MEASUREMENT WITH PERSONS

Theory, Methods, and Implementation Areas

Edited by
BIRGITTA BERGLUND
GIOVANNI B. ROSSI
JAMES T. TOWNSEND
LESLIE R. PENDRILL

Psychology Press
Taylor & Francis Group
New York London

Psychology Press
Taylor & Francis Group
711 Third Avenue
New York, NY 10017

Psychology Press
Taylor & Francis Group
27 Church Road
Hove, East Sussex BN3 2FA

Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number: 978-1-84872-939-1 (Hardback)

Library of Congress Cataloging-in-Publication Data

Measurements with persons : theory, methods, and implementation areas / editors, Birgitta Berglund ... [et al.].
p. cm. -- (Scientific psychology series)
Includes bibliographical references and index.
ISBN 978-1-84872-939-1 (hardback)
1. Perception--Mathematical models. 2. Senses and sensation--Mathematical models. 3. Human information processing--Mathematical models. I. Berglund, Birgitta. II. Title. III. Series.

BF311.M4325 2011
153.7028'7--dc22 2011008495

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the Psychology Press Web site at
http://www.psypress.com

9 Mathematical foundations of Universal Fechnerian Scaling

Ehtibar N. Dzhafarov

Purdue University
West Lafayette, Indiana

and

Swedish Collegium for Advanced Study
Uppsala, Sweden

9.1 Introduction

The main idea of Fechner's original theory (Fechner, 1860, 1877, 1887) can be described as follows (see Figure 9.1). If stimuli are represented by real numbers (measuring stimulus intensities, or their spatial or temporal extents), the *subjective distance* from a stimulus a to a stimulus $b > a$ is computed by *cumulating* from a to b, through all intermediate values, a measure of *dissimilarity* of every stimulus x from its "immediate" neighbors on the right. A modern rendering of Fechner's theory (Dzhafarov, 2001) defines the dissimilarity between x and $x + dx$ as

$$D\left(x,\, x+dx\right)=c\left(\gamma\left(x,\, x+dx\right)-\frac{1}{2}\right), \qquad \cancel{(9.1)}$$

where $\gamma\,(x, y)$ is a *psychometric function*

$$\gamma\,(x, y) = \Pr\,[y \text{ is judged to be greater than } x] \qquad \cancel{(9.2)}$$

with no *"constant error"* (i.e., $\gamma\,(x, x) = 1/2$), and c is a constant allowed to vary from one stimulus continuum to another. Assuming that $\gamma\,(x, y)$ is differentiable, and putting

$$\frac{D\left(x,\, x+dx\right)}{dx}=\left.\frac{\partial\gamma\left(x,y\right)}{\partial y}\right|_{y=x}=F\left(x\right),$$

the Fechnerian distance from a to $b \geq a$ becomes

$$G\left(a,b\right)=\int_a^b F\left(x\right)dx\,.$$

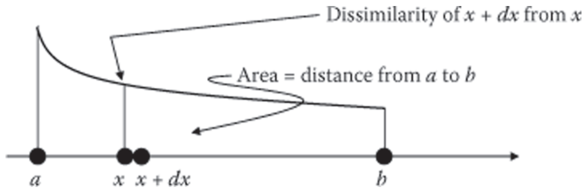

Figure 9.1 Fechner's main idea. To compute the subjective (Fechnerian) distance from a to b on a stimulus continuum, one cumulates (here, integrates) the dissimilarity of x from its infinitesimally close neighbors on the right as x changes from a to b.

In particular, if

$$F(x) = \frac{k}{x},$$

which is a rigorous form of *Weber's law*,

$$G(a,b) = k \log \frac{b}{a}.$$

We get the celebrated Fechner's law by setting a at the "absolute threshold" x_0,

$$S(x) = k \log \frac{x}{x_0},$$

where $S(x)$ can be referred to as the *magnitude of the sensation* caused by stimulus x. If $F(x)$ happens to be different from k/x, the expressions $G(a, b)$ and $S(x)$ are modified accordingly. Thus, from

$$F(x) = \frac{k}{x^{1-\beta}}, \; 1 \geq \beta > 0,$$

one gets

$$G(a,b) = \frac{k}{\beta}\left(b^{\beta} - a^{\beta}\right)$$

for the subjective distance from a to b, and

$$S(x) = \frac{k}{\beta}\left(x^{\beta} - x_0^{\beta}\right)$$

for the sensation magnitude of x. In this rendering Fechner's theory is impervious to the mathematical (Luce & Edwards, 1958) and experimental (Riesz, 1933) critiques levied against it (for details see Dzhafarov, 2001, and Dzhafarov & Colonius, 1999). The main idea of this interpretation was proposed by Pfanzagl (1962), and then independently reintroduced in Creelman (1967), Falmagne (1971), and Krantz (1971) within the framework of the so-called "Fechner problem" (Luce & Galanter, 1963).

Fechner's theory launched the world-view (or "mind-view") of classical psychophysics, according to which perception is essentially characterized by a collection of unidimensional continua representable by axes of nonnegative real numbers. Each continuum corresponds to a certain "sensory quality" (loudness, spatial extent, saturation, etc.) any two values of which, *sensory magnitudes*, are comparable in terms of "*less than or equal to*." Moreover, each such continuum has a primary physical correlate, an axis of nonnegative reals representing intensity, or spatiotemporal extent of a particular physical attribute: the sensory attribute is related to its physical correlate monotonically and smoothly, starting from the value of the absolute threshold. This mind-view has been dominant throughout the entire history of psychophysics (Stevens, 1975), and it remains perfectly viable at present (see, e.g., Luce, 2002, 2004).

There is, however, another mind-view, also derived from Fechner's idea of computing distances from local dissimilarity measures, dating back to Helmholtz's (1891) and Schrödinger's (1920, 1920/1970, 1926/1970) work on color spaces. Physically, colors are functions relating radiometric energy to wavelength, but even if their representation by means of one of the traditional color diagrams (such as CIE or Munsell) is considered their physical description, and even if the subjective representation of colors is thought of in terms of a finite number of unidimensional attributes (such as, in the case of aperture colors, their hue, saturation, and brightness), the mapping of physical descriptions into subjective ones is clearly that of one multidimensional space into another. In this context the notions of sensory magnitudes ordered in terms of "greater–less" and of psychophysical functions become artificial, if applicable at all. The notion of subjective dissimilarity, by contrast, acquires the status of a natural and basic concept, whose applicability allows for but does not presuppose any specific system of color coordinates, either physical or subjective. The natural operationalization of the discrimination of similar colors in this context is their judgment in terms of "*same or different*," rather than "greater or less." (For a detailed discussion of the "greater–less" versus "same–different" comparisons, see Dzhafarov, 2003a.)

This mind-view has been generalized in the theoretical program of *Multidimensional Fechnerian Scaling* (Dzhafarov, 2002a–d; Dzhafarov & Colonius, 1999, 2001). The scope of this differential-geometric program is restricted to stimulus spaces representable by *open connected regions of Euclidean n-space* (refer to Figure 9.2 for an illustration.). This space is supposed to be endowed with a probability-of-different function

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

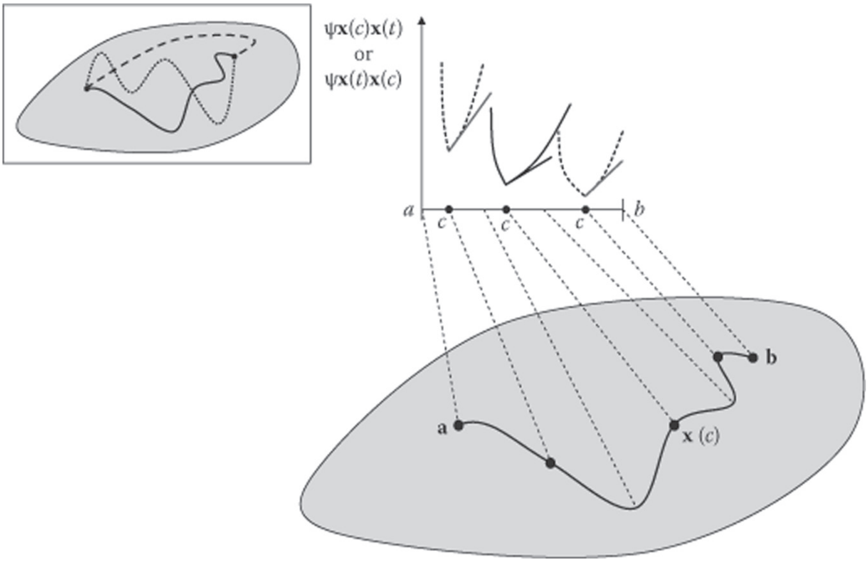

Figure 9.2 A continuously differentiable path $\mathbf{x}(t)$ (thick curve) is shown as a mapping of an interval $[a, b]$ (horizontal line segment) into an area of Euclidean space (gray area). For any point $c \in [a, b]$ there is a function $t \mapsto \psi(\mathbf{x}(c), \mathbf{x}(t))$ defined for all $t \in [a, b]$ (shown by *V*-shaped curves for three positions of c). The derivative of $\psi(\mathbf{x}(c), \mathbf{x}(t))$ at $t = c+$ (the slope of the tangent line at the minimum of the *V*-shaped curve) is taken for the value of $F(\mathbf{x}(c), \dot{\mathbf{x}}(c))$, and the integral of this function from a to b is taken for the value of length of the path. The inset at the left top corner shows that one should consider the lengths for all such paths from **a** to **b**, and take their infimum as the (generally asymmetric) distance $G\mathbf{ab}$. The overall, symmetric distance $G^*\mathbf{ab}$ is computed as $G\mathbf{ab} + G\mathbf{ba}$. [The lengths of paths can be alternatively computed by differentiating $\psi(\mathbf{x}(t), \mathbf{x}(c))$ rather than $\psi(\mathbf{x}(c), \mathbf{x}(t))$. Although this generally changes the value of $G\mathbf{ab}$, it makes no difference for the value of $G^*\mathbf{ab}$.]

Any two points **a**, **b** in such a space can be connected by a *continuously differentiable path* $\mathbf{x}(t)$ defined on a segment of reals $[a, b]$. The "length" of this path can be defined by means of the following construction. Assume that

$$\psi(\mathbf{x}, \mathbf{x}) < \min\{\psi(\mathbf{x}, \mathbf{y}), \psi(\mathbf{y}, \mathbf{x})\}$$

for all distinct **x**, **y**, and that for any $c \in [a, b]$ the discrimination probability $\psi(\mathbf{x}(c), \mathbf{x}(t))$ has a positive right-hand derivative at $t = c+$,

$$\left.\frac{\mathrm{d}\psi\big(\mathbf{x}(c), \mathbf{x}(t)\big)}{\mathrm{d}t}\right|_{t=c+} = F\big(\mathbf{x}(c), \dot{\mathbf{x}}(c)\big).$$

The function $F(\mathbf{x}(t), \dot{\mathbf{x}}(t))$ is referred to as a *submetric function*, and the differential $F(\mathbf{x}(t), \dot{\mathbf{x}}(t))\mathrm{d}t$ serves as the local dissimilarity between $\mathbf{x}(t)$ and $\mathbf{x}(t) + \dot{\mathbf{x}}(t)\mathrm{d}t$. Assuming further that F is continuous, we define the length of the path $\mathbf{x}(t)$ as the integral

$$D\left(\mathbf{x}\left[a,b\right]\right)=\int_a^b F\left(\mathbf{x}(t),\dot{\mathbf{x}}(t)\right)\mathrm{d}t.$$

Applying this to all continuously differentiable paths connecting **a** to **b** and finding the *infimum* of their *D*-lengths, one defines the *(asymmetric) Fechnerian distance* G**ab** from **a** to **b** (a function which satisfies all metric axioms except for symmetry). The *overall (symmetrical) Fechnerian distance* G^***ab** between **a** and **b** is computed as G**ab** + G**ba**. Although this description is schematic and incomplete it should suffice for introducing one line of generalizing Fechnerian Scaling: dispensing with unidimensionality but retaining the idea of cumulation of local dissimilarities.

A further line of generalization is presented in Dzhafarov and Colonius (2005b, 2006c). It is designated as Fechnerian Scaling of Discrete Object Sets and applies to stimulus spaces comprised of "isolated entities," such as schematic faces, letters of an alphabet, and the like (see Figure 9.3). Each pair ($\mathbf{x}$, $\mathbf{y}$) of such stimuli is assigned a probability ψ ($\mathbf{x}$, $\mathbf{y}$) with which they are judged to be different from each other. Schematizing and simplifying as before, the local discriminability measure is defined as

$$D\,(\mathbf{x},\mathbf{y}) = \psi\,(\mathbf{x},\mathbf{y}) - \psi\,(\mathbf{x},\mathbf{x}),$$

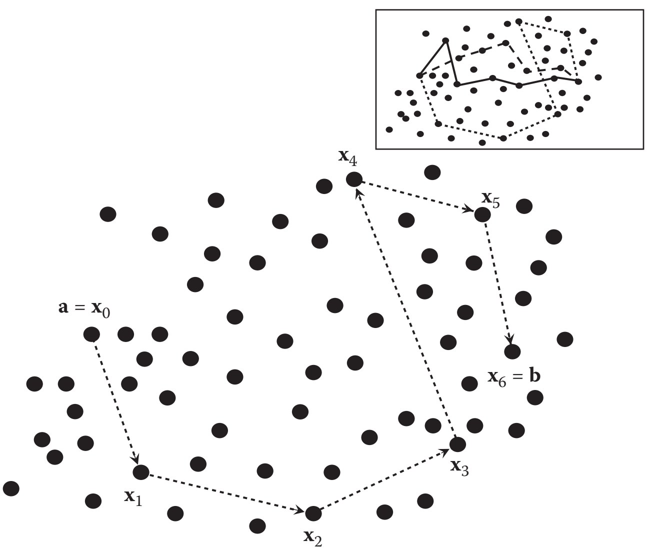

Figure 9.3 Given a chain of points $\mathbf{x}_0$, $\mathbf{x}_1$, …, $\mathbf{x}_k$ leading from **a** to **b**, the dissimilarities between its successive elements are summed (cumulated). In a discrete space, the (generally asymmetric) distance G**ab** from **a** to **b** is computed as the infimum of the cumulated dissimilarities over all chains leading from **a** to **b**. The symmetrical distance G^***ab** between **a** and **b** is computed as G**ab** + G**ba**.

and the (asymmetric) Fechnerian distance $G(\mathbf{a}, \mathbf{b})$ is defined as the infimum of

$$\sum_{i=0}^{k} D\left(\mathbf{x}_i, \mathbf{x}_{i+1}\right)$$

computed across all possible finite chains of stimuli

$$\mathbf{a} = \mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_k, \mathbf{x}_{k+1} = \mathbf{b}$$

connecting **a** to **b**. Here the deviation from Fechner's original theory is greater than in the Multidimensional Fechnerian Scaling: we dispense not only with unidimensionality, but also with the "infinitesimality" of dissimilarities being cumulated. But the idea of computing dissimilarities from discrimination probabilities and obtaining distances by some form of dissimilarity cumulation is retained.

The purpose of this work is to present a sweeping generalization of Fechner's theory which is applicable to *all possible* stimulus spaces endowed with "same-different" discrimination probabilities. This theory, called *Universal Fechnerian Scaling* (UFS), is presented in the trilogy of papers Dzhafarov and Colonius (2007), Dzhafarov (2008a), and Dzhafarov (2008b). We follow these papers closely, but omit proofs, examples, and technical explanations. Our focus is on the mathematical foundations of UFS, which are formed by an abstract theory called *Dissimilarity Cumulation* (DC): it provides a general definition of a *dissimilarity function* and shows how this function is used to impose on stimulus sets topological and metric properties.

The potential sphere of applicability of UFS is virtually unlimited. The ability to decide whether two entities are the same or different is the most basic faculty of all living organisms and the most basic requirement of artificial perceiving systems, such as intelligent robots. The perceiving system may be anything from an organism to a person to a group of consumers or voters to an abstract computational procedure. The stimuli may be anything from letters of alphabet (from the point of view of grammar school children) to different lung dysfunctions represented by X-ray films (from the point of view of a physician) to brands of a certain product (from the point of view of a group of consumers) to political candidates or propositions (from the point of view of potential voters) to competing statistical models (from the point of view of a statistical fitting procedure). Thus, if stimuli are several lung dysfunctions each represented by a potentially infinite set of X-ray films, a physician or a group of physicians can be asked to tell if two randomly chosen X-ray films do or do not indicate one and the same dysfunction. As a result each pair of dysfunctions is assigned the probability with which their respective X-ray representations are judged to indicate different ailments. If stimuli are competing statistical models, the probability with which models **x** and **y** are "judged" to be different can be estimated by the probability with which a dataset generated by the model **x** allows one to reject the model **y** (see Dzhafarov & Colonius, 2006a, for details). The questions to the perceiving system can be formulated in a variety of forms: "Are **x** and **y** the same (overall)?" or "Do **x** and **y** differ in respect to *A*?" or "Do **x** and **y** differ if one ignores their difference in property *B*?" or "Do **x** and **y** belong to one and the same category (from a given list)?", and so on. Note the difference from the other known scaling procedure based on discrimination probabilities, Thurstonian Scaling (Thurstone, 1927a,b). This procedure only deals with the probabilities with which one stimulus is judged to have more of a particular property (such as attractiveness, brightness, loudness, etc.) than another stimulus. The use of these probabilities therefore requires that the investigator know in advance which properties are relevant, that these properties be semantically unidimensional (i.e., assessable in terms of "greater–less"), and that the perception of the stimuli be entirely determined by these properties. No such assumptions are needed in UFS. Moreover, in the concluding section of the chapter it is mentioned that the discrimination probabilities may very well be replaced with other pairwise judgments of "subjective difference" between two stimuli, and that the theory can even be applied beyond the context of pairwise judgments altogether, for example, to categorization judgments. It is also mentioned there that the dissimilarity cumulation procedure can be viewed as an alternative to the nonmetric versions of Multidimensional Scaling, applying therefore in all cases in which one can use the latter.*

9.2 Psychophysics of discrimination

We observe the following notation conventions. Boldface lowercase letters, $\mathbf{a}$, $\mathbf{b}'$, $\mathbf{x}$, $\mathbf{y}_n$, …, always denote elements of a set of stimuli. Stimuli are merely names (qualitative entities), with no algebraic operations defined on them. Real-valued functions of one or more arguments that are elements of a stimulus set are indicated by strings without parentheses:

$$\psi\mathbf{ab},\ D\mathbf{abc}, D\mathbf{X}_n,\ \Psi^{(1)}\mathbf{ab},\ \ldots.$$

9.2.1 Regular Minimality and canonical representations

Here, we briefly recapitulate some of the basic concepts and assumptions underlying the theory of same–different discrimination probabilities. A toy example in Figure 9.4 provides an illustration. A detailed description and examples can be found in Dzhafarov (2002d, 2003a) and Dzhafarov and Colonius (2005a, 2006a).

The arguments $\mathbf{x}$ and $\mathbf{y}$ of the discrimination probability function

$$\psi^*\mathbf{xy} = \Pr\,[\mathbf{x} \text{ and } \mathbf{y} \text{ are judged to be different}]$$

* As a data-analytic procedure, UFS is implemented (as of September 2009) in three computer programs: the R-language package "fechner" described in Ünlü, Kiefer, and Dzhafarov (2009) and available at CRAN; a MATLAB-based program FSCAMDS developed at Purdue University and available at http://www1.psych.purdue.edu/~ehtibar/Links.html; and a MATLAB toolbox developed at Oldenburg University and available at http://www.psychologie.uni-oldenburg.de/stefan.rach/31856.html.

ψ^*	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_5$	$\mathbf{x}_6$	$\mathbf{x}_7$
$\mathbf{y}_1$	0.6	0.1	0.6	0.1	0.6	0.8	0.8
$\mathbf{y}_2$	0.9	0.8	0.9	0.8	0.9	0.1	0.1
$\mathbf{y}_3$	1	1	0.5	1	0.5	0.6	0.6
$\mathbf{y}_4$	0.5	1	0.7	1	0.7	1	1
$\mathbf{y}_5$	0.5	1	0.7	1	0.7	1	1
$\mathbf{y}_6$	0.5	1	0.7	1	0.7	1	1
$\mathbf{y}_7$	0.5	1	0.7	1	0.7	1	1

$\tilde{\psi}$	$\mathbf{x}_a$	$\mathbf{x}_b$	$\mathbf{x}_c$	$\mathbf{x}_d$
$\mathbf{y}_a$	0.6	0.6	0.1	0.8
$\mathbf{y}_b$	0.9	0.9	0.8	0.1
$\mathbf{y}_c$	1	0.5	1	0.6
$\mathbf{y}_d$	0.5	0.7	1	1

Observation area 1, $\mathfrak{S}_1$	$\mathbf{x}_c$	$\mathbf{x}_d$	$\mathbf{x}_b$	$\mathbf{x}_a$
Observation area 2, $\mathfrak{S}_2$	$\mathbf{y}_a$	$\mathbf{y}_b$	$\mathbf{y}_c$	$\mathbf{y}_d$
Common labels, $\mathfrak{S}$	**a**	**b**	**c**	**d**

ψ	**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

Figure 9.4 A toy example used in Dzhafarov & Colonius (2006a). The transformation from ($\mathfrak{S}_1^*$, $\mathfrak{S}_2^*$,ψ^*) to ($\mathfrak{S}_1$, $\mathfrak{S}_2$, $\tilde{\psi}$) is the result of "lumping together" psychologically equal stimuli (e.g., the stimuli $\mathbf{y}_4$, $\mathbf{y}_5$, $\mathbf{y}_6$, $\mathbf{y}_7$ are psychologically equal in $\mathfrak{S}_2^*$, stimuli $\mathbf{x}_2$ and $\mathbf{x}_4$ are psychologically equal in $\mathfrak{S}_1^*$). The space ($\mathfrak{S}_1$, $\mathfrak{S}_2$, $\tilde{\psi}$) satisfies the Regular Minimality condition (the minimum in each row is also the minimum in its column) because of which ($\mathfrak{S}_1$, $\mathfrak{S}_2$, $\tilde{\psi}$) can be canonically transformed into ($\mathfrak{S}$,ψ), by means of the transformation table shown in between.

belong to two distinct observation areas,

$$\psi^*: \mathfrak{S}_1^* \times \mathfrak{S}_2^* \to [0, 1].$$

Thus, $\mathfrak{S}_1^*$ (the first observation area) may represent stimuli presented chronologically first or on the left, whereas $\mathfrak{S}_2^*$ (the second observation area) designates stimuli presented, respectively, chronologically second or on the right. The adjectives "first" and "second" refer to the ordinal positions of stimulus symbols within a pair ($\mathbf{x}$, $\mathbf{y}$).

For $\mathbf{x},\mathbf{x}' \in \mathfrak{S}_1^*$, we say that the two stimuli are *psychologically equal* (or metameric) if $\psi^*\mathbf{xy} = \psi^*\mathbf{x}'\mathbf{y}$ for any $\mathbf{y} \in \mathfrak{S}_2^*$. Analogously, the psychological equality for $\mathbf{y}$, $\mathbf{y}' \in \mathfrak{S}_2^*$ is defined by $\psi^*\mathbf{xy} = \psi^*\mathbf{xy}'$, for any $\mathbf{x} \in \mathfrak{S}_1^*$. It is always possible to "*reduce*" the observation areas, that is, relabel their elements so that psychologically equal stimuli receive identical labels and are no longer distinguished. The discrimination probability function ψ^* can then be redefined as

$$\tilde{\psi} : \mathfrak{S}_1 \times \mathfrak{S}_2 \to [0, 1].$$

The law of Regular Minimality is the statement that there are functions $\mathbf{h} : \mathfrak{S}_1 \to \mathfrak{S}_2$ and $\mathbf{g} : \mathfrak{S}_2 \to \mathfrak{S}_1$ such that

$$(P_1)\quad \tilde{\psi}\mathbf{x}[\mathbf{h}(\mathbf{x})] < \tilde{\psi}\mathbf{xy} \text{ for all } \mathbf{y} \neq \mathbf{h}(\mathbf{x})$$

$$(P_2)\quad \tilde{\psi}[\mathbf{g}(\mathbf{y})]\mathbf{y} < \tilde{\psi}\mathbf{xy} \text{ for all } \mathbf{x} \neq \mathbf{g}(\mathbf{y})$$

$$(P_3)\quad \mathbf{h} \equiv \mathbf{g}^{-1}$$

Stimulus $\mathbf{y} = \mathbf{h}(\mathbf{x}) \in \mathfrak{S}_2$ is called the *Point of Subjective Equality* (PSE) for $\mathbf{x} \in \mathfrak{S}_1$; analogously, $\mathbf{x} = \mathbf{g}(\mathbf{y}) \in \mathfrak{S}_1$ is the PSE for $\mathbf{y} \in \mathfrak{S}_2$. The law of Regular Minimality states therefore that every stimulus in each of the (reduced) observation areas has a unique PSE in the other observation area, and that $\mathbf{y}$ is the PSE for $\mathbf{x}$ if and only if $\mathbf{x}$ is the PSE for $\mathbf{y}$. In some contexts the law of regular minimality is an empirical assumption, but it can also serve as a criterion for a properly defined stimulus space. For a detailed discussion of the law and its critiques see Dzhafarov (2002d, 2003a, 2006), Dzhafarov and Colonius (2006a), and Ennis (2006).

Due to the law of Regular Minimality, one can always relabel the stimuli in $\mathfrak{S}_1$ or $\mathfrak{S}_2$ so that any two mutual PSEs receive one and the same label. In other words, one can always bijectively map $\mathfrak{S}_1 \to \mathfrak{S}$ and $\mathfrak{S}_2 \to \mathfrak{S}$ so that $\mathbf{x} \mapsto \mathbf{a}$ and $\mathbf{y} \mapsto \mathbf{a}$ if and only if $\mathbf{x} \in \mathfrak{S}_1$ and $\mathbf{y} \in \mathfrak{S}_2$ are mutual PSEs: $\mathbf{y} = \mathbf{h}(\mathbf{x})$, $\mathbf{x} = \mathbf{g}(\mathbf{y})$. The set of labels $\mathfrak{S}$ is called a *canonically transformed* stimulus set. Its elements too, for simplicity, are referred to as stimuli. The discrimination probability function $\tilde{\psi}$ can now be presented in a *canonical form*,

$$\psi : \mathfrak{S} \times \mathfrak{S} \to [0, 1],$$

with the property

$$\psi\mathbf{aa} < \min\{\psi\mathbf{ab}, \psi\mathbf{ba}\}$$

for any $\mathbf{a}$ and $\mathbf{b} \neq \mathbf{a}$. Note that the first and the second $\mathbf{a}$ in $\psi\mathbf{aa}$ may very well refer to physically different stimuli (equivalence classes of stimuli): hence one should exercise caution in referring to $\psi\mathbf{aa}$ as the probability with which $\mathbf{a}$ is discriminated from "itself."

9.2.2. From discrimination to dissimilarity

For the canonically transformed function ψ, the *psychometric increments of the first and second kind* are defined as, respectively,

$$\Psi^{(1)}\mathbf{ab} = \psi\mathbf{ab} - \psi\mathbf{aa}$$

and

$$\Psi^{(2)}\mathbf{ab} = \psi\mathbf{ba} - \psi\mathbf{aa}.$$

Due to the canonical form of ψ these quantities are always positive for $\mathbf{b} \neq \mathbf{a}$.

The main assumption of UFS about these psychometric increments is that both of them are dissimilarity functions. The meaning of this statement becomes clear later, after a formal definition of a dissimilarity function is given.

Denoting by D either $\Psi^{(1)}$ or $\Psi^{(2)}$ one can compute the (generally asymmetric) Fechnerian distance $G\mathbf{ab}$ by considering all possible finite chains of stimuli $\mathbf{x}_1 \ldots \mathbf{x}_k$ for all possible k and putting

$$G\mathbf{ab} = \inf_{k, x_1 \ldots x_k} \left[D\mathbf{a}\mathbf{x}_1 + D\mathbf{x}_1\mathbf{x}_2 + \ldots + D\mathbf{x}_k\mathbf{b} \right].$$

The overall Fechnerian distance is then computed as

$$G^*\mathbf{ab} = G_1\mathbf{ab} + G_1\mathbf{ba}.$$

This quantity can be interpreted as the infimum of D-lengths of all finite closed loops that contain points $\mathbf{a}$ and $\mathbf{b}$. That is,

$$G^*\mathbf{ab} = \inf_{\substack{k, \mathbf{x}_1 \ldots \mathbf{x}_k \\ l, \mathbf{y}_1 \ldots \mathbf{y}_l}} \left[D\mathbf{a}\mathbf{x}_1 + D\mathbf{x}_1\mathbf{x}_2 + \ldots + D\mathbf{x}_k\mathbf{b} + D\mathbf{b}\mathbf{y}_1 + D\mathbf{y}_1\mathbf{y}_2 + \ldots + D\mathbf{y}_l\mathbf{a} \right]$$

It is easy to see that the D-length of any given loop remains invariant if $D \equiv \Psi^{(1)}$ is replaced with $D \equiv \Psi^{(2)}$ and the loop is traversed in the opposite direction. The value of $G^*\mathbf{ab}$ therefore does not depend on which of the two psychometric increments is taken for D. Henceforth we tacitly assume that D may be replaced with either $\Psi^{(1)}$ or $\Psi^{(2)}$, no matter which.

9.3 Dissimilarity Cumulation theory

9.3.1 Topology and uniformity

To explain what it means for a function $D : \mathfrak{S} \times \mathfrak{S} \to \mathbb{R}$ to be a dissimilarity function, we begin with a more general concept. Function $D : \mathfrak{S} \times \mathfrak{S} \to \mathbb{R}$ is a (uniform) *deviation function* if it has the following properties: for any $\mathbf{a}, \mathbf{b} \in \mathfrak{S}$ and any sequences $\mathbf{a}_n$, $\mathbf{a}'_n$, $\mathbf{b}_n$, $\mathbf{b}'_n$ in $\mathfrak{S}$,

$[D1.]$ $\mathbf{a} \neq \mathbf{b} \Rightarrow D\mathbf{ab} > 0$;

$[D2.]$ $D\mathbf{aa} = 0$;

$[D3.]$ (Uniform Continuity) If $D\mathbf{a}_n\mathbf{a}'_n \to 0$ and $D\mathbf{b}_n\mathbf{b}'_n \to 0,$ then $D\mathbf{a}'_n\mathbf{b}'_n - D\mathbf{a}_n\mathbf{b}_n \to 0.$

See Figure 9.5 for an illustration of Property $D3$. If D is a symmetric metric, then it is a deviation function, with the uniform continuity property holding as a theorem.

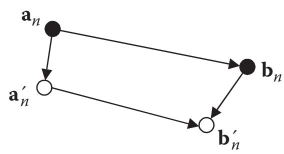

Figure 9.5 An illustration for property $D3$ (uniform continuity). Consider an infinite sequence of quadrilaterals $\mathbf{a}_1\mathbf{a}'_1\mathbf{b}'_1\mathbf{b}_1$, $\mathbf{a}_2\mathbf{a}'_2\mathbf{b}'_2\mathbf{b}_2$, …, such that the D-lengths of the sides $\mathbf{a}_n\mathbf{a}'_n$ and $\mathbf{b}_n\mathbf{b}'_n$ (directed as shown by the arrows) converge to zero. Then the difference between the D-lengths of the sides $\mathbf{a}_n\mathbf{b}_n$ and $\mathbf{a}'_n\mathbf{b}'_n$ (in the direction of the arrows) converges to zero.

If D is an asymmetric metric, then it is a deviation function if and only if it additionally satisfies the "invertibility in the small" condition,

$$D\mathbf{a}_n\mathbf{a}'_n \to 0 \Rightarrow D\mathbf{a}'_n\mathbf{a}_n \to 0.$$

In the following the term *metric* (or *distance*), unless specifically qualified as symmetric, always refers to an asymmetric metric (distance) invertible in the small.

D induces on $\mathfrak{S}$ the notion of convergence: we define $\mathbf{a}_n \leftrightarrow \mathbf{b}_n$ to mean $D\mathbf{a}_n\mathbf{b}_n \to 0$. The notation is unambiguous because convergence $\leftrightarrow$ is an equivalence relation (i.e.,it is reflexive, symmetric, and transitive). In particular, $\mathbf{a}_n \leftrightarrow \mathbf{a}$ means both $D\mathbf{a}\mathbf{a}_n$ 0 and $D\mathbf{a}_n\mathbf{a} \to 0$. The convergence $(\mathbf{a}_n^1, \ldots, \mathbf{a}_n^k) \leftrightarrow (\mathbf{b}_n^1, \ldots, \mathbf{b}_n^k)$ can be defined, e.g., by $\max_i D\mathbf{a}_n^i\mathbf{b}_n^i \to 0$.

A *topological basis* on $\mathfrak{S}$ is a family of subsets of $\mathfrak{S}$ covering $\mathfrak{S}$ and satisfying the following property (Kelly, 1955, p. 47): if $\mathfrak{a}$ and $\mathfrak{b}$ are within the basis, then for any $\mathbf{x} \in \mathfrak{a} \cap \mathfrak{b}$ the basis contains a set $\mathfrak{c}$ that contains $\mathbf{x}$. Given a topological basis on $\mathfrak{S}$, the *topology* on $\mathfrak{S}$ (a family of *open sets* "based" on this basis) is obtained by taking all possible unions of the subsets comprising the basis (including the empty set, which is the union of an empty class of such subsets). Deviation D induces on $\mathfrak{S}$ a topology based on

$$\mathfrak{B}_D(\mathbf{x},\varepsilon) = \{\mathbf{y} \in \mathfrak{S} : D\mathbf{x}\mathbf{y} < \varepsilon\}$$

taken for all $\mathbf{x} \in \mathfrak{S}$ and all real $\varepsilon > 0$. We call this topology (based on $\mathfrak{B}_D$-balls) the *D-topology.*

These topological considerations, as it turns out, can be strengthened: D induces on $\mathfrak{S}$ not only a topology but a more restrictive structure, called *uniformity*. Recall (Kelly, 1955, p. 177) that a family of subsets of $\mathfrak{S} \times \mathfrak{S}$ forms a *basis for a uniformity* on $\mathfrak{S}$ if it satisfies the following four properties: if $\mathfrak{A}$ and $\mathfrak{B}$ are members of the basis, then

1. $\mathfrak{A}$ includes as its subset $\Delta = \{(\mathbf{x}, \mathbf{x}) : \mathbf{x} \in \mathfrak{S}\}$.
2. $\mathfrak{A}^{-1} = \{(\mathbf{y}, \mathbf{x}) : (\mathbf{x}, \mathbf{y}) \in \mathfrak{A}\}$ includes as its subset a member of the basis.

3. For some member $\mathfrak{C}$ of the basis, $\{(\mathbf{x}, \mathbf{z}) \in \mathfrak{S}^2 : \text{for some } \mathbf{y}, (\mathbf{x}, \mathbf{y}) \in \mathfrak{C} \wedge (\mathbf{y}, \mathbf{z}) \in \mathfrak{C}\} \subset \mathfrak{A}$.
4. $\mathfrak{A} \cap \mathfrak{B}$ includes as its subset a member of the basis.

Given a uniformity basis on $\mathfrak{S}$, the uniformity on $\mathfrak{S}$ ("based" on this basis) is obtained by taking each member of the basis and forming its unions with all subsets of $\mathfrak{S} \times \mathfrak{S}$.A member of a uniformity is called an *entourage*. Deviation D induces on $\mathfrak{S}$ a uniformity based on entourages

$$\mathfrak{U}_D(\varepsilon) = \{(\mathbf{x}, \mathbf{y}) \in \mathfrak{S}^2 : D\mathbf{xy} < \varepsilon\}$$

taken for all real $\varepsilon > 0$. This uniformity satisfies the so-called separation axiom:

$$\cap_\varepsilon \mathfrak{U}_D(\varepsilon) = \{(\mathbf{x}, \mathbf{y}) \in \mathfrak{S}^2 : \mathbf{x} = \mathbf{y}\}.$$

We call this uniformity the *D-uniformity*. The D-topology is precisely the topology induced by the D-uniformity (Kelly, 1955, p. 178):

$$\mathfrak{B}_D(\mathbf{x},\varepsilon) = \{\mathbf{y} \in \mathfrak{S} : (\mathbf{x}, \mathbf{y}) \in \mathfrak{U}_D(\varepsilon)\}$$

is the restriction of the basic entourage $\mathfrak{U}_D(\varepsilon)$ to the pairs ($\mathbf{x}$ = const, $\mathbf{y}$).

9.3.2 Chains and dissimilarity function

Chains in space $\mathfrak{S}$ are finite sequences of elements, written as strings: **ab**, **abc**, $\mathbf{x}_1 \ldots \mathbf{x}_k$, etc. Note that the elements of a chain need not be pairwise distinct. A chain of *cardinality* k (a k-chain) is the chain with k elements (vertices), hence with $k - 1$ links (edges). For completeness, we also admit an empty chain, of zero cardinality. We use the notation

$$D\mathbf{x}_1 \ldots \mathbf{x}_k = \sum_{i=1}^{k-1} D\mathbf{x}_i\mathbf{x}_{i+1},$$

and call it the D-length of the chain $\mathbf{x}_1 \ldots \mathbf{x}_k$.

If the elements of a chain are not of interest, it can be denoted by a boldface capital, such as $\mathbf{X}$, with appropriate ornaments. Thus, $\mathbf{X}$ and $\mathbf{Y}$ are two chains, $\mathbf{XY}$ is their concatenation, $\mathbf{aXb}$ is a chain connecting $\mathbf{a}$ to $\mathbf{b}$. The cardinality of chain $\mathbf{X}$ is denoted $|\mathbf{X}|$. Unless otherwise specified, within a sequence of chains, $\mathbf{X}_n$, the cardinality $|\mathbf{X}_n|$ generally varies: $\mathbf{X}_n = \mathbf{x}_1^n \ldots \mathbf{x}_{k_n}^n$.

A uniform deviation function D on $\mathfrak{S}$ is a uniform dissimilarity (or, simply, dissimilarity) function on $\mathfrak{S}$ if it has the following property:

[D4.] for any sequence of chains $\mathbf{a}_n\mathbf{X}_n\mathbf{b}_n$,

$$D\mathbf{a}_n\mathbf{X}_n\mathbf{b}_n \to 0 \Rightarrow D\mathbf{a}_n\mathbf{b}_n \to 0.$$

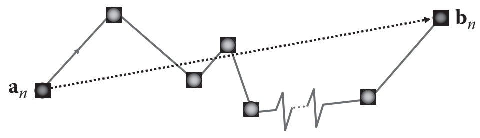

Figure 9.6 An Illustration for Property $D4$ (chain property). Consider an infinite sequence of chains $\mathbf{a}_1\mathbf{X}_1\mathbf{b}_1, \mathbf{a}_2\mathbf{X}_2\mathbf{b}_2, \ldots$, such that $|\mathbf{X}_n|$ increases beyond bounds with $n \to \infty$, and $D\mathbf{a}_n\mathbf{X}_n\mathbf{b}_n$ converges to zero. Then $D\mathbf{a}_n\mathbf{b}_n$ (the D-length of the dotted arrow) converges to zero too.

See Figure 9.6 for an illustration. If D is a metric, then D is a dissimilarity function as a trivial consequence of the triangle inequality.

9.3.3 Fechnerian distance

The set of all possible chains in $\mathfrak{S}$ is denoted by $C_{\mathfrak{S}}$, or simply C. We define function $G\mathbf{ab}$ by

$$G\mathbf{ab} = \inf_{\mathbf{X}\in C} D\mathbf{aXb}.$$

$G\mathbf{ab}$ is a metric, and $G^*\mathbf{ab} = G\mathbf{ab} + G\mathbf{ba}$ is a symmetric metric (also called "overall"). We say that the metric G and the overall metric G^* are induced by the dissimilarity D. Clearly, $G^*\mathbf{ab}$ can also be defined by

$$G^*\mathbf{ab} = \inf_{(\mathbf{X},\mathbf{Y})\in C^2} D\mathbf{aXbYa} = \inf_{(\mathbf{X},\mathbf{Y})\in C^2} D\mathbf{bXaYb}.$$

9.3.4 Topology and uniformity on ($\mathfrak{S}$,G)

It can be shown that

$$D\mathbf{a}_n\mathbf{b}_n \to 0 \Leftrightarrow G\mathbf{a}_n\mathbf{b}_n \to 0,$$

and

$$\mathbf{a}_n \leftrightarrow \mathbf{b}_n \Leftrightarrow G\mathbf{a}_n\mathbf{b}_n \to 0 \Leftrightarrow G\mathbf{b}_n\mathbf{a}_n \to 0 \Leftrightarrow G^*\mathbf{a}_n\mathbf{b}_n = G^*\mathbf{b}_n\mathbf{a}_n \to 0.$$

As a consequence, G induces on $\mathfrak{S}$ a topology based on sets

$$\mathfrak{B}_G(\mathbf{x},\varepsilon) = \{\mathbf{y} \in \mathfrak{S} : G\mathbf{xy} < \varepsilon\}$$

taken for all $\mathbf{x} \in \mathfrak{S}$ and positive ε. This topology coincides with the D-topology. Analogously, G induces on $\mathfrak{S}$ a uniformity based on the sets

$$\mathfrak{U}_G(\varepsilon) = \{(\mathbf{x}, \mathbf{y}) \in \mathfrak{S}^2 : G\mathbf{xy} < \varepsilon\}$$

taken for all positive ε. This uniformity coincides with the D-uniformity. The metric G is uniformly continuous in $(\mathbf{x}, \mathbf{y})$, i.e., if $\mathbf{a}'_n \leftrightarrow \mathbf{a}_n$ and $\mathbf{b}'_n \leftrightarrow \mathbf{b}_n$, then

$$G\mathbf{a}'_n\mathbf{b}'_n - G\mathbf{a}_n\mathbf{b}_n \to 0.$$

The space $(\mathfrak{S},D)$ being uniform and metrizable, we get its standard topological characterization (see, e.g., Hocking & Young, 1961, p. 42): it is a *completely normal space*, meaning that its singletons are closed and any its two separated subsets $\mathfrak{A}$ and $\mathfrak{B}$ (i.e., such that $\overline{\mathfrak{A}} \cap \mathfrak{B} = \mathfrak{A} \cap \overline{\mathfrak{B}} = \varnothing$) are contained in two disjoint open subsets.

The following is an important fact which can be interpreted as that of internal consistency of the metric G induced by means of dissimilarity cumulation: once $G\mathbf{ab}$ is computed as the infimum of the D-length across all chains from **a** to **b**, the infimum of the G-length across all chains from **a** to **b** equals $G\mathbf{ab}$:

$$D\mathbf{a}\mathbf{X}_n\mathbf{b} \to G\mathbf{ab} \Rightarrow G\mathbf{a}\mathbf{X}_n\mathbf{b} \to G\mathbf{ab},$$

where we use the notation for cumulated G-length analogous to that for D-length,

$$G\mathbf{x}_1\ldots\mathbf{x}_k = \sum_{i=1}^{k-1} G\mathbf{x}_i\mathbf{x}_{i+1}.$$

Extending the traditional usage of the term, one can say that G is an *intrinsic metric*. This is an extension because traditionally the notion of intrinsic metric presupposes the existence of paths (continuous images of segments of reals) and their lengths. In subsequent sections we consider special cases of dissimilarity cumulation in which the intrinsicality of G does acquire its traditional meaning.

9.4 Dissimilarity Cumulation in arc-connected spaces

9.4.1 Path and their lengths

Because the notion of uniform convergence in the space $(\mathfrak{S},D)$ is well-defined,

$$\mathbf{a}_n \leftrightarrow \mathbf{b}_n \Leftrightarrow D\mathbf{a}_n\mathbf{b}_n \to 0,$$

we can meaningfully speak of continuous and uniformly continuous functions from reals into $\mathfrak{S}$.

Let $\mathbf{f} : [a, b] \to \mathfrak{S}$, or $\mathbf{f}\,|\,[a, b]$, be some continuous (hence uniformly continuous) function with $\mathbf{f}(a) = \mathbf{a}$, $\mathbf{f}(b) = \mathbf{b}$, where **a** and **b** are not necessarily distinct. We call such a function a *path* connecting **a** to **b**. A space is called *arc-connected* (or *path-connected*) if any two points in it can be connected by a path. Even though arcs have not yet been introduced, the terms "arc-connected" and "path-connected" are synonymous, because $(\mathfrak{S},D)$ is a Hausdorff space, so if two points in it are connected by a path they are also connected by an arc (see, e.g., Hocking & Young, 1961, pp. 116–117). Hereafter we assume that $(\mathfrak{S},D)$ is an arc-connected space.

Choose an arbitrary *net* on $[a, b]$,

$$\mu = (a = x_0 \le x_1 \le \ldots \le x_k \le x_{k+1} = b),$$

where the x_i's need not be pairwise distinct. We call the quantity

$$\delta\mu = \max_{i=0,1,\ldots,k} \left(x_{i+1} - x_i\right)$$

the net's *mesh*. As $\delta\mu_n \to 0$, nets μ_n provide a progressively better approximation for $[a, b]$.

Given a net $\mu = (x_0, x_1, \ldots, x_k, x_{k+1})$, any chain $\mathbf{X} = \mathbf{x}_0\mathbf{x}_1 \ldots \mathbf{x}_k\mathbf{x}_{k+1}$ (with the elements not necessarily pairwise distinct, and $\mathbf{x}_0$ and $\mathbf{x}_{k+1}$ not necessarily equal to $\mathbf{a}$ and $\mathbf{b}$) can be used to form a *chain-on-net*

$$\mathbf{X}^\mu = ((x_0, \mathbf{x}_0), (x_1, \mathbf{x}_1), \ldots, (x_k, \mathbf{x}_k), (x_{k+1}, \mathbf{x}_{k+1})).$$

Denote the class of all such chains-on-nets $\mathbf{X}^\mu$ (for all possible pairs of a chain $\mathbf{X}$ and a net μ of the same cardinality) by $\mathcal{M}_a^b$. Note that a chain-on-net is not a function from $\{x : x \text{ is an element of } \mu\}$ into $\mathfrak{S}$, for it may include pairs $(x_i = x, \mathbf{x}_i)$ and $(x_j = x, \mathbf{x}_j)$ with $\mathbf{x}_i \neq \mathbf{x}_j$. Note also that within a given context $\mathbf{X}^\mu$ and $\mathbf{X}^\nu$ denote one and the same chain on two nets, whereas $\mathbf{X}^\mu$, $\mathbf{Y}^\mu$ denote two chains on the same net.

We define the *separation* of the chain-on-net $\mathbf{X}^\mu = ((x_0, \mathbf{x}_0), \ldots, (x_{k+1}, \mathbf{x}_{k+1})) \in \mathcal{M}_a^b$ from a path $\mathbf{f}|[a, b]$ as

$$\sigma_\mathbf{f}\left(\mathbf{X}^\mu\right) = \max_{x_i \in \mu} D\mathbf{f}\left(x_i\right)\mathbf{x}_i.$$

For a sequence of paths $\mathbf{f}_n|[a, b]$, any sequence of chains-on-nets $\mathbf{X}_n^{\mu_n} \in \mathcal{M}_a^b$ with $\delta\mu_n \to 0$ and $\sigma_{\mathbf{f}_n}(\mathbf{X}_n^{\mu_n}) \to 0$ is referred to as a sequence *converging with* $\mathbf{f}_n$. We denote such convergence by $\mathbf{X}_n^{\mu_n} \to \mathbf{f}_n$. In particular, $\mathbf{X}_n^{\mu_n} \to \mathbf{f}$ for a fixed path $\mathbf{f}|[a, b]$ means that $\delta\mu_n \to 0$ and $\sigma_\mathbf{f}(\mathbf{X}_n^{\mu_n}) \to 0$: in this case we can say that $\mathbf{X}_n^{\mu_n}$ converges *to* $\mathbf{f}$. See Figure 9.7 for an illustration.

We define the *D-length* of $\mathbf{f}|[a, b]$ as

$$D\mathbf{f} = \liminf_{\mathbf{X}^\mu \to \mathbf{f}} D\mathbf{X} = \liminf_{\substack{\delta\mu \to 0 \\ \sigma_\mathbf{f}\left(\mathbf{X}^\mu\right) \to 0}} D\mathbf{X},$$

where all $\mathbf{X}^\mu \in \mathcal{M}_a^b$.

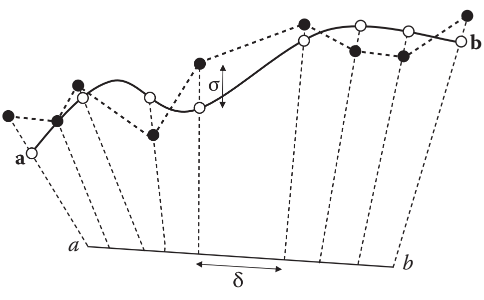

Figure 9.7 A chain-on-net $\mathbf{X}^\mu$ is converging to a path $\mathbf{f}$ as $\sigma = \sigma_\mathbf{f}(\mathbf{X}^\mu) \to 0$ and $\delta = \delta\mu \to 0$.

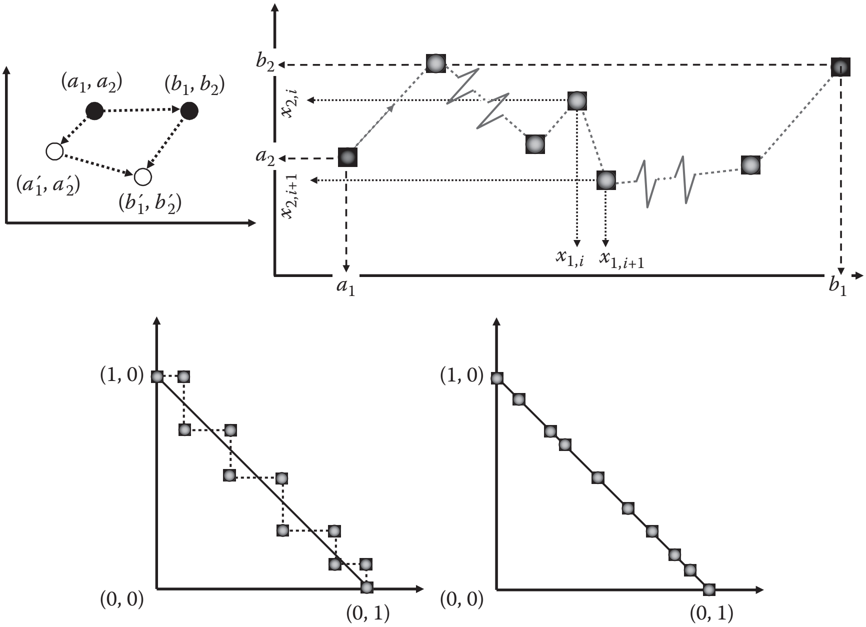

Figure 9.8 A demonstration of the fact that inscribed chains are not suffcient for D-length computations. The D from (a_1,a_2) to (b_1,b_2) is defined as $|a_1 - b_1| + |a_2 - b_2| + \min\{|a_1 - b_1|, |a_2 - b_2|\}$. It is a dissimilarity function, as illustrated in the top panels. Bottom left: the staircase chain has the cumulated dissimilarity 2, and 2 is the true D-length of the hypotenuse. Bottom right: the inscribed chain has the cumulated dissimilarity 3.

Given a path $\mathbf{f}|\,[a, b]$, the class of the chains-on-nets $\mathbf{X}^\mu$ such that $\delta\mu < \delta$ and $\sigma_\mathbf{f}(\mathbf{X}^\mu) < \varepsilon$ is nonempty for all positive δ and ε, because this class includes appropriately chosen *inscribed* chains-on-nets

$$((a, \mathbf{a}), (x_1, \mathbf{f}(x_1)), \ldots, (x_k, \mathbf{f}(x_k)), (b, \mathbf{b})).$$

Here, obviously, $\sigma_\mathbf{f}(\mathbf{X}^\mu)$ is identically zero. Note, however, that with our definition of D-length one generally cannot confine one's consideration to the inscribed chains-on-nets only (see Figure 9.8).

Let us consider some basic properties of paths. For any path $\mathbf{f}|\,[a, b]$ connecting $\mathbf{a}$ to $\mathbf{b}$,

$$D\mathbf{f} \geq G\mathbf{ab}.$$

That is, the D-length of a path is bounded from below by $G\mathbf{ab}$. There is no upper bound for $D\mathbf{f}$; this quantity need not be finite. Thus, it is shown below that when D is a metric, the notion of $D\mathbf{f}$ coincides with the traditional notion of path length; and examples of paths whose length, in the traditional sense, is infinite, are well-known (see, e.g., Chapter 1 in Papadopoulos, 2005). We call a path *D-rectifiable* if its D-length is finite.

We next note the additivity property for path length. For any path $\mathbf{f}|[a, b]$ and any point $z \in [a, b]$,

$$D\mathbf{f}|[a, b] = D\mathbf{f}|[a, z] + D\mathbf{f}[z, b].$$

$D\mathbf{f}$ for any path $\mathbf{f}|[a, b]$ is nonnegative, and $D\mathbf{f} = 0$ if and only if $\mathbf{f}$ is constant (i.e., $\mathbf{f}([a, b])$ is a singleton).

The quantity

$$\sigma_{\mathbf{f}}(\mathbf{g}) = \max_{x \in [a,b]} D\mathbf{f}(x)\mathbf{g}(x)$$

is called the *separation* of path $\mathbf{g}|[a, b]$ from path $\mathbf{f}|[a, b]$. Two sequences of paths $\mathbf{f}_n$ and $\mathbf{g}_n$ are said to be (*uniformly*) *converging* to each other if $\sigma_{\mathbf{f}_n}(\mathbf{g}_n) \to 0$. Due to the symmetry of the convergence in $\mathfrak{S}$, this implies $\sigma_{\mathbf{g}_n}(\mathbf{f}_n) \to 0$, so the definition and terminology are well-formed. We symbolize this by $\mathbf{f}_n \to \mathbf{g}_n$. In particular, if $\mathbf{f}$ is fixed then a sequence $\mathbf{f}_n$ converges to $\mathbf{f}$ if $\sigma_{\mathbf{f}}(\mathbf{f}_n) \to 0$. We present this convergence as $\mathbf{f}_n \to \mathbf{f}$. Note that if $\mathbf{f}_n \to \mathbf{f}$, the endpoints $\mathbf{a}_n = \mathbf{f}_n(a)$ and $\mathbf{b}_n = \mathbf{f}_n(b)$ generally depend on n and differ from, respectively $\mathbf{a} = \mathbf{f}(a)$ and $\mathbf{b} = \mathbf{f}(b)$.

The following very important property is called the *lower semicontinuity* of D-length (as a function of paths). For any sequence of paths $\mathbf{f}_n \to \mathbf{f}$,

$$\liminf_{n \to \infty} D\mathbf{f}_n \geq D\mathbf{f}.$$

9.4.2 G-lengths

Because the metric G induced by D in accordance with

$$G\mathbf{ab} = \inf_{\mathbf{X}} D\mathbf{aXb}$$

is itself a dissimilarity function, the *G-length* of a path $\mathbf{f} : [a, b] \to \mathfrak{S}$ should be defined as

$$G\mathbf{f} = \liminf_{\substack{\mathbf{X}^\mu \in \mathcal{M}_a^b \\ \mathbf{X}^\mu \xrightarrow{G} \mathbf{f}}} G\mathbf{X},$$

where (putting $\mathbf{X} = \mathbf{x}_0\mathbf{x}_1 \ldots \mathbf{x}_k\mathbf{x}_{k+1}$),

$$G\mathbf{X} = \sum_{i=0}^{k} G\mathbf{x}_i\mathbf{x}_{i+1},$$

and the convergence $\mathbf{X}^\mu \xrightarrow{G} \mathbf{f}$ (where μ is the net $a = x_0, x_1, \ldots, x_k, x_{k+1} = b$ corresponding to $\mathbf{X}$) means the conjunction of $\delta\mu \to 0$ and

$$\sigma^*_{\mathbf{f}}\left(\mathbf{X}^{\mu}\right) = \max_{i=0,\ldots,k+1} G\mathbf{f}\left(x_i\right)\mathbf{x}_i \to 0\,.$$

It is easy to see, however, that $\mathbf{X}^{\mu} \xrightarrow{G} \mathbf{f}$ and $\mathbf{X}^{\mu} \to \mathbf{f}$ are interchangeable:

$$\mathbf{X}^{\mu} \to \mathbf{f} \Leftrightarrow \mathbf{X}^{\mu} \xrightarrow{G} \mathbf{f}\,.$$

Because G is a metric, we also have, by a trivial extension of the classical theory (e.g., Blumenthal, 1953),

$$G\mathbf{f} = \sup G\mathbf{Z}$$

with the supremum taken over all inscribed chains-on-nets $\mathbf{Z}^{\nu}$; moreover,

$$G\mathbf{f} = \lim_{n\to\infty} G\mathbf{Z}_n$$

for any sequence of inscribed chains-on-nets $\mathbf{Z}_n^{\nu_n}$ with $\delta\nu_n \to 0$.

As it turns out, these traditional definitions are equivalent to our definition of G-length. Moreover the D-length and G-length of a path are always equal: for any path $\mathbf{f}$,

$$D\mathbf{f} = G\mathbf{f}.$$

9.4.3 Other properties of D-length for paths and arcs

The properties established in this section parallel the basic properties of path length in the traditional, metric-based theory (Blumenthal, 1953; Blumenthal & Menger, 1970; Busemann, 2005). We note first the uniform continuity of length traversed along a path: for any D-rectifiable path $\mathbf{f}|\,[a, b]$ and $[x, y] \subset [a, b]$, $D\mathbf{f}|\,[x, y]$ is uniformly continuous in (x, y), nondecreasing in y and nonincreasing in x (see Figure 9.9).

The next issue we consider is the (in)dependence of the D-length of a path on the path's parametrization. The D-length of a path is not determined by its image $\mathbf{f}([a, b])$ alone but by the function $\mathbf{f} : [a, b] \to \mathfrak{S}$. Nevertheless two paths $\mathbf{f}|\,[a, b]$ and $\mathbf{g}|\,[c, d]$ with one and the same image do have the same D-length if they are related to each other in a certain way. Specifically, this happens if $\mathbf{f}$ and $\mathbf{g}$ are each others' reparametrizations, by which we mean that, for some nondecreasing and onto (hence continuous) mapping $\phi : [c, d] \to [a, b]$,

$$\mathbf{g}(x) = \mathbf{f}(\phi(x)),\ x \in [c, d].$$

Note that we use a "symmetrical" terminology (*each other's* reparametrizations) even though the mapping ϕ is not assumed to be invertible. If it is invertible, then it is an increasing homeomorphism, and then it is easy to see that $D\mathbf{f} = D\mathbf{g}$. This equality extends to the general case (see Figure 9.10).

We define an *arc* as a path that can be reparametrized into a homeomorphic path. In other words, $\mathbf{g}|\,[c, d]$ is an arc if one can find a nondecreasing and onto (hence

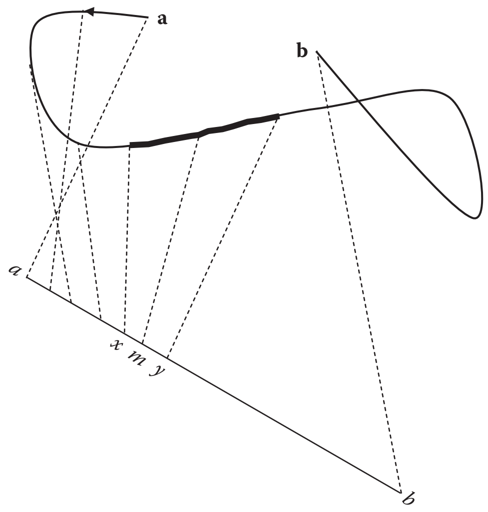

Figure 9.9 Uniform continuity of length: as x and y get closer to each other, the length of the corresponding piece of the path converges to zero.

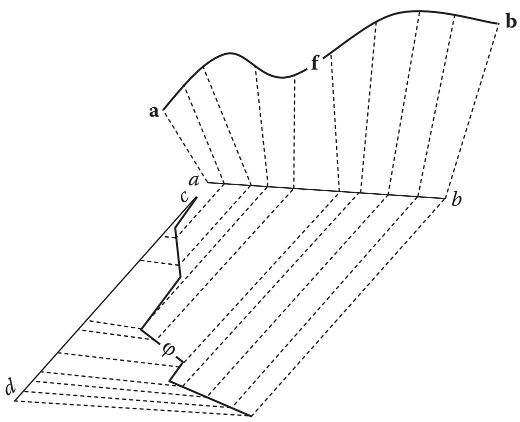

Figure 9.10 The path **f** on $[a, b]$ can be reparametrized without its length affected into a path on $[c, d]$ mapped onto $[a, b]$ by a nondecreasing function ϕ.

continuous) mapping $\phi : [c, d] \to [a, b]$, such that, for some one-to-one and continuous (hence homeomorphic) function $\mathbf{f} : [a, b] \to \mathfrak{S}$,

$$\mathbf{g}(x) = \mathbf{f}(\phi(x)),$$

for any $x \in [c, d]$. It can be shown (by a nontrivial argument) that any path contains an arc with the same endpoints and the D-length that cannot exceed the D-length of

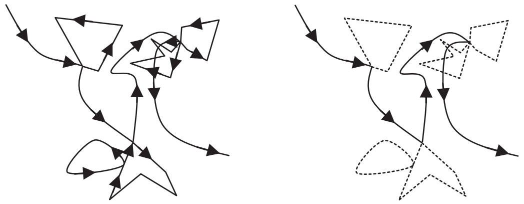

Figure 9.11 One can remove closed loops from a path and be left with a shorter arc.

the path (see Figure 9.11). Stated rigorously, let $\mathbf{f}|[a, b]$ be a D-rectifiable path connecting **a** to **b**. Then there is an arc $\mathbf{g}|[a, b]$ connecting **a** to **b**, such that

$$\mathbf{g}([a, b]) \subset \mathbf{f}([a, b]),$$

and

$$D\mathbf{g}|[a, b] \leq D\mathbf{f}|[a, b],$$

where the inequality is strict if $\mathbf{f}|[a, b]$ is not an arc. This result is important, in particular, in the context of searching for shortest paths connecting one point to another (Section 9.4.4): in the absence of additional constraints this search can be confined to arcs only.

9.4.4 Complete dissimilarity spaces with intermediate points

A dissimilarity space $(\mathfrak{S}, D)$ is said to be a space *with intermediate points* if for any distinct **a**, **b** one can find an **m** such that $\mathbf{m} \notin \{\mathbf{a}, \mathbf{b}\}$ and $D\mathbf{amb} \leq D\mathbf{ab}$ (see Figure 9.12). This notion generalizes that of *Menger convexity* (Blumenthal, 1953, p. 41; the term itself is due to Papadopoulos, 2005). If D is a metric, the space is Menger-convex if, for any distinct **a**, **b**, there is a point $\mathbf{m} \notin \{\mathbf{a}, \mathbf{b}\}$ with $D\mathbf{amb} = D\mathbf{ab}$. (The traditional definition is given for symmetric metrics but it can be easily extended.)

Recall that a space is called *complete* if every Cauchy sequence in it converges to a point. Adapted to $(\mathfrak{S}, D)$, the completeness means that given a sequence of points $\mathbf{x}_n$ such that

$$\lim_{\substack{k\to\infty \\ l\to\infty}} D\mathbf{x}_k\mathbf{x}_l = 0,$$

there is a point **x** in $\mathfrak{S}$ such that

$$\mathbf{x}_n \leftrightarrow \mathbf{x}.$$

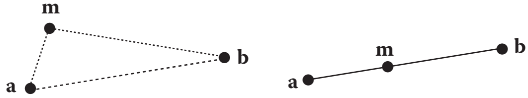

Figure 9.12 Point **m** is intermediate to **a** and **b** if $D\mathbf{amb} \leq D\mathbf{ab}$. E.g., if D is Euclidean distance (right panel), any **m** on the straight line segment connecting **a** to **b** is intermediate to **a** and **b**.

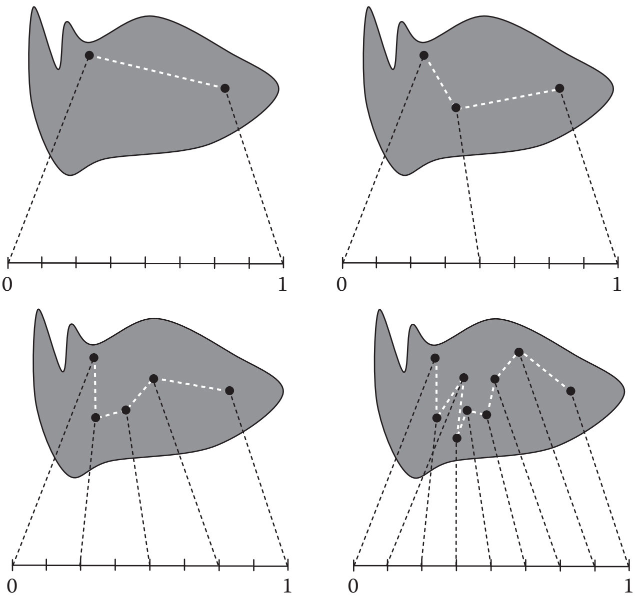

Figure 9.13 In a complete space with intermediate points any points **a** and **b** can be connected by chains whose cardinality increases beyond bounds and the dissimilarity between successive elements converged to zero. As a result the chains converge, pointwise and in length, to an arc whose length is not greater than D**ab**.

Blumenthal (1953, pp. 41–43) proved that if a Menger-convex space is complete then **a** can be connected to **b** by a *geodesic arc*, that is, an arc **h** with D**h** = D**ab** (where D is a symmetric metric). As it turns out, this result can be generalized to nonmetric dissimilarity functions, in the following sense: in a complete space with intermediate points, any **a** can be connected to any **b** by an arc **f** with

$$D\mathbf{f} \leq D\mathbf{ab}.$$

See Figure 9.13 for an illustration. It follows that G**ab** in such a space can be viewed as the infimum of lengths of all arcs connecting **a** to **b**. Put differently, in a complete space with intermediate points the metric G induced by D is intrinsic, in the traditional sense of the word.

9.5 Conclusion

Let us summarize. Universal Fechnerian Scaling is a theory dealing with the computation of subjective distances from pairwise discrimination probabilities. The theory is applicable to all possible stimulus spaces subject to the assumptions that (a) discrimination probabilities satisfy the law of Regular Minimality, and (b) the two canonical psychometric increments of the first and second kind, $\Psi^{(1)}$ and $\Psi^{(2)}$, are dissimilarity functions.

A dissimilarity function $D\mathbf{ab}$ (where D can stand for either $\Psi^{(1)}$ or $\Psi^{(2)}$) for pairs of stimuli in a canonical representation is defined by the following properties:

$D1$. $\mathbf{a} \neq \mathbf{b} \Rightarrow D\mathbf{ab} > 0$;
$D2$. $D\mathbf{aa} = 0$;
$D3$. If $D\mathbf{a}_n\mathbf{a}'_n \to 0$ and $D\mathbf{b}_n\mathbf{b}'_n \to 0$, then $D\mathbf{a}'_n\mathbf{b}'_n - D\mathbf{a}_n\mathbf{b}_n \to 0$; and
$D4$. for any sequence $\mathbf{a}_n\mathbf{X}_n\mathbf{b}_n$, where $\mathbf{X}_n$ is a chain of stimuli, $D\mathbf{a}_n\mathbf{X}_n\mathbf{b}_n \to 0 \Rightarrow D\mathbf{a}_n\mathbf{b}_n \to 0$.

It allows us to impose on the stimulus space the (generally asymmetric) Fechnerian metric $G\mathbf{ab}$, computed as as the infimum of $D\mathbf{aXb}$ across all possible chains $\mathbf{X}$ inserted between $\mathbf{a}$ and $\mathbf{b}$. The overall (symmetric) Fechnerian distance $G^*\mathbf{ab}$ between $\mathbf{a}$ and $\mathbf{b}$ is defined as $G\mathbf{ab} + G\mathbf{ba}$. This quantity does not depend on whether one uses $\Psi^{(1)}$ or $\Psi^{(2)}$ in place of D.

The dissimilarity D imposes on stimulus space a topology and a uniformity structure that coincide with the topology and uniformity induced by the Fechnerian metric G (or G^*). The metric G is uniformly continuous with respect to the uniformity just mentioned. Stimulus space is topologically characterized as a completely normal space.

The Dissimilarity Cumulation theory can be specialized to arc-connected spaces with no additional constraints imposed either on these spaces or on the type of paths. We have seen that the path length can be defined in terms of a dissimilarity function as the limit inferior of the lengths of appropriately chosen chains converging to paths. Unlike in the classical metric based theory of path length, the converging chains generally are not confined to inscribed chains only: the vertices of the converging chains are allowed to "jitter and meander" around the path to which they are converging. Given this difference, however, most of the basic results of the metric-based theory are shown to hold true in the dissimilarity-based theory.

The dissimilarity-based length theory properly specializes to the classical one when the dissimilarity in question is itself a metric (in fact without assuming that this metric is symmetric). In this case the limit inferior over all converging chains coincides with that computed over the inscribed chains only. It is also the case that the length of any path computed by means of a dissimilarity function remains the same if the dissimilarity function is replaced with the metric it induces.

We have considered a class of spaces in which the metrics induced by the dissimilarity functions defined on these spaces are intrinsic: which means that the distance between two given points can be computed as the infimum of the lengths of all arcs connecting these points. We call them spaces with intermediate points, the concept generalizing that of the metric-based theory's Menger convexity.

All of this shows that the properties $D3$ and $D4$ of a dissimilarity function rather than the symmetry and triangle inequality of a metric are essential in dealing with the notions of path length and intrinsic metrics.

In conclusion, it should be mentioned that the notion of dissimilarity and the theory of dissimilarity cumulation has a broader field of applicability than just discrimination functions. Thus, it seems plausible to assume that means or medians of direct *numerical estimates of pairwise dissimilarities*, of the kind used in Multidimensional Scaling (MDS, see, e.g., Borg & Groenen, 1997), can be viewed as dissimilarity values in the technical sense of the present theory. This creates the possibility of using the dissimilarity cumulation procedure as a data-analytic technique alternative to (and, in some sense, generalizing) MDS. Instead of nonlinearly transforming dissimilarity estimates $D\mathbf{ab}$ into distances of a preconceived kind (usually, Euclidean distances in a low-dimensional Euclidean space) one can use dissimilarity cumulation to compute distances $G^*\mathbf{ab}$ from untransformed $D\mathbf{ab}$ and then see if these stimuli are isometrically (i.e., without changing the distances $G^*\mathbf{ab}$ among them) embeddable in a low-dimensional Euclidean space (or another geometric structure with desirable properties). This approach can be used even if the dissimilarity estimates are nonsymmetric. A variety of modifications readily suggest themselves, such as taking into account only small dissimilarities in order to reduce the dimensionality of the resulting Euclidean representation.

Another line of research links the theory of dissimilarity cumulation with *information geometry* (see, e.g., Amari & Nagaoka, 2000) and applies to the categorization paradigm. Here, each stimulus $\mathbf{a}$ is characterized by a vector of probabilities $(a_1, \ldots, a_k)$,

$$\sum_{i=1}^{k} a_i = 1,$$

where a_i indicates the probability with which $\mathbf{a}$ is classified (by an observer or a group of people) into the ith category among certain $k > 1$ mutually exclusive and collectively exhaustive categories. It can be shown, to mention one application, that the square root of the symmetrical version of the Kullback–Leibler divergence measure (Kullback & Leibler, 1951),

$$D\mathbf{ab} = \sqrt{\mathrm{Div}_{\mathrm{KL}}\mathbf{ab}} = \sqrt{\sum_{i=1}^{k} (a_i - b_i) \log \frac{a_i}{b_i}},$$

is a (symmetric) dissimilarity function on any closed subarea of the area

$$\left\{ \mathbf{x} = (x_1, \ldots, x_k) : x_1 > 0, \ldots, x_k > 0, \sum_{i=1}^{k} x_i = 1 \right\}.$$

The stimuli $\mathbf{x}$ can also be viewed as belonging to a $(k - 1)$-dimensional unit sphere, with coordinates $\sqrt{x_1}, \ldots, \sqrt{x_k}$. The cumulation of $D\mathbf{ab}$ leads to the classical for information geometry spherical metric in any spherically convex area of the stimulus space (i.e., an area which with any two stimuli it contains also contains the smaller arc of the great circle connecting them). In those cases where the spherical convexity is not satisfied (e.g., if the sphere has gaps with no stimuli, or stimuli form a discrete set), the computation of the distances along great circles has to be replaced with more general computations using finite chains of stimuli.

Acknowledgment

This research has been supported by NSF grant SES 0620446 and AFOSR grants FA9550-06-1-0288 and FA9550-09-1-0252. I am grateful to my long-term collaborator Hans Colonius who, among other things, opened to me the wealth of the German-language literature on the subject. I am grateful to James T. Townsend and Devin Burns who critically read the first draft of the chapter and suggested improvements.

References

Amari, S., & Nagaoka, H. (2000). *Methods of information geometry*. Providence, RI: American Mathematical Society.

Blumenthal, L. M. (1953). *Theory and applications of distance geometry*. London: Oxford University.

Blumenthal, L. M., & Menger, K. (1970). *Studies in geometry*. San Francisco: W.H. Freeman.

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

Busemann, H. (2005). *The geometry of geodesics*. Mineola, NY: Dover.

Creelman, C. D. (1967). Empirical detectability scales without the jnd. *Perceptual & Motor Skills, 24,* 1079–1084.

Dzhafarov, E. N. (2001). Fechnerian psychophysics. In N. J. Smelser and P. B. Baltes (Eds.) *International encyclopedia of the social and behavioral sciences* (Vol. 8, pp. 5437–5440). New York: Pergamon Press.

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. [see Dzhafarov, E. N. (2006). Corrigendum to "Thurstonian-type representations for 'same–different' discriminations: Probabilistic decisions and interdependent images." *Journal of Mathematical Psychology*, *50,* 511.]

Dzhafarov, E. N. (2004). Perceptual separability of stimulus dimensions: A Fechnerian approach. In C. Kaernbach, E. Schröger, & H. Müller (Eds.), *Psychophysics beyond sensation: Laws and invariants of human cognition* (pp. 9–26). Mahwah, NJ: Erlbaum.

Dzhafarov, E. N. (2006). On the law of regular minimality: Reply to Ennis. *Journal of Mathematical Psychology*, *50,* 74–93.

Dzhafarov, E. N. (2008a). Dissimilarity cumulation theory in arc-connected spaces. *Journal of Mathematical Psychology*, *52,* 73–92. [see Dzhafarov, E. N. (2009). Corrigendum to "Dissimilarity cumulation theory in arc-connected spaces". *Journal of Mathematical Psychology, 53,* 300.]

Dzhafarov, E. N. (2008b). Dissimilarity cumulation theory in smoothly connected spaces. *Journal of Mathematical Psychology*, *52,* 93–115.

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.

Dzhafarov, E. N., & Colonius, H. (2006a). Regular minimality: A fundamental law of discrimination. In H. Colonius & E. N. Dzhafarov (Eds.), *Measurement and representation of sensations* (pp. 1–46). Mahwah, NJ: Erlbaum.

Dzhafarov, E. N., & Colonius, H. (2006b). Generalized Fechnerian scaling. In H. Colonius & E. N. Dzhafarov (Eds.), *Measurement and representation of sensations* (pp. 47–88). Mahwah, NJ: Erlbaum.

Dzhafarov, E. N., & Colonius, H. (2006c). Reconstructing distances among objects from their discriminability. *Psychometrika*, *71,* 365–386.

Dzhafarov, E. N., & Colonius, H. (2007). Dissimilarity Cumulation theory and subjective metrics. *Journal of Mathematical Psychology*, *51,* 290–304.

Falmagne, J. C. (1971). The generalized Fechner problem and discrimination. *Journal of Mathematical Psychology*, *8,* 22–43.

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

Fechner, G. T.(1877). *In Sachen der psychophysik* [*In the matter of psychophysics*]. Leipzig: Breitkopf & Härtel.

Fechner, G. T. (1887). Über die psychischen massprinzipien und das Webersche gesetz [On the principles of mental measurement and Weber's law]. *Philosophische Studien*, *4,* 161–230.

Helmholtz, H. von. (1891). Versucheinererweiterten anwendungdes Fechnerschen gesetzes im farbensystem [An attempt at a generalized application of Fechner's Law to the color system]. *Zeitschrift für die Psychologie und die Physiologie der Sinnesorgane*, *2,* 1–30.

Hocking, J. H. & Young, G. S. (1961). *Topology*. Reading, MA: Addison-Wesley.

Kelly, J. L. (1955). *General topology*. Toronto: Van Nostrand.

Krantz, D. (1971). Integration of just-noticeable differences. *Journal of Mathematical Psychology, 8,* 591–599.

Kullback, S., & Leibler, R. A. (1951). On information and suffciency. *Annals of Mathematical Statistics*, *22,* 79–86.

Luce, R. D. (2002). A psychophysical theory of intensity proportions, joint presentations, and matches. *Psychological Review*, *109,* 520–532.

Luce, R. D. (2004). Symmetric and asymmetric matching of joint presentations. *Psychological Review*, *111,* 446–454.

Luce, R. D., & Edwards W. (1958). The derivation of subjective scales from just noticeable differences. *Psychological Review, 65,* 222–237.

Luce, R. D., & Galanter, E. (1963). Discrimination. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1, 191–244). New York: Wiley.

Papadopoulos, A. (2005). *Metric spaces, convexity and nonpositive curvature*. Zurich, Switzerland: European Mathematical Society.

Pfanzagl, J. (1962). Über die stochastische Fundierung des psychophysischen Gesetzes [On stochastic foundations of the psychophysical law]. *Biometrische Zeitschrift*, *4,* 1–14.

Riesz, R. R. (1933). The relationship between loudness and the minimum perceptible increment of intensity. *Journal of the Acoustical Society of America*, *4,* 211–216.

Schrödinger, E. von. (1920). Farbenmetrik [Color metrics]. *Zeitschrift für Physik*, 12, 459– 466.

Schrödinger, E. von. (1920/1970). Outline of a theory of color measurement for daylight vision. In D. L. MacAdam (Ed.), *Sources of color science* (pp. 397–447, 481–520).Cambridge, MA: MIT Press.

Schrödinger, E. von. (1926/1970). Thresholds of color differences. In D. L. MacAdam (Ed.), *Sources of color science* (pp. 183–193). Cambridge, MA: MIT Press.

Stevens, S. S. (1975). Psychophysics: Introduction to its Perceptual, Neural, and Social Prospects. New York: Wiley.

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.

Ünlü, A., Kiefer, T., & Dzhafarov, E. N. (2009). Fechnerian scaling in R: The package Fechner. *Journal of Statistical Software*, *31,* 1–24.