Reconstructing Distances Among Objects from Their Discriminability

Ehtibar N. Dzhafarov[1]

Purdue University and

Swedish Collegium for Advanced Studies in Social Sciences

Hans Colonius

Oldenburg University, Germany

[1]Address correspondence to Ehtibar N. Dzhafarov, Department of Psychological Sciences, Purdue University, 703 Third Street, West Lafayette, IN 47907-2081 (e-mail:ehtibar@purdue.edu) (ph: 765/4942376).
This research was supported by the NSF grant SES 0318010 (E.D.), Humboldt Research Award (E.D.), Humboldt Foundation grant DEU/1038348 (H.C.& E.D.), and DFG grant Co 94/5 (H.C.).

Abstract

We describe a principled way of imposing a metric representing dissimilarities on any discrete set of stimuli (symbols, handwritings, consumer products, X-ray films, etc.), given the probabilities with which they are discriminated from each other by a perceiving system, such as an organism, person, group of experts, neuronal structure, technical device, or even an abstract computational algorithm. In this procedure one does not have to assume that discrimination probabilities are monotonically related to distances, or that the distances belong to a predefined class of metrics, such as Minkowski. Discrimination probabilities do not have to be symmetric, the probability of discriminating an object from itself need not be a constant, and discrimination probabilities are allowed to be 0's and 1's. The only requirement that has to be satisfied is *Regular Minimality*, a principle we consider the defining property of discrimination: for ordered stimulus pairs (a, b), b is least frequently discriminated from a if and only if a is least frequently discriminated from b. Regular Minimality generalizes one of the weak consequences of the assumption that discrimination probabilities are monotonically related to distances: the probability of discriminating a from a should be less than that of discriminating a from any other object. This special form of Regular Minimality also underlies such traditional analyses of discrimination probabilities as Multidimensional Scaling and Cluster Analysis.

KEYWORDS:. continuous stimulus space, discrete stimulus space, discrimination, Fechnerian Scaling of Discrete Object Sets (FSDOS), Multidimensional Fechnerian Scaling (MDFS), Non-constant Self-Dissimilarity, Regular Minimality, psychometric function, same-different judgments, subjective distance.

1. Introduction

1. Example. We begin with a toy example, to be used throughout to illustrate various points. Let there be a set of four distinct objects, $\{A, B, C, D\}$ (say, four pictures or symbols), and let them be presented pairwise side-by-side (AA, AB, BB, CD, etc.). A "perceiver" (say, a child) has to answer the question "Are these two objects different or the same?" Each of the 16 pairs in this experiment is presented R times, R being large enough to form reliable estimates of the probabilities $\psi(x, y)$ with which objects x and y are judged to be different, $x, y \in \{A, B, C, D\}$. The results of such an experiment may look like in Table 1, where the rows represent, say, the objects presented on the left, and the columns the objects presented on the right.

Table 1: A toy example: discrimination probabilities in a four-element set.

$M1$	A	B	C	D
A	0.1	0.8	0.6	0.6
B	0.8	0.1	0.9	0.9
C	1	0.6	0.5	1
D	1	1	0.7	0.5

This experimental paradigm may have numerous variants (different randomization schemes, presence or absence of feedback, successive rather than simultaneous presentations of object pairs, etc.), but the essential feature is that at the end we get a matrix whose entries $\psi(x, y)$ can be interpreted as *probabilities with which a given perceiver judges the row objects, x, to be different from the column objects, y.*

The purpose of this paper is to describe a computational procedure, *Fechnerian Scaling of Discrete Object Sets (FSDOS)*, which when applied to such matrices produces a matrix of distances we call *Fechnerian.* Intuitively, they reflect the degree of subjective dissimilarity among the objects, "from the point of view" of the perceiver. In addition, FSDOS produces the set of what we call *geodesic loops*, the shortest (in some well-defined sense) chains of objects leading from one given object to another and back. Thus, when applied to our matrix $M1$, FSDOS yields the two matrices shown in Table 2. For instance, the geodesic loop connecting A and B is $A \to C \to B \to A$, whereas the geodesic loop connecting A and C is $A \to C \to A$. The lengths of these loops (whose computation is explained later) is taken to be the Fechnerian distances between A and B and between A and C, respectively. We see in $G1$ that the Fechnerian distance between A and B is 1.3 times the Fechnerian distance between A and C.

Table 2: Geodesic loops (matrix $L1$) and Fechnerian distances (matrix $G1$) computed from matrix $M1$ of Table 1.

$L1$	A	B	C	D
A	A	$ACBA$	ACA	ADA
B	$BACB$	B	BCB	$BDCB$
C	CAC	CBC	C	CDC
D	DAD	$DCBD$	DCD	D

$G1$	A	B	C	D
A	0	1.3	1	1
B	1.3	0	0.9	1.1
C	1	0.9	0	0.7
D	1	1.1	0.7	0

Our choice of $M1$ in Table 1 illustrates the fact that FSDOS does not presuppose that $\psi(x,x)$ is the same for all x (*constant self-dissimilarity*), or that $\psi(x,y) = \psi(y,x)$ (*symmetry*). Also, FSDOS allows some or even all the probabilities to equal 1's and 0's. The procedure, however, is based on the assumption that discrimination probabilities satisfy what we call the *Regular Minimality requirement*: in our example this means that $\psi(x,x)$ is always less than both $\psi(x,y)$ and $\psi(y,x)$, for any $y \neq x$ (as explained in Section 3.6, the general formulation of Regular Minimality is weaker).

2. Experimental paradigm. In general, the experimental paradigm we deal with involves a set of objects $\{s_1, s_2, ..., s_N\}$, $N > 1$, presented two at a time to a perceiver whose task is to respond to each ordered pair (x, y) by one of two answers, interpretable as "x and y are the same" and "x and y are different." As a result, each ordered pair (x, y) is assigned (an estimate of) the probability

$$\psi(x, y) = \Pr[\text{the perceiver judges } x \text{ and } y \text{ in } (x, y) \text{ to be different objects}], \tag{1}$$

where $x, y \in \{s_1, s_2, ..., s_N\}$. The "perceiver" is a technical term whose meaning can vary. In psychophysics it usually means a biological organism or a person to whom each pair is presented repeatedly (Dzhafarov & Colonius, 2005a; Indow, 1998; Indow, Robertson, von Grunau, & Fielder, 1992; Zimmer & Colonius, 2000). In other applications the "perceiver" can be a group of people whose individual responses to a given pair of objects are treated as replications of this pair (Rothkopf, 1957; Wish, 1967). With some additional assumptions, the term can also designate a neuronal system reacting differently when a stimulus changes and when it does not change (Izmailov, Dzhafarov, & Zimachev, 2001).

Discrimination probabilities $\psi(x, y)$ occupy a special place among available measures of pairwise dissimilarity. The ability of telling two objects apart or identifying them as being the same (in some respect or overall) seems to be the most basic cognitive ability in biological perceivers and the most basic requirement of intelligent technical systems. For our purposes it is convenient to use the term "perceiver" in the maximally broad meaning, including even the cases when objects $\{s_1, s_2, ..., s_N\}$ are purely conceptual entities and the "perceiver" designates a computational procedure whose inputs are ordered pairs (s_i, s_j) and whose outputs are interpretable as responses "same" and "different." The Fechnerian distances among $\{s_1, s_2, ..., s_N\}$ then are pairwise dissimilarities "from the point of view" of this computational procedure. To give an example of such a situation, let $\{A, B, C, D\}$ in Table 1 designate four statistical models whose parameters are completely specified by fitting them to a given data set s. Let there be a certain statistical criterion, q, that allows one to reject or retain any of the models A, B, C, D when applied to any data set s' having the same format as s. Then the entries $\psi(x, y)$ of matrix $M1$ could represent the probabilities with which model y is rejected (by criterion q) when applied to a data set s' generated by model x. The Fechnerian distances in matrix *G1* of Table2 then can be interpreted as dissimilarities among the four models "from the point of view" of the procedure specified by data set s and criterion q. Similar examples can be constructed for a variety of other applications, such as similarities among molecules, bird songs, and other strings of elements as discussed in Sankoff and Kruskal (1999).

The precise meaning of the response categories "same" and "different" also may vary depending on the context: "x is the same as y" may mean that x and y appear physically identical (i.e., it is the same object presented in two different locations or at two different times), or it may mean that they appear to belong to the same category or have the same source. In the latter case it is the categories or sources that are viewed as objects $\{s_1, s_2, ..., s_N\}$, whereas the "replications" of a pair (s_i, s_j) are pairs of examples or instances of these categories or sources. Thus, in matrix $M1$ of Table 1 the objects A, B, C, D might designate four lung dysfunctions, each represented by a set of X-ray films. The fact that $\psi(A, B) = 0.8$ in this case means that randomly chosen examples of A paired with randomly chosen examples of B are judged (by a physician) to be representing different dysfunctions in 80% of cases.

3. To prevent confusion. The paradigm just described (pairwise presentations, same-different judgments) must not be confused with two other experimental paradigms that produce matrices superficially similar to Table 1.

One of these paradigms is pairwise presentations with *greater-less judgments*. This paradigm underlies the classical procedure of Thurstonian scaling (Thurstone, 1927): the perceiver is given a semantically unidimensional property (such as pleasantness, usefulness, loudness, etc.), and in response to every pair (x, y) chosen from $\{s_1, s_2, ..., s_N\}$ the perceiver determines which of the two stimuli is "greater" (has more of this property). In the psychophysical literature the probabilities $\xi(x, y)$ with which the second object, y, is judged to be greater than the first one, x, are often referred to as discrimination probabilities, the same term as we use for $\psi(x, y)$ in (1). Note that the determination of the difference or sameness of two objects may but need not involve any designated properties, unidimensional or otherwise. For a detailed comparison of the same-different and greater-less judgments see Dzhafarov (2002d, 2003a).

The other paradigm is that of *identification*: stimuli from a set $\{s_1, s_2, ..., s_N\}$ are presented *one at a time*, and the perceiver's task is to identify the presentation by a normatively preassigned "stimulus' name". The results of such an experiment can be presented in the form of a stimulus-response *confusion* matrix, with rows representing objects and columns object names. The entries of this matrix $\eta(x, y)$ are conditional probabilities of the perceiver replying to x by the name normatively assigned to y. Clearly, $\sum_{j=1}^{N} \eta(s_i, s_j) = 1$. In contrast, in the same-different paradigm the discrimination probabilities $\psi(s_i, s_j)$ can, logically speaking, attain any set of $N \times N$ values (e.g., all of them can be equal to 1). The Regular Minimality constraint mentioned earlier is an empirical assumption, rather than a mathematical necessity.

With some *additional assumptions*, the FSDOS procedure described in this paper can in fact be applied to stimulus-response confusion matrices (as outlined in the concluding section) and matrices of probabilities for greater-less judgments (as described in Dzhafarov & Colonius, 1999;

Dzhafarov, 2002b). These applications, however, are not focal for this paper.

4. Regular Minimality, nonconstant self-dissimilarity, and asymmetry. Classical Multidimensional Scaling (*MDS*, Borg & Groenen, 1997; Kruskal & Wish, 1978) when applied to discrimination probabilities serves as a convenient reference against which to consider FSDOS. MDS is based on the assumption that for some metric $d(x,y)$ (distance function) and some increasing transformation f,

$$\psi(x,y) = f(d(x,y)). \tag{2}$$

This is a prominent instance of what is called the *probability-distance hypothesis* in Dzhafarov (2002b). To remind, the defining properties of a metric d are: (A) $d(a,b) \geq 0$; (B) $d(a,b) = 0$ if and only if $a = b$; (C) $d(a,c) \leq d(a,b) + d(b,c)$; (D) $d(a,b) = d(b,a)$. In addition one assumes in MDS that metric d belongs to a predefined class, usually the class of power-function Minkowski metrics with exponents between 1 and 2. It immediately follows from (A), (B), (D), and the monotonicity of f that for any distinct x and y, $\psi(x,y) = \psi(y,x)$ (Symmetry), $\psi(x,x) = \psi(y,y)$ (Constant Self-Dissimilarity), and $\psi(x,x)$ is less than both $\psi(x,y)$ and $\psi(y,x)$ (Regular Minimality). The problem for MDS is that the properties of symmetry and, more important, constant self-dissimilarity are systematically violated in experimental data. For continuous stimulus spaces (colors, line segments, two-dot apparent motions, pure tones) this has been demonstrated in experiments reported in Dzhafarov & Colonius (2005a), Indow (1998), Indow et al. (1992), and Zimmer & Colonius (2000). For discrete object spaces an example is provided in Table 3 (for now refer to the parenthesized numbers only) representing Rothkopf's (1957) study of discrimination probabilities among 36 Morse codes. As one can see, the Morse code for digit 6 was judged different from itself by 15% of respondents, but only by 6% for digit 9. Digits 4 and 5 were discriminated from each other in 83% of cases when 5 was presented first in the two-code sequence, but in only 58% when 5 was presented second.

Table 3: A 10×10 excerpt from Rothkopf's (1957) 36×36 Morse code data. Shown are Fechnerian distances (first number in each cell), percentages of "different" judgments for row→column Morse code pair sequences (in parentheses), and corresponding closed-loop geodesics (bottom strings). The 10-code subset is chosen so that it forms a self-contained subspace of the 36 codes: a geodesic loop for any two of its elements is contained within the subset.

	B	0	1	3	4	5	6	7	8	9
B	0 (16) B	151 (88) B0B	142 (83) B1B	95 (60) B35B	97 (68) B4B	16(26) B5B	57 (57) B565B	77 (83) B5675B	140 (96) B567875B	157 (96) B975B
0	151 (95) 0B0	0 (16) 0	48 (37) 010	160 (92) 030	150 (90) 040	147 (92) 050	127 (81) 0670	99 (68) 070	61 (43) 080	73 (45) 090
1	142 (86) 1B1	48 (38) 101	0 (11) 1	132 (80) 131	164 (95) 141	147 (86) 151	125 (80) 161	128 (79) 171	106 (84) 1081	121 (89) 10901
3	95 (81) 35B3	160 (95) 303	132 (74) 313	0 (11) 3	68 (58) 343	95 (56) 35B3	127 (68) 363	145 (90) 3673	165 (97) 383	169 (97) 393
4	97 (55) 4B4	150 (86) 404	164 (90) 414	68 (31) 434	0 (10) 4	106 (58) 45B4	138 (76) 4564	160 (90) 474	171 (84) 484	174 (95) 494
5	16 (20) 5B5	147 (85) 505	147 (86) 515	95 (76) 5B35	106 (83) 5B45	0 (14) 5	41 (31) 565	61 (86) 5675	124 (95) 567875	143 (86) 5975
6	57(67) 65B56	127 (78) 6706	125 (71) 616	127 (85) 636	138 (88) 6456	41 (39) 656	0 (15) 6	44 (30) 676	92 (80) 6786	118 (87) 678986
7	77 (77) 75B567	99 (58) 707	128 (71) 717	145 (84) 7367	160 (91) 747	61 (40) 7567	44 (40) 767	0 (11) 7	63 (39) 787	83 (74) 7897
8	140 (86) 875B5678	61 (43) 808	106 (61) 8108	165 (88) 838	171 (96) 848	124 (89) 875678	92 (58) 8678	63 (44) 878	0 (9) 8	26 (22) 898
9	157 (97) 975B9	73 (50) 909	121 (74) 90109	169 (89) 939	174 (95) 949	143 (78) 9759	118 (83) 986789	83 (48) 9789	26 (19) 989	0 (6) 9

At the same time we see that every diagonal probability in this table is less than the off-diagonal probabilities in its row and in its column. This means that Regular Minimality is satisfied, and this is the only property required by FSDOS (generally, in a weakened form). For continuous stimulus spaces Regular Minimality holds (though, as discussed later, not necessarily in this simplest form) in all data sets mentioned earlier. In the context of continuous stimulus spaces the combination of Regular Minimality with nonconstant self-dissimilarity has been shown (Dzhafarov, 2002d) to impose stringent constraints on the possible shapes of functions $\psi(x, y)$, some of which have been experimentally corroborated (Dzhafarov & Colonius, 2005a). The same two properties have also been shown (Dzhafarov, 2003a, b) to have surprisingly strong consequences for *modeling* of discrimination probabilities. They rule out, in particular, the possibility of modeling $\psi(x, y)$ in a continuous stimulus space by means of "well-behaved" random representations of x and y in some perceptual space (e.g., multivariate normal distributions with parameters smoothly depending on stimuli, in combination with any decision rule).

It appears that prior to Dzhafarov (2002d) Regular Minimality for discrimination probabilities has not been formulated as a basic property of discrimination, independent of its other properties, such as constant self-dissimilarity. The violations of symmetry and constant self-dissimilarity, however, have long since been noted. Tversky's (1977) contrast model and Krumhansl's (1978) distance-and-density scheme are two best known theoretical schemes dealing with these issues. Some non-classical versions of MDS are based on these models (e.g., DeSarbo et al., 1992; Weeks & Bentler, 1982). We do not review these approaches here, as a detailed comparison of Tversky's and Krumhansl's ideas with those of Fechnerian Scaling is beyond the scope of this paper. We note only that the "uncertainty blobs" model proposed in Dzhafarov (2003b) leads to a mathematical expression which is similar to that of Krumhansl's (1978) main formula.

2. Background

FSDOS is an outgrowth of the general theory of Fechnerian Scaling originally proposed for continuous stimulus spaces in a primarily psychophysical context (Dzhafarov 2002a, b, c, d; 2003a, b; Dzhafarov & Colonius, 1999, 2001, 2005a, b). The historical reasons for associating this theory with G. T. Fechner (1801-1887) are given in Dzhafarov & Colonius (1999). We begin with a brief simplified account of the Fechnerian theory for a special class of continuous stimulus spaces, from which its extension to discrete object sets will follow in a natural way.

1. Basics: Two observation areas and Regular Minimality. *Stimulus space* (or *object space*) is a set S of all objects of a particular kind (say, all audible simple tones, or all letters of an alphabet) endowed with a discrimination probability function $\psi(x, y)$, $x, y \in S$. The reason we can distinguish (x, y) from (y, x) and treat (x, x) as a pair rather than a single object is that in pairwise presentations the two stimuli generally belong to two distinct *observation areas*. In psychophysical applications this usually refers to spatial arrangement (say, one stimulus is on the left, the other on the right,) or temporal order (first-second). The perception of a stimulus may depend on which of the two observation areas it belongs to. In the case of conceptual objects and "paper-and-pencil perceivers" the term observation area refers to the asymmetries in the computational procedure. Thus, in our example with statistical models, a data set can be generated by model x and fitted by model y, or vice versa.

The most fundamental property of discrimination probabilities is Regular Minimality, which we present for now in its simplest (so-called *canonical*) form: for any $x \neq y$,

$$\psi(x, x) < \min\{\psi(x, y), \psi(y, x)\}. \tag{3}$$

It should be noted from the outset that the logic of Fechnerian Scaling is very different from that of MDS in the following respect: Fechnerian distances are computed *within* rather than *across* the two observation areas. The Fechnerian distance between a and b does not mean a distance between a presented first (or on the left) and b presented second (on the right). Rather, we should logically distinguish $G^{(1)}(a,b)$, the distance between a and b in the first observation area, from $G^{(2)}(a,b)$, the distance between a and b in the second observation area. This must not come as a surprise: a and b in the first observation area are generally perceived differently from a and b in the second observations area. As it turns out, however, if Regular Minimality is satisfied in the canonical form, (3), then it follows from the general theory that $G^{(1)}(a,b) = G^{(2)}(a,b)$ (details below, Sections 2.3, 3.4).

2. Oriented Fechnerian distances in continuous spaces. MDFS (*Multidimensional Fechnerian Scaling*) is Fechnerian Scaling on a stimulus set that can be represented by an open connected region E of n-dimensional ($n \geq 1$) real-valued vectors, such that $\psi(x,y)$ is continuous with respect to its Euclidean topology. This means that $(x_k, y_k) \to (x,y)$ implies $\psi(x_k, y_k) \to \psi(x,y)$. Fechnerian Scaling has been developed for continuous spaces of a much more general structure (Dzhafarov & Colonius, 2005a), but a brief overview of MDFS should suffice in providing motivation for FSDOS.

Refer to Fig. 1. Any points $a, b \in E$ can be connected by a smooth arc $x(t)$, a piecewise continuously differentiable mapping of an interval $[\alpha, \beta]$ of reals into E, with $x(\alpha) = a$, $x(\beta) = b$. The main intuitive idea underlying Fechnerian Scaling is that (A) any point $x(t)$, $t \in [\alpha, \beta)$, can be assigned a local measure of its difference from its "immediate neighbors," $x(t+dt)$; (B) by integrating this local difference from α to β one can obtain the "psychometric length" of the arc $x(t)$; and (C) by taking the infimum of psychometric lengths across all possible smooth arcs connecting a to b one obtains the distance from a to b in space E.

As argued in Dzhafarov and Colonius (1999), this intuitive scheme can be viewed as the essence of Fechner's original theory for unidimensional stimulus continua (Fechner, 1860). The

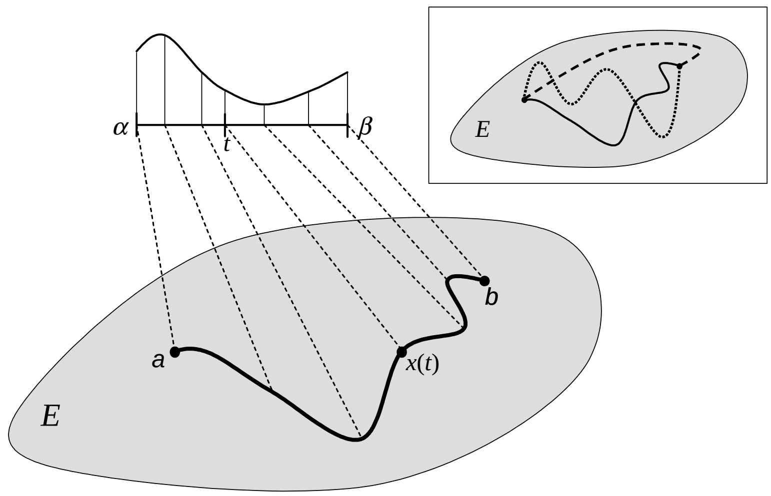

Figure 1: The underlying idea of MDFS. $[\alpha, \beta]$ is a real interval, $a \to x(t) \to b$ a smooth arc. The psychometric length of this arc is the integral of "local difference" of $x(t)$ from $x(t+dt)$, shown by vertical spikes along $[\alpha, \beta]$. The inset shows that one should compute the psychometric lengths for all possible smooth arcs leading from a to b. Their infimum is the oriented Fechnerian distance from a to b.

implementation of this idea in MDFS is as follows (see Fig. 2). As t for a smooth arc $x(t) : [\alpha, \beta] \to E$ moves from α to β, the value of self-discriminability $\psi(x(t), x(t))$ may vary (nonconstant self-dissimilarity). Therefore, to see how distinct $x(t)$ is from $x(t+dt)$ it would not suffice to look at $\psi(x(t), x(t+dt))$ or $\psi(x(t+dt), x(t))$; one should compute instead the increments in discriminability $\psi(x(t), x(t+dt)) - \psi(x(t), x(t))$ and $\psi(x(t+dt), x(t)) - \psi(x(t), x(t))$. These increments, denoted $\phi^{(1)}(x(t), x(t+dt))$ and $\phi^{(2)}(x(t), x(t+dt))$, respectively, are positive due to the Regular Minimality property. They are referred to as *psychometric differentials* of the *first kind* (or in the first observation area) and *second kind* (in the second observation area), respectively.

The assumptions of MDFS guarantee that the cumulation of $\phi^{(1)}(x(t), x(t+dt))$ from $t = \alpha$ to $t = \beta$ always yields a positive quantity. We call this quantity the *psychometric length* of arc

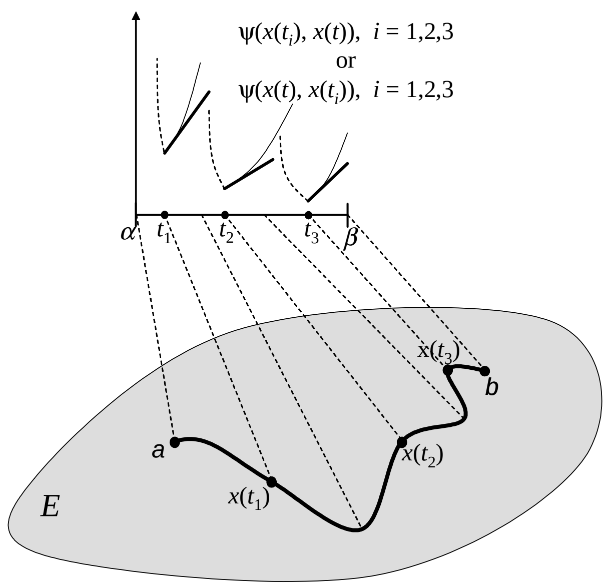

Figure 2: The "local difference" of $x(t)$ from $x(t+dt)$ (as $dt \to 0+$) at a given point, $t = t_i$, is the slope of the tangent line drawn to $\psi(x(t_i), x(t))$ or to $\psi(x(t), x(t_i))$ at $t = t_i+$. Using $\psi(x(t_i), x(t))$ yields derivatives of the first kind, using $\psi(x(t), x(t_i))$ yields derivatives of the second kind. Their integration from α to β yields oriented Fechnerian distances of, respectively, first and second kind (from a to b).

$x(t)$ of the first kind, and denote it $L^{(1)}[a \to x \to b]$ (where we use the suggestive notation for arc x connecting a to b). The infimum $G_1(a,b)$ of the psychometric lengths $L^{(1)}[a \to x \to b]$ across all possible smooth arcs connecting a to b satisfies all properties of a distance except for symmetry: (A) $G_1(a,b) \geq 0$; (B) $G_1(a,b) = 0$ if and only if $a = b$; (C) $G_1(a,c) \leq G_1(a,b) + G_1(b,c)$; but it is not necessarily true that $G_1(a,b) = G_1(b,a)$. We call $G_1(a,b)$ the *oriented Fechnerian distance of the first kind* from a to b. By repeating the whole construction with $\phi^{(2)}(x(t), x(t+dt))$ in place of $\phi^{(1)}(x(t), x(t+dt))$ we will get the psychometric lengths $L^{(2)}[a \to x \to b]$ of the second kind and, as their infima, the oriented Fechnerian distances $G_2(a,b)$ of the second kind (from a to b).

The following observation provides additional justification for computing the oriented Fechnerian distances in the way just outlined. A metric d (symmetrical or oriented) on some set S is called *intrinsic* if $d(a,b)$ for any $a,b \in S$ equals the infimum of the lengths of all "allowable" (in our case, smooth) arcs connecting a and b. The oriented Fechnerian distances $G_1(a,b)$ and $G_2(a,b)$ are intrinsic in this sense. In reference to the classical MDS, all Minkowski metrics are (symmetrical) intrinsic metrics. Assume now that the discrimination probabilities $\psi(x,y)$ on E can be obtained from some symmetrical intrinsic distance d on E by means of (2), with f being a continuous increasing function. It is sufficient to assume that (2) holds for small values of d only. Then, as proved in Dzhafarov (2002b), $d \equiv G_1 \equiv G_2$: $\psi(x,y)$ cannot monotonically and continuously depend on any (symmetrical) intrinsic metric other than the Fechnerian one. The latter in this case is symmetrical, and its two kinds G_1 and G_2 coincide.[1] The classical MDS and its modification proposed in Shepard and Carroll (1966), Tenenbaum, de Silva, & Langford

[1] Fechnerian distances are unique up to multiplication by a positive constant. Equation $d \equiv G_1 \equiv G_2$ therefore could more generally be written as $d \equiv kG_1 \equiv kG_2$, $k > 0$. Throughout this paper we ignore the trivial distinction between different multiples of Fechnerian metrics.

(2000), and Roweis & Saul (2000) fall within this category of models. For continuous spaces, therefore, MDS and MDFS are not simply compatible, the former in fact implies the latter (under the assumption of intrinsicality but without confining the class of metrics d to Minkowski ones). Fechnerian computations, however, are also applicable when the probability-distance hypothesis is false (as we know it generally to be).

3. Overall Fechnerian distances in continuous spaces. The asymmetry of the oriented Fechnerian distances lacks operational meaning. It is easy to understand why $\psi(x,y) \neq \psi(y,x)$: stimulus x in the two cases belongs to two different observation areas and can therefore be perceived differently (the same being true for y). In $G_1(a,b)$, however, a and b belong to the same (first) observation area, and the non-coincidence of $G_1(a,b)$ and $G_1(b,a)$ prevents one from interpreting either of them as a reasonable measure of perceptual dissimilarity between a and b (in the first observation area, "from the point of view" of a given perceiver). The same consideration applies, of course, to G_2. In MDFS this difficulty is resolved by taking as a measure of perceptual dissimilarity the *overall Fechnerian distances* $G_1(a,b) + G_1(b,a)$ and $G_2(a,b) + G_2(b,a)$. What justifies this particular choice of symmetrization is the remarkable fact that

$$G_1(a,b) + G_1(b,a) = G_2(a,b) + G_2(b,a) = G(a,b), \tag{4}$$

where the overall Fechnerian distance $G(a,b)$ (we need not now specify of which kind) can be easily checked to satisfy all properties of a metric (Dzhafarov, 2002d; Dzhafarov & Colonius, 2005a). Caution should be exercised though: the observation-area-invariance of the overall Fechnerian distance is predicated on the *canonical form* of Regular Minimality, (3). In a more general case, as explained in Section 3.6, $G_1(a, b) + G_1(b, a)$ equals $G_2(a', b') + G_2(b', a')$ if a and a' (as well as b and b') are "points of subjective equality," not necessarily physically identical.

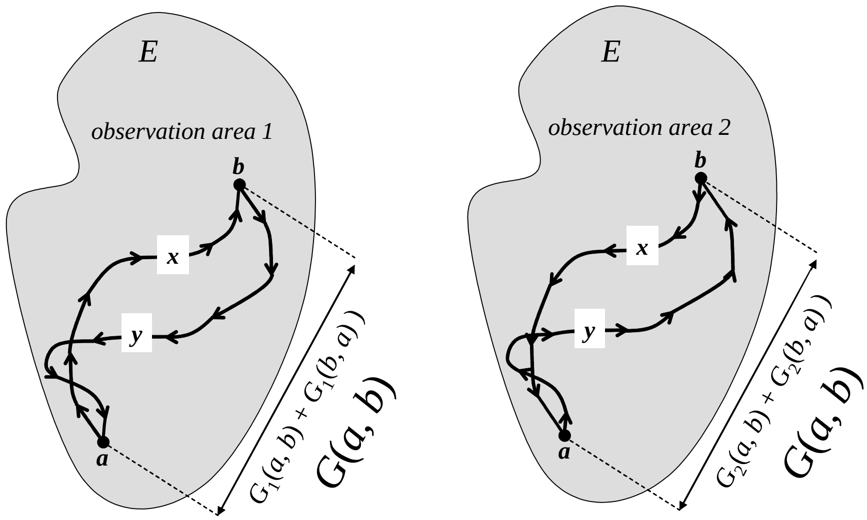

Figure 3: Illustration for the Second Main Theorem: the psychometric length of the first kind of a closed loop from a to b and back equals the psychometric length of the second kind for the same loop traversed in the opposite direction. This leads to the equality of the overall Fechnerian distances in the two observation areas.

Equation (4) is an immediate consequence of the following proposition (Dzhafarov, 2002d; Dzhafarov & Colonius, 2005a): for any smooth arcs $a \to x \to b$ and $b \to y \to a$,

$$L^{(1)}[a \to x \to b] + L^{(1)}[b \to y \to a] = L^{(2)}[a \to y \to b] + L^{(2)}[b \to x \to a]. \quad (5)$$

Put differently, the psychometric length of the first kind for any *closed loop* containing a and b equals the psychometric length of the second kind for the same closed loop but traversed in the opposite direction. Together (5) and its corollary (4) constitute what we call the Second Main Theorem of MDFS (see Fig. 3). This theorem plays a critical role in extending the continuous theory to discrete and other, more complex object spaces (Dzhafarov & Colonius, 2005b).

3. FSDOS

1. Discrete object spaces. Recall that a space of stimuli (objects) is a set S of all objects of a particular kind endowed with a discrimination probability function $\psi(x, y)$. For any distinct $x, y \in S$ we define *psychometric increments* of the first and second kind (or, in the first and second observation areas) as, respectively,

$$\phi^{(1)}(x, y) = \psi(x, y) - \psi(x, x), \quad \phi^{(2)}(x, y) = \psi(y, x) - \psi(x, x). \tag{6}$$

Due to Regular Minimality, (3), $\phi^{(\iota)}(x, y) > 0$, $\iota = 1, 2$. A space S is called *discrete* if, for any $x \in S$, $\inf_y \left[\phi^{(\iota)}(x, y)\right] > 0$, $\iota = 1, 2$. In other words, the psychometric increments of both kinds from x to other objects cannot fall below some positive quantity ("get arbitrarily close" to x). Clearly, objects in a discrete space cannot be connected by arcs (continuous images of intervals of reals).

2. Main idea. To understand Fechnerian computations in discrete spaces, return for a moment to a continuous spaces E (Section 2.2). Consider a smooth arc $x(t) : [\alpha, \beta] \to E$, $x(\alpha) = a$, $x(\beta) = b$, as in Fig. 4. We know that its psychometric length $L^{(\iota)}[a \to x \to b]$ of the ιth kind ($\iota = 1, 2$) is obtained by cumulating psychometric differentials of the same kind from α to β. It is also possible, however, to approximate $L^{(\iota)}[a \to x \to b]$ by partitioning $[\alpha, \beta]$ into $\alpha = t_0, t_1, ..., t_k, t_{k+1} = \beta$ and computing the sum of the chained psychometric increments

$$L^{(\iota)}[x(t_0), x(t_1), ..., x(t_{k+1})] = \sum_{i=0}^{k} \phi^{(\iota)}(x(t_i), x(t_{i+1})), \quad \iota = 1, 2 \tag{7}$$

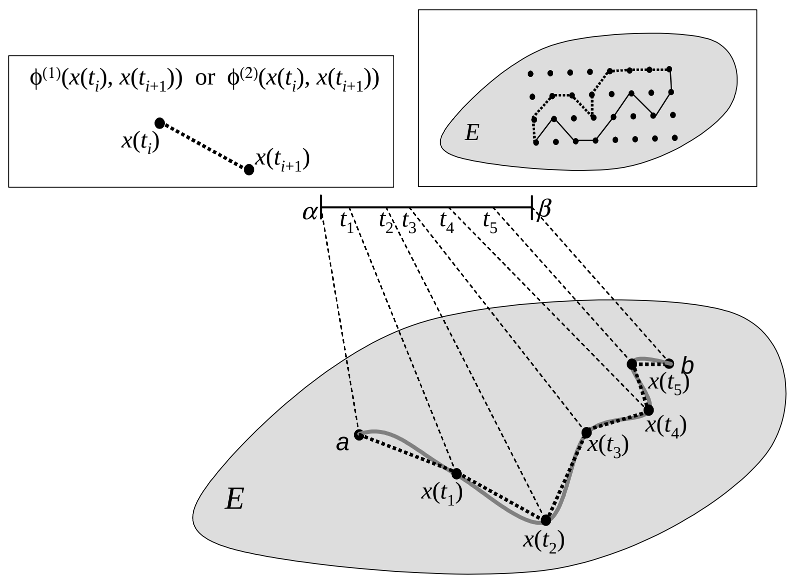

Figure 4: The psychometric length of the first (second) kind of an arc can be approximated by the sum of psychometric increments of the first (second) kind chained along the arc. The right insert shows that if E is represented by a dense grid of points, the Fechnerian computations involve taking all possible chains leading from one point to another through successions of immediately neighboring points.

As shown in Dzhafarov and Colonius (2005a), by progressively refining the partitioning this sum can be made as close to the value of $L^{(\iota)}[a \to x \to b]$ as one wishes. In practical computations, E can be represented by a sufficiently dense discrete grid of points. In view of the result just mentioned, the oriented Fechnerian distance $G_\iota(a,b)$ ($\iota = 1,2$) in this case can be approximated by (A) considering all possible chains of successive neighboring points leading from a to b, (B) computing sums (7) for each of these chains, and (C) taking the smallest value.

This almost immediately leads to the algorithm for Fechnerian computations in discrete spaces. The main difference is that in discrete spaces we have no physical ordering of objects to rely on: every point in a discrete space can be viewed as a "neighbor" of any other point. Consequently, in place of "all possible chains of successive neighboring points leading from a to b" one has to consider simply *all possible chains of points leading from a to b* (see Fig. 5).

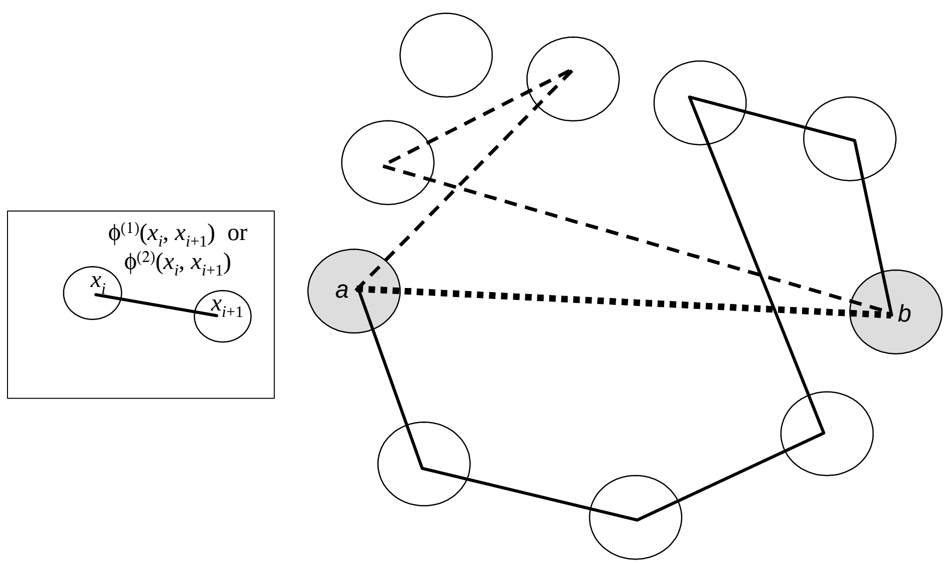

Figure 5: In a discrete space (10 elements whereof are shown in an arbitrary spatial arrangement) Fechnerian computations are performed by taking sums of psychometric increments (of the first or second kind, as shown in the inset) for all possible chains leading from one point to another.

3. Illustration. Returning to our toy example (Table 1), let us compute the Fechnerian distance between, say, objects D and B. The stimulus space here is $\{A, B, C, D\}$, and we have five different chains in this space which are comprised of distinct (nonrecurring) objects and lead from D to B: $DB, DAB, DCB, DACB, DCAB$. We begin by computing their psychometric lengths of the first kind, $L^{(1)}[DB]$, $L^{(1)}[DAB]$, etc. By analogy with (7), $L^{(1)}[DCAB]$, for example, is computed as $L^{(1)}[DCAB] = \phi^{(1)}(D,C)+\phi^{(1)}(C,A)+\phi^{(1)}(A,B)$. Using the definition of $\phi^{(1)}(x,y)$ in (6), $L^{(1)}[DCAB] = [\psi(D,C) - \psi(D,D)] + [\psi(C,A) - \psi(C,C)] + [\psi(A,B) - \psi(A,A)] = [0.7 - 0.5] + [1.0 - 0.5] + [0.8 - 0.1] = 1.4$. Repeating this procedure for all our five chains, we find out that the smallest value is $L^{(1)}[DCB] = \phi^{(1)}(D,C) + \phi^{(1)}(C,B) = [\psi(D,C) - \psi(D,D)] + [\psi(C,B) - \psi(C,C)] = 0.3$. Note that this value is smaller than the length of the one-link chain DB: $L^{(1)}[DB] = \phi^{(1)}(D,B) = \psi(D,B) - \psi(D,D) = 0.5$. The chain DCB can be called a *geodesic chain* connecting D to B. Its length is taken to be the oriented Fechnerian distance of the first kind from D to B, $G_1(D,B) = 0.3$. (Generally there can be more than one geodesic chain, of the same length, for a given pair of objects.)

Consider now the same five chains but viewed in the opposite direction, that is, all chains in $\{A, B, C, D\}$ leading from B to D, and compute for these chains the psychometric lengths of the first kind: $L^{(1)}[BD]$, $L^{(1)}[BAD]$, etc. Having done this we will find out that this time the shortest chain is the one-link chain BD, with the length $L^{(1)}[BD] = \phi^{(1)}(B,D) = \psi(B,D) - \psi(B,B) = 0.8$. The geodesic chain from B to D therefore is BD, and $G_1(B,D) = 0.8$.

Using the same logic as for continuous stimulus spaces, we now compute the (symmetrical) overall Fechnerian distance between D and B by adding the two oriented distances "to and fro," $G(D,B) = G(B,D) = G_1(D,B) + G_1(B,D) = 0.3 + 0.8 = 1.1$. This is the value we find in cells (D,B) and (B,D) of matrix $G1$ in Table 2. The concatenation of the two geodesic chains DCB

and BD forms the *geodesic loop* between D and B, which we find in cells (D,B) and (B,D) of matrix $L1$ in Table 2. This loop, of course, can be written in three different ways depending on which of its three distinct elements we choose to begin and end with. The convention we adopt is to begin and end with the row object: $DCBD$ in cell (D,B) and $BDCB$ in cell (B,D). Note that the overall Fechnerian distance $G(D,B)$ and the corresponding geodesic loop could also be found by computing psychometric lengths for all 25 possible closed loops containing objects D and B in space $\{A,B,C,D\}$ and finding the smallest. This, however, would be a more wasteful procedure.

The reason we do not need to add the qualification "of the first kind" to the designations of the overall Fechnerian distance $G(D,B)$ and the geodesic loop $DCBD$ is that precisely the same value of $G(D,B)$ and the same geodesic loop (only traversed in the opposite direction) are obtained if the computations are performed with psychometric increments of the second kind. For chain $DCAB$, for example, the psychometric length of the second kind, using the definition of $\phi^{(2)}$ in (6), is computed as $L^{(2)}[DCAB] = \phi^{(2)}(D,C) + \phi^{(2)}(C,A) + \phi^{(2)}(A,B) = [\psi(C,D) - \psi(D,D)] + [\psi(A,C) - \psi(C,C)] + [\psi(B,A) - \psi(A,A)] = 1.3$. Repeating this computation for all our five chains leading from D to B, the shortest chain will be found to be DB, with the length $L^{(2)}[DB] = \phi^{(2)}(D,B) = \psi(B,D) - \psi(D,D) = 0.4$, taken to be the value of $G_2(D,B)$, the oriented Fechnerian distance form D to B of the second kind. For the same five chains but viewed as leading from B to D, the shortest chain is BCD, with the length $L^{(2)}[BCD] = \phi^{(2)}(B,C) + \phi^{(2)}(C,D) = [\psi(C,B) - \psi(B,B)] + [\psi(D,C) - \psi(C,C)] = 0.7$, taken to be the value of $G_2(B,D)$. Their sum is $G(D,B) = G(B,D) = G_2(D,B) + G_2(B,D) = 0.4 + 0.7 = 1.1$, the same value for the overall Fechnerian distance as before (even though the oriented distances are different). The geodesic loop obtained by concatenating the geodesic chains DB and BCD is also the same as we find in matrix $L1$ in cells (D,B) and (B,D), but read from right to left: $DBCD$

in cell (D, B) and $BCDB$ in cell (B, D).

The complete formulation of the convention adopted in $L1$ therefore is as follows: the geodesic loop in cell (x, y) begins and ends with x and is read from left to right for the computations of the first kind, and from right to left for the computations of the second kind (yielding one and the same result, the overall Fechnerian distance between x and y).

4. Procedure of FSDOS.[2] It is clear that any finite set $S = \{s_1, s_2, ..., s_N\}$ endowed with probabilities $p_{ij} = \psi(s_i, s_j)$ forms a discrete space in the sense of our formal definition. As this case is of the greatest interest in empirical applications, in the following we will confine our discussion to finite object spaces. All our statements, however, unless specifically qualified, apply to discrete object spaces of arbitrary cardinality. The procedure below is described as if one knew the probabilities p_{ij} on the population level. If sample sizes do not warrant this approximation, the procedure should ideally be repeated with a large number of matrices p_{ij} that are statistically retainable given a matrix of frequency estimates $\hat{p}_{ij}$. We return to this issue in the concluding section.

The computation of Fechnerian distances G_{ij} among $\{s_1, s_2, ..., s_N\}$ proceeds in several steps. The first step is to check for Regular Minimality: for any i and all $j \neq i$, $p_{ii} < \min\{p_{ij}, p_{ji}\}$. If Regular Minimality is violated (on the population level), FSDOS will not work. Put differently, given a matrix of frequency estimates $\hat{\psi}(s_i, s_j)$, one should use statistically retainable matrices of probabilities p_{ij} that do satisfy Regular Minimality; and if no such matrices can be found, FSDOS is not applicable. Having Regular Minimality verified, we compute psychometric increments of

[2] **An algorithmic description of FSDOS as well as a computer program implementing it can be downloaded from http://www.psych.purdue.edu/~ehtibar. It is written in Matlab 6.0 and MS Excel XP. (The program also performs some computations not discussed in this paper.)**

the first and second kind, $\phi^{(1)}(s_i, s_j) = p_{ij} - p_{ii}$, $\phi^{(2)}(s_i, s_j) = p_{ji} - p_{ii}$, which are positive for all $j \neq i$.

Consider now a chain of objects $s_i = x_1, x_2, ..., x_k = s_j$ leading from s_i to s_j, with $k \geq 2$. The psychometric length of the first kind for this chain, $L^{(1)}[x_1, x_2, ..., x_k]$, is defined as $L^{(1)}[x_1, x_2, ..., x_k] = \sum_{m=1}^{k} \phi^{(1)}(x_m, x_{m+1})$. The set of different psychometric lengths across all possible chains of distinct elements connecting s_i to s_j being finite, it contains a minimum value $L^{(1)}_{\min}(s_i, s_j)$. (The consideration can always be confined to chains $(x_1, x_2, ..., x_k)$ of distinct elements, because if $x_l = x_m$, $l < m$, the length cannot increase if the subchain $(x_{l+1}, ..., x_m)$ is removed.) This value is called the oriented Fechnerian distance of the first kind from object s_i to object s_j: $G_1(s_i, s_j) = L^{(1)}_{\min}(s_i, s_j)$. G_1 satisfies all properties of a metric, except for symmetry: (A) $G_1(s_i, s_j) \geq 0$; (B) $G_1(s_i, s_j) = 0$ if and only if $i = j$; (C) $G_1(s_i, s_j) \leq G_1(s_i, s_m) + G_1(s_m, s_j)$; but in general, $G_1(s_i, s_j) \neq G_1(s_j, s_i)$. Properties (A) and (B) trivially follow from the fact that for $i \neq j$, $G_1(s_i, s_j)$ is the smallest of several positive quantities, $L^{(1)}[x_1, x_2, ..., x_k]$. Property (C) follows from the observation that the chains leading from s_i to s_j through a fixed s_k form a proper subset of all chains leading from s_i to s_j.[3] In accordance with the general logic of Fechnerian Scaling, $G_1(s_i, s_j)$ is interpreted as the oriented Fechnerian distance from s_i to s_j in the first observation area. Any chain from s_i to s_j whose elements are distinct and whose length equals $G_1(s_i, s_j)$ is a geodesic chain from s_i to s_j. There may be more than one geodesic chain for given s_i, s_j. (Note that in the case of infinite discrete sets geodesic chains need not exist.)

The oriented Fechnerian distances $G_2(s_i, s_j)$ of the second kind (in the second observation

[3] For finite sets S we can always find the minimum of $L^{(1)}[x_1, x_2, ..., x_k]$ across all chains with fixed endpoints. For an infinite discrete S the minimum need not exist and $L^{(1)}_{\min}(a, b)$ should be replaced with $L^{(1)}_{\inf}(a, b)$, the infimum of $L^{(1)}[a = x_1, x_2, ..., x_k = b]$. The argument for properties (A) and (B) then should be modified: for $a \neq b$, $G_1(a, b) > 0$ because $L^{(1)}_{\inf}(a, b) \geq \inf_x \left[\phi^{(1)}(a, x)\right]$, and by definition of discrete object spaces, $\inf_x \left[\phi^{(1)}(a, x)\right] > 0$.

area) and the corresponding geodesic chains are computed analogously, using the chained sums of psychometric increments $\phi^{(2)}$ instead of $\phi^{(1)}$.

As argued in Section 2.1, the order of two objects in a given observation area has no operational meaning, and we add the two oriented distances, "to and fro," to obtain the (symmetrical) overall Fechnerian distances: $G_{ij} = G_1(s_i, s_j) + G_1(s_j, s_i) = G_{ji}$, and also $G_{ij} = G_2(s_i, s_j) + G_2(s_j, s_i) = G_{ji}$. Quantity G_{ij} clearly satisfies all the properties of a metric. The validation for this procedure is provided by the fact that

$$G_1(s_i, s_j) + G_1(s_j, s_i) = G_2(s_i, s_j) + G_2(s_j, s_i), \tag{8}$$

i.e., the distance G_{ij} between the ith and the jth objects does not depend on the observation area in which these objects are taken. The proof of this fact is a trivial corollary of the following statement, which is of interest on its own sake: for any two chains $s_i = x_1, x_2, ..., x_k = s_j$ and $s_i = y_1, y_2, ..., y_l = s_j$ (connecting s_i to s_j),

$$L^{(1)}[x_1, x_2, ..., x_k] + L^{(1)}[y_l, y_{l-1}, ..., y_1] = L^{(2)}[y_1, y_2, ..., y_l] + L^{(2)}[x_k, x_{k-1}, ..., x_1]. \tag{9}$$

Indeed, denoting $p'_{ij} = \psi(x_i, x_j)$ and $p''_{ij} = \psi(y_i, y_j)$,

$$\begin{aligned} L^{(1)}[x_1, x_2, ..., x_k] + L^{(1)}[y_l, y_{l-1}, ..., y_1] &= \sum_{m=1}^{k-1}\left(p'_{m,m+1} - p'_{ii}\right) + \sum_{m=1}^{l-1}\left(p''_{m+1,m} - p''_{m+1,m+1}\right), \\ L^{(2)}[y_1, y_2, ..., y_l] + L^{(2)}[x_k, x_{k-1}, ..., x_1] &= \sum_{m=1}^{l-1}\left(p''_{m+1,m} - p''_{m,m}\right) + \sum_{m=1}^{k-1}\left(p'_{m,m+1} - p'_{m+1,m+1}\right). \end{aligned}$$

Subtracting the second equation from the first we get $p'_{kk} - p'_{11} + p''_{11} - p''_{kk}$, which is zero because

$p'_{11} = p''_{11} = p_{ii}$ and $p'_{kk} = p''_{kk} = p_{jj}$, where, we recall, $p_{ij} = \psi(s_i, s_j)$. Together (8) and (9) provide a simple version of the Second Main Theorem of Fechnerian Scaling, mentioned earlier.

An equivalent way of defining G_{ij} is to consider all *closed loops* $x_1, x_2, ..., x_n, x_1$ ($n \geq 2$) containing two given objects s_i, s_j: G_{ij} is the shortest of the psychometric lengths computed for all such loops. Note that the psychometric length of a loop depends on the direction in which it is traversed: generally, $L^{(1)}(x_1, x_2, ..., x_n, x_1) \neq L^{(1)}(x_1, x_n, ..., x_2, x_1)$, and $L^{(2)}(x_1, x_2, ..., x_n, x_1) \neq L^{(2)}(x_1, x_n, ..., x_2, x_1)$. The Second Main Theorem tells us, however, that $L^{(1)}(x_1, x_2, ..., x_n, x_1) = L^{(2)}(x_1, x_n, ..., x_2, x_1)$, that is, any closed loop in the first observation area has the same length in the second observation area if traversed in the opposite direction. In particular, if $x_1, x_2, ..., x_n, x_1$ is a geodesic (i.e., shortest) loop containing the objects s_i, s_j in the first observation area (the concatenation of the geodesic chains connecting s_i to s_j and s_j to s_i), then the same loop is a geodesic loop in the second observation area, if traversed in the opposite direction, $x_1, x_n, ..., x_2, x_1$.

5. Two examples. The algorithm just described was used to compute Fechnerian distances and geodesic loops for the 36×36 data set reported in Rothkopf (1957) and the 32×32 data reported in Wish (1967). Only small subsets of these object sets are shown in Tables 3 and 4, chosen because they form "self-contained" *subspaces*: any two elements of each subset can be connected by a geodesic loop lying entirely within the subset. The discrimination probabilities satisfy Regular Minimality in the canonical form: the main diagonal values in the two tables are both row and column minima.[4] Recall our convention on presenting geodesic loops. Thus, in Table

[4] In the complete 32×32 matrix reported in Wish (1967) there are two violations of Regular Minimality, both due to a single value, $\hat{p}_{TV} = 0.03$: this value is the same as $\hat{p}_{VV}$ and smaller than $\hat{p}_{TT} = 0.06$ (using the labeling described in Table 4). As we used Wish's data for illustration purposes only, we simply replaced $\hat{p}_{TV} = 0.03$ with $p_{TV} = 0.07$, putting $p_{ij} = \hat{p}_{ij}$ for the rest of the data. Chi-square deviation of thus defined matrix of p_{ij} from the matrix of $\hat{p}_{ij}$ is negligibly small. A comprehensive procedure should have involved a repeated generation of statistically retainable p_{ij} matrices subject to Regular Minimality.

Table 4: A 10×10 excerpt from Wish's (1967) 32×32 data matrix. Stimuli were 5-element sequences $T_1P_1T_2P_2T_3$, where T stands for a tone (short or long) and P stands for a pause (1 or 3 units long). We arbitrarily labeled the stimuli $A, B, ..., Z, 0, 1, ..., 5$, in the order they are presented in Wish's paper. The format of the table and the criterion for choosing this particular subset of 10 stimuli are the same as in Table 3.

	S	U	W	X	0	1	2	3	4	5
S	0 (6) S	32 (16) SUS	72 (38) SWS	89 (45) SUXS	57 (35) S0S	119 (73) SU1WS	112 (81) SU2US	128 (70) SUX3XS	119 (89) SUX4WS	138 (97) SUX5XS
U	32 (28) USU	0 (6) U	76 (44) UWU	79 (24) UXWU	89 (59) US0SU	107 (56) U1WU	80 (49) U2U	116 (51) UX31WU	107 (71) UX4WU	128 (69) UX5XWU
W	72 (44) WSW	76 (42) WUW	0 (4) W	30 (11) WXW	119 (78) WS0W	55 (40) W1W	122 (79) W2XW	67 (55) WX31W	58 (48) WX4W	79 (83) WX5XW
X	89 (64) XSUX	79 (71) XWUX	30 (26) XWX	0 (3) X	123 (86) X0X	67 (51) X31WX	94 (73) X2X	39 (27) X3X	45 (31) X4X	49 (44) X5X
0	57 (34) 0S0	89 (55) 0SUS0	119 (56) 0WS0	123 (46) 0X0	0 (6) 0	113 (52) 010	71 (39) 020	143 (69) 0130	95 (39) 040	132 (95) 0250
1	119 (84) 1WSU1	107 (75) 1WU1	55 (22) 1W1	67 (33) 1WX31	113 (70) 101	0 (3) 1	109 (69) 121	31 (17) 131	72 (40) 141	08 (97) 135X31
2	112 (81) 2USU2	80 (44) 2U2	122 (62) 2XW2	94 (31) 2X2	71 (45) 202	109 (50) 212	0 (7) 2	116 (41) 232	92 (35) 242	74 (26) 252
3	128 (94) 3XSUX3	116 (85) 31WUX3	67 (44) 31WX3	39 (17) 3X3	143 (85) 3013	31 (19) 313	116 (84) 323	0 (2) 3	84 (63) 3X4X3	77 (47) 35X3
4	119 (89) 4WSUX4	107 (73) 4WUX4	58 (26) 4WX4	45 (20) 4X4	95 (65) 404	72 (38) 414	92 (67) 424	84 (45) 4X3X4	0 (3) 4	68 (49) 454
5	138 (100) 5XSUX5	128 (94) 5XWUX5	79 (74) 5XWX5	49 (11) 5X5	132 (83) 5025	108 (95) 5X3135	74 (58) 525	77 (67) 5X35	68 (25) 545	0 (3) 5

3 the geodesic chain from letter B to digit 8 in the first observation area is $B \to 5 \to 6 \to 7 \to 8$ and that from 8 to B is $8 \to 7 \to 5 \to B$. In the second observation area the geodesic chains should be read from right to left: $8 \leftarrow 7 \leftarrow 5 \leftarrow B$ from B to 8, and $B \leftarrow 5 \leftarrow 6 \leftarrow 7 \leftarrow 8$ from 8 to B. The oriented Fechnerian distances are $G_1(B, 8) = .70$, $G_1(8, B) = .70$, $G_2(B, 8) = .77$, and $G_2(8, B) = .63$, yielding $G(8, B) = 1.40$.

Note that Fechnerian distances G_{ij} are not monotonically related to discrimination probabilities p_{ij}: there is no functional relationship between the two because the computation of G_{ij} for a *given* (i, j) involves p_{ij} values for *all* (i, j). Nor are the oriented Fechnerian distances $G_1(s_i, s_j)$ and $G_2(s_i, s_j)$ monotonically related to psychometric increments $p_{ij} - p_{ii}$ and $p_{ji} - p_{ii}$, due to the existence of longer-than-one-link geodesic chains. There is, however, a strong positive correlation between p_{ij} and G_{ij}:[5] 0.94 for Rothkopf's data and 0.89 for Wish's data (Pearson correlation for the entire matrices, 36×36 and 32×32). This indicates that the probability-distance hypothesis, even if known to be false mathematically, may still be acceptable as a crude approximation. We may expect consequently that MDS-distances could provide crude approximations to the Fechnerian distances. That the adjective "crude" cannot be dispensed with is indicated by the relatively low values of Kendall's correlation between p_{ij} and G_{ij}: 0.76 for Rothkopf's data and 0.68 for Wish's data.

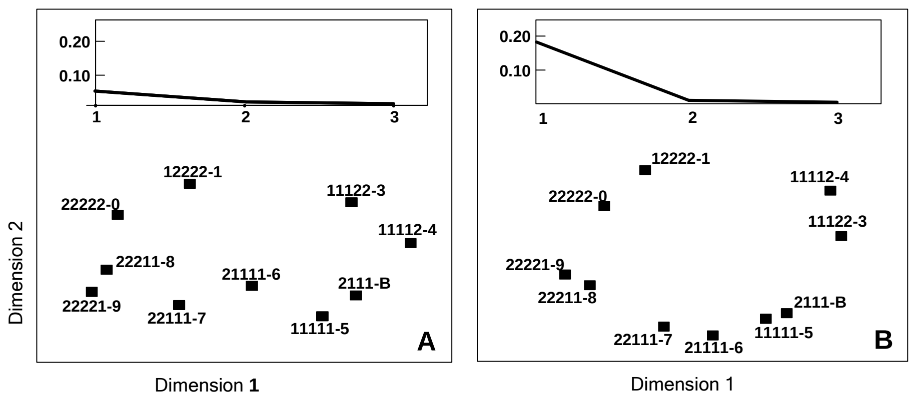

Figure 6: Two-dimensional Euclidean representations for discrimination probabilities (nonmetric MDS, Panel A) and for Fechnerian distances (metric MDS, Panel B) in Table 3. The MDS program used is PROXSCAL 1.0 in SPSS 11.5, minimizing raw stress. Sequence of 1's and 2's preceding a dash is the Morse code for the symbol following the dash. Insets are scree plots (normalized raw stress versus number of dimensions).

MDS can be used in conjunction with FSDOS, as a follow-up analysis once Fechnerian distances have been computed. A nonmetric version of MDS can be applied to Fechnerian distances (as opposed to discrimination probabilities directly) simply to provide a rough graphical representation for matrices like in Tables 3 and 4. More interestingly, a *metric* version of MDS can be applied to Fechnerian distances to test the hypothesis that Fechnerian distances, not restricted

[5] We are grateful to Associate Editor for pointing this out to us.

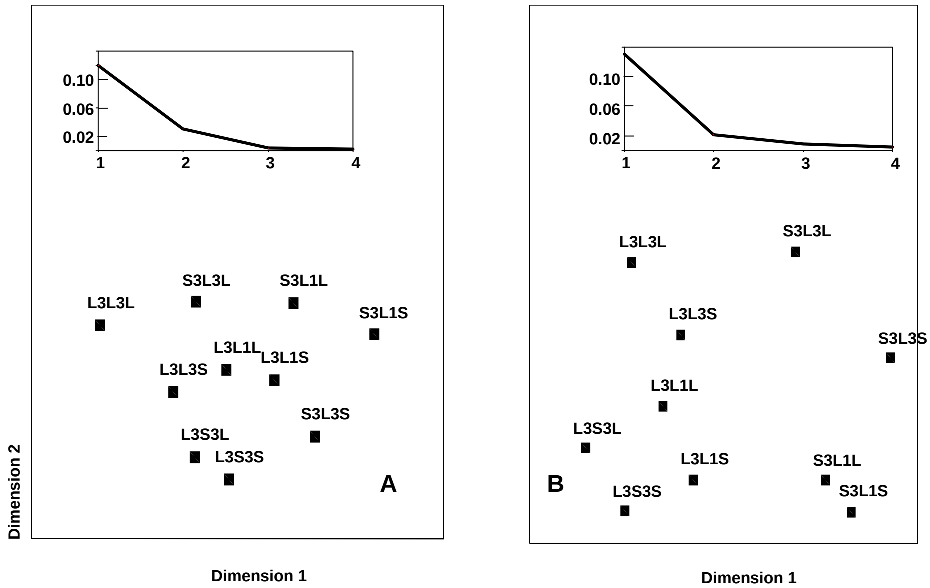

Figure 7: Same as Fig. 6, but for discrimination probabilities (nonmetric MDS, Panel A) and for Fechnerian distances (metric MDS, Panel B) in Table 4. *L* stands for long tone, *S* for short tone, while digits 1 an 3 show the lengths of the two pauses.

a priori to any particular class (except for being intrinsic), can be approximated by Euclidean distances (more generally, power-function Minkowski ones); the degree of approximation for any given dimensionality is measured by the achieved stress value. Geometrically, metric MDS on Fechnerian distances is an attempt to *isometrically embed* (i.e., map without distorting pairwise distances) the discrete object space in a low-dimensional Euclidean (or Minkowskian) space. Figures 6 and 7 provide a comparison of the metric MDS on Fechnerian distances with nonmetric MDS performed on discrimination probabilities directly, for the 10×10 submatrices presented in Tables 3 and 4. Using the value of normalized raw stress as our criterion, the two-dimensional solution is almost equally good in both analyses. Therefore to the extent we consider the traditional MDS solution acceptable, we can view the Fechnerian distances in these two cases as being approximately Euclidean. The configurations of points obtained by metric MDS on Fechnerian distances and nonmetric MDS on discrimination probabilities are more similar in Fig. 6 than in Fig. 7, indicating that MDS-distances provide a better approximation to Fechnerian distances in the former case. This may reflect the fact that the Kendall correlation between the probabilities and Fechnerian distances for Rothkopf's data is higher than for Wish's data (0.76 vs 0.68). A detailed comparison of the configurations provided by the two analyses, as well as such related issues as interpretation of axes are, however, beyond the scope of this paper.

6. General form of Regular Minimality. In continuous stimulus spaces it often happens that for a fixed value of x, $\psi(x, y)$ achieves its minimum not at $y = x$ but at some other value of y; and for a fixed value of y, $\psi(x, y)$ achieves its minimum at some value other than $x = y$. It has been noticed since Fechner (1860), for example, that when x and x are presented in a succession, the second stimulus often seems larger (bigger, brighter, etc.) than the first: this is the classical phenomenon of "time error." It follows that in a successive pair (x, y) the two elements maximally resemble each other when y is physically smaller than x. Borrowing the terminology from the theory of greater-less comparisons, Dzhafarov (2002d, 2003a) proposed to call the value of $\arg\min_y \psi(x, y)$ (i.e., the value of y at which $\psi(x, y)$ achieves its minimum, for a fixed x) the *point of subjective equality*, *PSE*, for x; and analogously, the value of $\arg\min_x \psi(x, y)$ is called the PSE for y.

Using this terminology, the general formulation of the Regular Minimality principle is as follows: (a) every x in the first observation area has a unique PSE in the second observation area; (b) every y in the second observation area has a unique PSE in the first observations area; and (c) y is the PSE for x if and only if x is the PSE for y. For a detailed discussion, see Dzhafarov (2002d, 2003a) and Dzhafarov and Colonius (2005a).

Table 5: A matrix of discrimination probabilities satisfying Regular Minimality in a non-canonical form. A canonical transformation of this matrix yields matrix $M1$ of Table 1.

$M0$	a	b	c	d
a	0.6	0.6	**0.1**	0.8
b	0.9	0.9	0.8	**0.1**
c	1	**0.5**	1	0.6
d	**0.5**	0.7	1	1

Clearly, the formulation of Regular Minimality that we used so far, (3), is a special case: $\arg\min_y \psi(x,y) = x$ and $\arg\min_x \psi(x,y) = y$ (i.e., two physically identical stimuli are mutual PSEs). As we know, this form of Regular Minimality is called *canonical*. It is possible that in discrete stimulus spaces Regular Minimality always has this special form, but it need not be so a priori. It seems useful therefore to reformulate the algorithm and interpretation of FSDOS under the assumption that Regular Minimality holds in its general form. Consider the following modification of our toy example. Let the object space now be $\{a,b,c,d\}$ and the initial matrix look as in Table 5. It is easy to see that Regular Minimality holds here, although not in a canonical form: every row contains a single minimal cell, and this cell is also minimal in its column. We can make a list of mutual PSEs and relabel them, assigning one and the same label to every pair of PSEs: $(a,c) \to A$, $(b,d) \to B$, $(c,b) \to C$, $(d,a) \to D$. With this relabeling, the matrix $M0$ of Table 5 transforms into the matrix $M1$ of our initial toy example (Table 1), with Regular Minimality now holding in the canonical form, (3). Due to Regular Minimality this can always be achieved by an appropriate relabeling. Having performed the Fechnerian analysis on $M1$ and having computed the matrices $L1$ and $G1$ of Table2, we can return to the original labeling and present the Fechnerian distances and geodesic loops separately for the first and the second observation areas. A,B,C,D in $L1$ and $G1$ should be replaced with, respectively, a,b,c,d for the first observation area, and with

c, d, b, a for the second observation area. Denoting the overall Fechnerian distance G in the first and second observation areas by $G^{(1)}$ and $G^{(2)}$, respectively (not to be confused with the oriented Fechnerian distances distances G_1 and G_2), we will see, for instance, that $G^{(1)}(a, b)$ is 1.3, while $G^{(2)}(a, b)$ is 0.7, reflecting the fact that a, b are perceived differently in the two observation areas. On the other hand, $G^{(2)}(c, d)$ is 1.3., the same as $G^{(1)}(a, b)$. This reflects the fact that c, d in the second observation area are PSEs for, respectively, a, b in the first observation area.

4. Concluding Remarks

1. Statistical issues. In some applications the frequency estimates of $p_{ij} = \psi(s_i, s_j)$ are computed from samples sufficiently large to ignore statistical issues and treat FSDOS as being performed on essentially a population level. To a large extent this is how the theory of FSDOS is presented in this paper. The questions of finding the joint sampling distribution for Fechnerian distances $\hat{G}_{ij}$ $(i, j = 1, 2, ..., N)$ or joint confidence intervals for G_{ij} are beyond the scope of this paper. We can, however, outline a general approach. The estimators $\hat{P}_{ij}$ of the probabilities p_{ij} are obtained as

$$\hat{P}_{ij} = \frac{1}{R_{ij}} \sum_{k=1}^{R_{ij}} X_{ijk},$$

where $\{X_{ij1}, ..., X_{ijR_{ij}}\}$ are random variables representing binary responses ($1 = different$, $0 = same$). The index k may represent chronological trial numbers for (s_i, s_j), different examples of this pair, different respondents, etc. Random variables X_{ijk} and $X_{i'j'k'}$ can be treated as stochastically independent, provided $(i, j, k) \neq (i', j', k')$. Assuming that $\Pr[X_{ijk} = 1]$ does not vary too much as a function of k (i.e., ignoring such factors as fatigue, learning, and individual differences), $\hat{P}_{ij}$ may be viewed as independent normally distributed variables with means p_{ij} and variances $p_{ij}(1-p_{ij})/R_{ij}$, from which it would follow that the joint distribution of the psychometric lengths of all chains with distinct elements is asymptotically multivariate normal, with both the means and covariances being known functions of true probabilities p_{ij}. The problem then is reduced to finding the (asymptotic) joint sampling distribution of the minima of psychometric lengths with common terminal points. Realistically, the problem is more likely to be dealt with by means of Monte Carlo simulations.

Monte Carlo is also likely to be used for constructing joint confidence intervals for G_{ij}, given a matrix of $\hat{p}_{ij}$. The procedure consists of repeatedly replacing the latter with matrices of p_{ij} that are subject to Regular Minimality and deviate from $\hat{p}_{ij}$ less than some critical value (e.g., by the conventional chi-square criterion), and computing Fechnerian distances from each of these matrices.

2. Choice of object set. In some cases, as with Rothkopf's (1957) Morse codes, the set S of objects used in an experiment or computation may contain all objects of a given kind. If such a set is too large or infinite, however, one can only use a subset S' of the entire S. This gives rise to a problem: the Fechnerian distance $G(a,b)$ between objects $a, b \in S'$ will generally depend on what other objects are included in S'. In a psychophysical experiment, when pairs of objects are presented repeatedly to a single observer, adding a new object s to S' may change the pairwise discrimination probabilities $\psi(a,b)$ within the old subset. In a group experiment with each pair presented just once, or for the "paper-and-pencil" perceivers, adding s to S' may not change discrimination probabilities, but this will still add new loops containing any given $a, b \in S'$; as a result, the minimum psychometric length $L^{(\iota)}_{\min}(s_i, s_j)$ will generally decrease.[6]

[6] This decrease must not be interpreted as a decrease in subjective dissimilarity. As mentioned earlier, Fechnerian distances are determined up to multiplication by an arbitrary positive constant, which means that only *relative*

A formal approach to this issue is to simply state that the Fechnerian distance between two given objects is a relative concept: $G(a,b)$ shows how far apart the two objects are "from the point of view" of a given perceiver and *with respect to a given object set*. This approach may be sufficient in a variety of applications, especially in psychophysical experiments with repeated presentations of pairs to a single observer: one might hypothesize that the observer in such a situation gets adapted to the immediate context of the objects in play, effectively confining to it the subjective "universe of possibilities." A discussion of this "adaptation to subspace" hypothesis can be found in Dzhafarov and Colonius (2005a). Like in many other situations involving sampling, however (including, e.g., sampling of respondents in a group experiment), one may only be interested in a particular subset S' of objects to the extent it is representative of the entire set S of objects of a particular kind. In this case one faces two distinctly different questions. The first question is empirical: is S' large enough (well chosen enough) for its further enlargements not to lead to noticeable changes in discrimination probabilities within S'? This question is not FSDOS-specific, any other analysis of discrimination probabilities (e.g., MDS) will have to address it too. The second question is computational, and it is FSDOS-specific: provided the first question is answered in the affirmative, is S' large (well chosen) enough for its further enlargements not to lead to noticeable changes in Fechnerian distances within S'? A detailed discussion being outside the scope of this paper, we can only mention what seems to be an obvious approach: the affirmative answer to second question can be given if one can show, by means of an appropriate version of subsampling, that the exclusion of a few objects from S' does not lead to changes in Fechnerian distances within the remaining subset.

Fechnerian distances $G(a,b)/G(c,d)$ are meaningfully interpretable.

3. Other empirical procedures. The procedure of pairwise presentations with same-different judgments is the focal empirical paradigm for FSDOS. With some caution, however, FSDOS can also be applied to other empirical paradigms, such as that of *identification*, mentioned in Section 1.3: all objects $\{s_1, s_2, ..., s_N\}$ are associated with rigidly fixed, normative reactions $\{R_1, R_2, ..., R_N\}$ (e.g., fixed names), and the objects are presented one at a time. Such an experiment results in the stimulus-response confusion probabilities $\eta(s_i, s_j)$ with which reaction R_j (normatively reserved for s_j) is given to object s_i. FSDOS here can be applied under the additional assumption that $\eta(s_i, s_j)$ can be interpreted as $1 - \psi(s_i, s_j)$. Regular Minimality here means that each object s_i has a single modal reaction R_j (in the canonical form, R_i), and then any other object causes R_j less frequently than s_i does. Thus understood, Regular Minimality is satisfied, for example, in the data reported in Shepard (1957, 1958). We reproduce here one of the matrices from this work (Table6, rows are stimuli, columns normative responses, entries conditional probabilities $\eta(s_i, s_j)$), together with the matrix of Fechnerian distances. Geodesic loops are not shown because the space $\{A, B, ..., I\}$ here turns out to be a "Fechnerian simplex": a geodesic chain from a to b in this space is always the one-link chain (a, b).

In a variant of the identification procedure, the reactions may be *preference ranks* for objects $\{s_1, s_2, ..., s_N\}$, R_1 designating, say, the most preferred object, R_N the least preferred. Suppose that Regular Minimality holds in the following sense: each object has a modal (most frequent) rank, each rank has a modal object, and R_j is the modal rank for s_i if and only if s_i is the modal object for R_j. Then the frequency $\phi(s_i, R_j)$ can be taken as an estimate of $1 - \psi(s_i, s_j)$, and the data be subjected to FSDOS. The fact that these and similar procedures are used in a variety of areas (psychophysics, neurophysiology, consumer research, educational testing, political science), combined with the great simplicity of the algorithm for FSDOS, makes one hope that its potential

Table 6: $M2$: one of the identification probability matrices reported in Shepard (1957, 1958). $G2$: Fechnerian distances computed from matrix $M2$.

$M2$	A	B	C	D	E	F	G	H	I
A	.678	.148	.054	.03	.025	.02	.016	.011	.016
B	0.167	0.544	0.066	0.077	0.053	0.015	0.045	0.018	0.015
C	0.06	0.07	0.615	0.015	0.107	0.067	0.022	0.03	0.014
D	0.015	0.104	0.016	0.542	0.057	0.005	0.163	0.032	0.065
E	0.037	0.068	0.12	0.057	0.46	0.075	0.057	0.099	0.03
F	0.027	0.029	0.053	0.015	0.036	0.715	0.015	0.095	0.014
G	0.011	0.033	0.015	0.145	0.049	0.016	0.533	0.052	0.145
H	0.016	0.027	0.031	0.046	0.069	0.096	0.053	0.628	0.034
I	0.005	0.016	0.011	0.068	0.02	0.021	0.061	0.018	0.78

$G2$	A	B	C	D	E	F	G	H	I
A	0	0.907	1.179	1.175	1.076	1.346	1.184	1.279	1.437
B	0.907	0	1.023	0.905	0.883	1.215	0.999	1.127	1.293
C	1.179	1.023	0	1.126	0.848	1.21	1.111	1.182	1.37
D	1.175	0.905	1.126	0	0.888	1.237	0.767	1.092	1.189
E	1.076	0.883	0.848	0.888	0	1.064	0.887	0.92	1.19
F	1.346	1.215	1.21	1.237	1.064	0	1.217	1.152	1.46
G	1.184	0.999	1.111	0.767	0.887	1.217	0	1.056	1.107
H	1.279	1.127	1.182	1.092	0.92	1.152	1.056	0	1.356
I	1.437	1.293	1.37	1.189	1.19	1.46	1.107	1.356	0

application sphere may be very large.

References

Borg, I., & Groenen, P. (1997). *Modern multidimensional scaling.* New York: Springer-Verlag.

DeSarbo, W. S., Johnson, M. D., Manrai, A. K., Manrai, L. A., & Edwards, E. A. (1992) TSCALE: A new multidimensional scaling procedure based on Tversky's contrast model. *Psychometrika*, **57**, 43-70.

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. (2003a). Thurstonian-type representations for "same-different" discriminations: Deterministic decisions and independent images. *Journal of Mathematical Psychology,* **47**, 208-228.

Dzhafarov, E.N. (2003b). 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.

Dzhafarov, E.N., & Colonius, H. (2005a). Psychophysics without physics: A purely psychological theory of Fechnerian Scaling in continuous stimulus spaces. *Journal of Mathematical Psychology,* **49**, 1-50.

Dzhafarov, E.N., & Colonius, H. (2005b). Psychophysics without physics: Extension of Fechnerian Scaling from continuous to discrete and discrete-continuous stimulus spaces. *Journal of Mathematical Psychology,* **49**, 125-141.

Fechner, G. T. (1860). *Elemente der Psychophysik* [Elements of psychophysics]. Leipzig: Breitkopf & Härtel.

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.

Izmailov, Ch. A., Dzhafarov, E. N, & Zimachev, M.M. (2001). Luminance discrimination probabilities derived from the frog electroretinogram. In: E. Sommerfeld, R. Kompass, T. Lachmann (Eds.) *Fechner Day 2001* (pp. 206-211). Lengerich: Pabst Science Publishers.

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.

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling.* Beverly Hills: Sage.

Rothkopf, E. Z. (1957). A measure of stimulus similarity and errors in some paired-associate learning tasks. *Journal of Experimental Psychology*, **53**, 94-102.

Roweis, S. T. , & Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science,* **290**, 2323-2326.

Sankoff, D. & Kruskal, J. (1999). *Time warps, string edits, and macromolecules.* Stanford, CA: CSLI Publications.

Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, **22**, 325-345.

Shepard, R. N. (1958). Stimulus and response generalization: Tests of a model relating generalization to distance in psychological space. *Journal of Experimental Psychology*, **55**, 509-523.

Shepard, R. N., and Carroll, J. D. (1966). Parametric representation of nonlinear data structures. In P. R. Krishnaiah (Ed.), *Multivariate Analysis* (pp. 561-592). New York: Academic Press.

Tenenbaum, J. B., de Silva, V., & Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science,* **290**, 2319-2323.

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

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

Weeks, D. G., & Bentler, P. M. (1982). Restricted multidimensional scaling models for asymmetric proximities. *Psychometrika*, **47**, 201-208.

Wish, M. (1967). A model for the perception of Morse code-like signals. *Human Factors*, **9**, 529-540.

Zimmer, K., & Colonius, H. (2000). Testing a new theory of Fechnerian scaling: The case of auditory intensity discrimination. *Journal of the Acoustical Society of America*, **108**, 2596.