and Review*, *6*, 224–238.

Thurstone, L. L. (1927a). Psychophysical analysis. *American Journal of Psychology*, *38*, 368–389.

Thurstone, L. L. (1927b). A law of comparative judgments. *Psychological Review*, *34*, 273–286.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327–352.

Yellott, J. I. (1997). Preference models and irreversibility. In: A.A.J. Marley (Ed.), *Choice, decision and measurement: Essays in honor of R. Duncan Luce* (pp. 131–151). Mahwah, NJ: Erlbaum.

Zimmer, K., & Colonius, H. (2000). Testing a new theory of Fechnerian scaling: the case of auditory intensity discrimination. Paper presented at the 140th Meeting of the Acoustical Society of America, Newport Beach, CA, USA.

Zinnes, J. L., & MacKey, D. B. (1983). Probabilistic multidimensional scaling: complete and incomplete data. *Psychometrika*, *48*, 27–48.