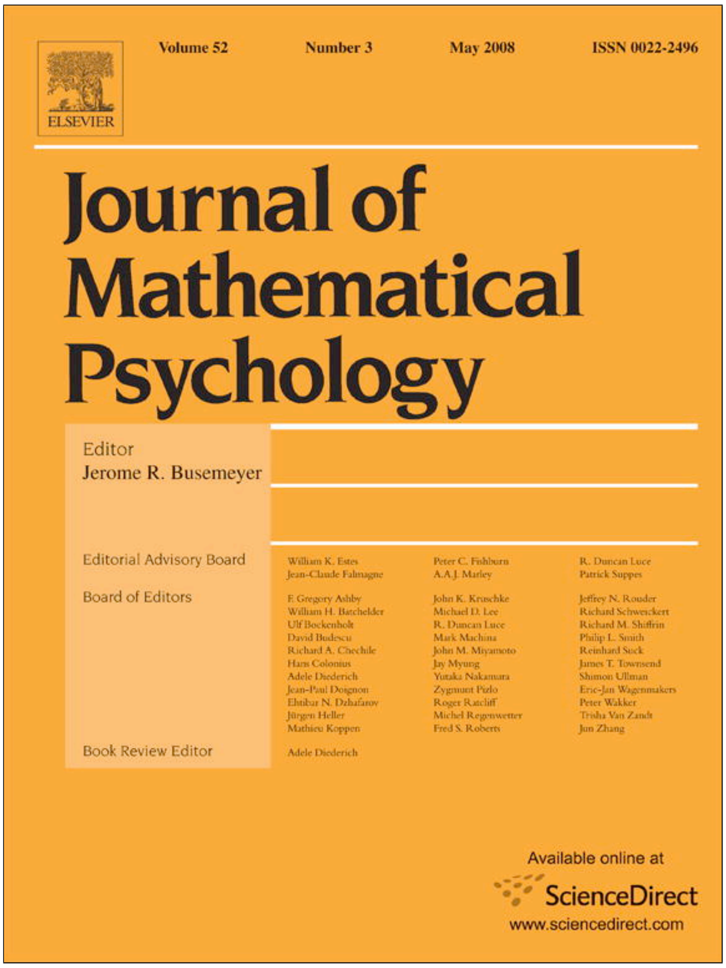
Volume 52
Number 3
May 2008
ISSN 0022-2496
ELSEVIER
Journal of Mathematical Psychology
Editor
Jerome R. Busemeyer
Editorial Advisory Board
William K. Estes
Jean-Claude Falmagne
Peter C. Fishburn
A.A.J. Marley
R. Duncan Luce
Patrick Suppes
Board of Editors
F. Gregory Ashby
William H. Batchelder
Ulf Bockenholt
David Budescu
Richard A. Chechile
Hans Colonius
Adele Diederich
Jean-Paul Doignon
Ehtibar N. Dzhafarov
Jürgen Heller
Mathieu Koppen
John K. Kruschke
Michael D. Lee
R. Duncan Luce
Mark Machina
John M. Miyamoto
Jay Myung
Yutaka Nakamura
Zygmunt Pizlo
Roger Ratcliff
Michel Regenwetter
Fred S. Roberts
Jeffrey N. Rouder
Richard Schweickert
Richard M. Shiffrin
Philip L. Smith
Reinhard Suck
James T. Townsend
Shimon Ullman
Eric-Jan Wagenmakers
Peter Wakker
Trisha Van Zandt
Jun Zhang
Book Review Editor
Adele Diederich
Available online at
ScienceDirect
www.sciencedirect.com

Available online at www.sciencedirect.com
ScienceDirect
Journal of Mathematical Psychology 52 (2008) 184–201
Journal of Mathematical Psychology
www.elsevier.com/locate/jmp

Measurement by subjective estimation: Testing for separable representations

Michele Bernasconi[a], Christine Choirat[b], Raffaello Seri[a,*]

[a] *Dipartimento di Economia, Università dell'Insubria, Via Monte Generoso, 71, I-21100 Varese, Italy*
[b] *Department of Quantitative Methods, School of Economics and Business Management, Universidad de Navarra, Edificio de Bibliotecas (Entrada Este), E-31080 Pamplona, Spain*

Received 8 September 2006; received in revised form 2 October 2007
Available online 8 April 2008

Abstract

Studying how individuals compare two given quantitative stimuli, say d_1 and d_2, is a fundamental problem. One very common way to address it is through *ratio estimation*, that is to ask individuals not to give values to d_1 and d_2, but rather to give their estimates of the ratio $p = d_1/d_2$. Several psychophysical theories (the best known being Stevens' power-law) claim that this ratio cannot be known directly and that there are cognitive distortions on the apprehension of the different quantities. These theories result in the so-called *separable representations* [Luce, R. D. (2002). A psychophysical theory of intensity proportions, joint presentations, and matches. *Psychological Review, 109*, 520–532; Narens, L. (1996). A theory of ratio magnitude estimation. *Journal of Mathematical Psychology, 40*, 109–788], which include Stevens' model as a special case. In this paper we propose a general statistical framework that allows for testing in a rigorous way whether the separable representation theory is grounded or not. We conclude in favor of it, but reject Stevens' model. As a byproduct, we provide estimates of the psychophysical functions of interest.

Keywords: Separable representations; Stevens' model; Psychophysical experiments

1. Introduction

A fundamental question in the behavioral sciences concerns the ability of individuals to evaluate quantities. In this paper we conduct a statistical analysis on data collected in two experiments aimed at answering this question. We focus on ratio evaluations in which individuals do not estimate absolute quantities, but relative ones. In our experiments subjects compare two stimuli, d_1 and d_2, and state in what proportion p they are with respect to each other.

This resembles what in psychophysics Stevens (1946, 1951) called ratio magnitude estimation.[1] More generally, the evaluation of ratios by individuals occurs in millions of everyday situations. These range from simple life experiences, like a friend claiming to have caught a fish which was twice as big as yours, to more complex decisions.[2] To what extent should we rely on the accuracy of the ratio assessments expressed in the various situations?

It is well-known that, according to Stevens (1946), subjective ratio estimations should be treated as any other form of scientific measurement and acknowledged as exhibiting the usual standard of scientific reliability. In particular, while Stevens always recognized the possibility of cognitive

* Corresponding author.
E-mail addresses: bernasconi@uninsubria.it (M. Bernasconi), ccchoirat@unav.es (C. Choirat), rseri@eco.uninsubria.it (R. Seri).

[1] See below for a more precise definition of the various experimental concepts and procedures used in psychophysics.

[2] For example, in economics, the simplest decision problem in condition of risk and uncertainty can be viewed as a situation in which an individual has to decide the price x he is willing to pay to participate in a gamble giving prize y with probability p and 0 otherwise. In many economic models, including the classical expected utility (von Neumann & Morgenstern, 1944), x solves a ratio expression given by:

$$\frac{u(x)}{u(y)} = W(p)$$

where $u(\cdot)$ is a utility index and $W(\cdot)$ is a probability weighting function (which is linear in expected utility, but possibly taking various forms of non-linearity under several non-expected utility models; see e.g. Prelec, 1998 and Tversky & Fox, 1995). As will become apparent below, the psychophysical forms studied in this paper have strong similarities with the above separable utility model (on this theme, see in particular Luce, 2002, 2005).

0022-2496/$ - see front matter © 2008 Published by Elsevier Inc.
doi:10.1016/j.jmp.2008.01.002

distortions in the individual apprehension of quantitative stimuli, his extensive application of subjective estimation experiments in several psychological domains led him to elaborate his famous psychophysical law, establishing that subjective value is a power function of physical value and that equal physical ratios produce equal psychological ratios (Stevens, 1957).

Stevens' approach has been criticized under various perspectives.[3] For mathematical psychologists one of its major drawbacks has always been seen in the lack of rigor and of proper mathematical and philosophical foundations justifying the proposition that, when assessing a ratio judgment, a "subject is, in a scientific sense, 'computing ratios' " (Narens, 1996, p. 109).

In recent years, however, there has been an important stream of research clarifying the conditions and giving various sets of axioms that can justify ratio estimations. An achievement of this literature has in particular been the axiomatization of various theories belonging to a class of so-called *separable representations* (see Narens, 1996, for a seminal paper; Luce, 2001a, 2002 and Narens, 2002 as more recent contributions).[4] Formally, we say that a separable representation holds in a ratio estimation if there exist a *psychophysical function* ψ and a *subjective weighting function* W such that the ratio p is in the following relation with d_1 and d_2:

$$\frac{\psi(d_1)}{\psi(d_2)} = W(p). \tag{1}$$

Eq. (1) corresponds to Narens' (1996) original model and incorporates the notion that various and independent distortions may occur both in the assessment of subjective intensities and in the determination of subjective ratios (see also Luce, 2002). In this perspective, Stevens' model corresponds to a very special case in which either ψ or W is a power function, with the other being the identity (see Section 2). Other models with specific restrictions or expressions for either the function ψ or W have been proposed in the literature.

The objective of this paper is to conduct a statistical analysis of separable representations, to test whether the representation holds in simple ratio estimation experiments and whether Stevens' power-law model is appropriate or other functional representations provide better descriptions of subjects' assessments.

Various recent papers have dealt with tests of specific properties underlying separable representations (including Ellermeier & Faulhammer, 2000, Zimmer, 2005 and Zimmer, Luce, & Ellermeier, 2001 and the series of Steingrimsson & Luce, 2005a,b, 2006, 2007) or tests of particular functional forms (Hollands & Dyre, 2000). In the course of the paper we give some accounts of the properties tested in the above experiments and of the results. However, it seems to us that the present paper is the first that conducts a formal and direct test of formula (1).

Our approach will in particular focus on functional (generally, non-parametric) restrictions of separable representations, rather than on underlying behavioral hypotheses. An advantage of our approach is that it allows for comparisons of models and for establishing which models best represent the data, in terms of *both* goodness-of-fit *and* simplicity or parsimony of representations: it is the approach of model selection that has been recently advocated also in experimental psychology by several authors, including Pitt, Myung, and Zhang (2002), and the various authors who have contributed to two recent issues which the Journal of Mathematical Psychology dedicated to the topic (among others, see Cutting, 2000 and Zucchini, 2000, in addition to the editors' introduction by Myung, Foster, & Browne, 2000 and Wagenmakers & Waldorp, 2006, for general overviews). The approach also permits us to introduce quite simply a stochastic element in the analysis, which is on the one side consistent with the obvious notion that no model of subjective measurement can be thought to hold deterministically (this point has also been acknowledged several times in the literature; see e.g. Luce, 1997 and Stevens, 1946), and on the other side provides an appropriate setting for developing a systematic and robust statistical analysis. A limitation of the approach is that the stochastic term has no explicit relationships with the behavioral properties underlying the separable representations, so that it cannot be directly interpreted in terms of mathematical psychology.

The paper is organized as follows. We start in the next section (Section 2) by giving a detailed classification of separable representations available in the literature; then we describe (in Section 3) the experiments we have conducted to test the different models. Section 4 develops the statistical framework, introducing the stochastic term and showing how the separable representation can be estimated non-parametrically (using polynomial regression) and how it can be tested. The results of the experiments are analyzed in Section 5. The last section (Section 6) brings the various themes of the paper together and concludes. Some technical material is consigned to the Appendix.

2. Theoretical models

Technically, in the theoretical psychophysical literature, the notion of *ratio estimation* does not enter directly in the axioms, which are instead based on the primitive concept of *ratio production* (see e.g. Steingrimsson & Luce, 2006, p. 16). In a ratio production, two stimuli d_1 and d_0, with $d_1 > d_0$, are given with a positive p, and then the respondent is asked to select the stimulus d_2 such that the "subjective" interval from d_0 to d_2 stands in the proportion p to the "subjective" interval from d_0 to d_1. Steingrimsson and Luce (2006) show that a natural interpretation of ratio estimation can be given within the framework of separable representations when $d_0 = 0$. They also

[3] See Michell, 1999 in particular Chapter 7), for a critical review of Stevens in a historical perspective.

[4] There are, however, some important differences in the axiomatic approach of Narens, on the one side, and that of Luce, on the other side; and although this paper doesn't directly deal with psychological primitive concepts and axioms, in the course of the paper we will give some references about the main differences between the two approaches.

Table 1
Models of ratio estimation

Acronym	Name	Formulation	Log–log formulation	Regression model
UNR	Unrestricted	$p = F(d_1, d_2)$	$\pi = f(\delta_1, \delta_2)$	$\pi = f(\delta_1, \delta_2) + \varepsilon$
SEP	Separable	$W(p) = \frac{\psi(d_1)}{\psi(d_2)}$	$w(\pi) = \Psi(\delta_1) - \Psi(\delta_2)$	$\pi = w^{-1}\left[\Psi(\delta_1) - \Psi(\delta_2)\right] + \varepsilon$
LUC	Luce when $p \geq 1$	$\omega \cdot \exp\left[\rho'(\ln p)^{\eta'}\right] = \frac{\psi(d_1)}{\psi(d_2)}$	$\ln\omega + \rho' \cdot \pi^{\eta'} = \Psi(\delta_1) - \Psi(\delta_2)$	$\pi = \left[\frac{\Psi(\delta_1) - \Psi(\delta_2) - \ln\omega}{\rho'}\right]^{\frac{1}{\eta'}} + \varepsilon$
RAT	Ratio	$p = \frac{\psi(d_1)}{\psi(d_2)}$	$\pi = \Psi(\delta_1) - \Psi(\delta_2)$	$\pi = \Psi(\delta_1) - \Psi(\delta_2) + \varepsilon$
STG	Stevens' generalized	$W(p) = \left(\frac{d_1}{d_2}\right)^{\kappa}$	$w(\pi) = \kappa \cdot (\delta_1 - \delta_2)$	$\pi = w^{-1}\left[\kappa \cdot (\delta_1 - \delta_2)\right] + \varepsilon$
STE	Stevens'	$p = \left(\frac{d_1}{d_2}\right)^{\kappa}$	$\pi = \kappa \cdot (\delta_1 - \delta_2)$	$\pi = \kappa \cdot (\delta_1 - \delta_2) + \varepsilon$
NAI	Naive	$p = \frac{d_1}{d_2}$	$\pi = \delta_1 - \delta_2$	$\pi = \delta_1 - \delta_2 + \varepsilon$

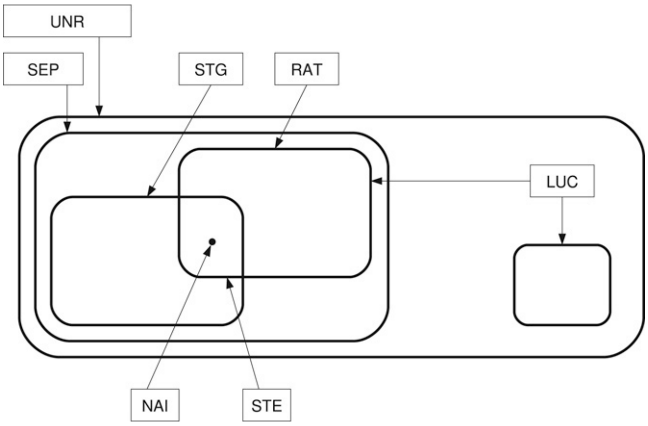

Fig. 1. Respective positions of the theoretical models.

draw attention on the relationship between the concept of ratio estimation and that of *magnitude estimation* most often used by the empirical psychophysical literature and stemming from the works of Stevens, in particular from his posthumous treatise (Stevens, 1975). One example of *magnitude estimation* is an experiment in which the respondent is provided with a standard stimulus d_0 and with a number ψ_0 called the modulus (typically greater than 1); and then to each stimulus d_2, the respondent is asked for its numerical values so that ratios are preserved, meaning that $\frac{\psi(d_2)}{\psi_0} = \frac{d_2}{d_0}$. As it is well known, Stevens thought that with this method one could recover the function ψ directly, but this is precisely to what most of his critics have objected and what modern mathematical psychology has tried to clarify. In the following, we present various models that have been developed in the literature to link the proposed stimuli d_1 and d_2 with the subjects' stated proportion p in a *ratio estimation* experiment. Each model is identified with a three letter acronym which will be used throughout the paper. The models' names, acronyms and deterministic formulations are also shown in the first three columns of Table 1. (The meaning of the last two columns of the Table are explained in Section 4.) Their respective relations are shown in Fig. 1.

Starting with the general *separable representation* of Eq. (1), we indicate it with SEP. As noted, several formulations of this kind have been proposed in the literature. In our opinion, the most lucid derivations of separable representations are due to Narens (in particular, Narens, 1996, 2002) and Luce (especially Luce, 2002, 2004). There are some important differences between the two approaches. Among the most important we observe that whereas Luce is closer in spirit to the psychophysical tradition, developing empirically testable assumptions for three psychophysical primitives (joint presentations of pairs of stimuli, a respondent's ordering of such pairs, and judgments about two pairs of stimuli being related as some proportion), Narens' aim is to provide conditions for ratio subjective estimations to be consistent with the representational theory of measurement, in the sense of carefully axiomatizing what Stevens might have meant, assuming that his method of magnitude estimation yielded the psychophysical functions directly.[5] Despite the differences,[6] an important similarity between the representations obtained by the two authors is that both endorse the idea that *two* independent distortions affect a ratio estimation: one of stimuli intensities, embodied in the *psychophysical function* ψ; and the other of numbers, entailed in the *weighting function* W. Were a subject able to give exactly unbiased estimates, both ψ and W would be everywhere linear, and formula (1) would reduce to $d_1/d_2 = p$. Given, however, the unlikelihood of this possibility according to psychophysicists, we refer to this extreme case as a naive model (NAI).

In an ordinary SEP model, the equality $W(1) = 1$ is generally supposed to hold, that is the individuals are able to correctly estimate ratios of equal stimuli (even this, however, is an assumption requiring some testing); moreover, ψ is defined up to a multiplicative constant so that we can suppose that $\psi(1) = 1$. At last, it is easily seen that both ψ and W are defined up to a power transformation (i.e., if ψ and W are functions for which a separable representation holds, then so are also ψ^r and W^r for any real $r \neq 0$). This is what is called in statistics an identification problem and will lead us to impose a restriction in our empirical investigation.

The original *Stevens' model* (STE) holds when W can be represented as the identity function and ψ is a power function.

[5] On the notion of "direct" measurement see also Narens (2006); while see below for the specific property which Narens (1996) showed Stevens' method required.

[6] Another relevant qualification regarding Luce's axiomatization is that, since it is based on joint presentation of stimuli, generally thought to refer to sensorial intensities (like auditory or visual), the separable representation in Eq. (1) is derived as a special case of a more general model which applies when one of the stimuli is 0 (namely, at the threshold intensity).

This case can be recovered as $\psi(d) = d^{\kappa}$ and $W(p) = p$; or, alternatively, as $\psi(d) = d$ and $W(p) = p^{\frac{1}{\kappa}}$. This second formulation will in fact be used for our empirical investigation because of its computational advantages (see the remarks after Eq. (5) in Section 4.2).

When uncovering what he calls *Stevens's Assumptions*, Narens (1996) states that the fact that W could be chosen such that $W(p) = p$ seems to be "anything more than a coincidence" (p. 110). In particular, he shows that for W to be represented linearly, it is necessary that subjective ratio judgments satisfy a given *multiplicative property*: namely, that if d_2 stands in proportion p to d_1 and d_3 stands in proportion q to d_2, then d_3 stands in proportion pq to d_1.[7, 8] A test of this property has been carried forward by Ellermeier and Faulhammer (2000) and Zimmer (2005) with negative results. Therefore, we will also consider a model that we call *Stevens' generalized model* (STG), in which ψ is a power function and W is left free to vary.

Another model that may be of some interest is obtained when W is the identity and ψ is left free to vary: namely, $p = \psi(d_1)/\psi(d_2)$. In this case the above multiplicative property holds (see Augustin, 2006) and the evaluation of a ratio of two quantities d_1 and d_2 is simply equivalent to the ratio of the separate evaluations of the quantities, so that evaluating relative quantities does not appear to be a different problem from evaluating absolute ones (see e.g. the discussion in Hollands & Dyre, 2000). We call this a *ratio model* (RAT).

Starting from representation (1), Luce (2001a, 2002) has developed a functional form for W similar to a specification which Prelec (1998) proposed in the context of utility theory for risky gambles. The original specification assumed $W(1) = 1$. The more recent specification we analyze here is:

$$W(p) = \omega \cdot \begin{cases} \exp\left[-\rho(-\ln p)^{\eta}\right] & p \in]0, 1] \\ \exp\left[\rho'(\ln p)^{\eta'}\right] & p \in]1, \infty[\end{cases} \tag{2}$$

with ρ, ρ', η and η' all greater than zero; and where ω may be different from one.

The attractiveness of the specification is that, depending on the combinations of the parameters, various shapes for W are possible: concave, convex, S-shaped or inverse S-shaped. In the context of preferences among gambles, an inverse S-shaped W under the assumption $\omega = 1$ (hence $W(1) = 1$) seems to hold (see Luce, 2000 and Prelec, 1998). In recent psychophysical experiments on loudness production, Steingrimsson and Luce (2007) and Zimmer (2005) have instead rejected the behavioral hypothesis underlying the specification with $W(1) = 1$; though the latter accepted one with $W(1) \neq 1$.[9] In the following, we will refer to a separable representation with W as in Eq. (2) as LUC. Since in our experiments we will restrict the analysis to ratio estimations where $p \geq 1$, we focus the attention on the second of the two expressions (see Table 1).

Other examples of functional forms for ψ and W are discussed by Luce, 2002, pp. 526–528) and further by Steingrimsson and Luce (2007). In any case we remark that our aim is not only to assess the form of ψ and W, but also to test whether the general SEP model holds true. Therefore, for the empirical analysis we also consider a general nonparametric alternative to the separable representation, which is simply a model given by $p = F(d_1, d_2)$ (UNR).

Before moving to explain how the various models will be tested, we now present the experimental design.

3. The experiments

Most experiments which have recently tested behavioral properties underlying separable representations have followed the psychophysical tradition of using sensorial stimuli, typically auditory or visual (see e.g. Ellermeier & Faulhammer, 2000, Steingrimsson & Luce, 2005a,b, 2006, 2007 and Zimmer, 2005). This is in line with the purpose of studying most directly the validity of Stevens' ratio scaling method when used to obtain subjective measurements of sensation magnitudes. Our perspective is here slightly different, as we are interested in studying the kind of judgmental process used by people when asked to 'compute' subjective ratio judgments. We conducted two experiments with two different stimuli. In the first experiment participants were given a map of Italy and they were asked in a sequence of 90 questions to indicate how many times a certain Italian city (City A) appearing on the map was according to them more distant from Milan than another Italian city (City B). In each question City A was actually more distant from Milan than City B. Subjects were informed of this. Each question was presented to subjects on a new page of a computer screen and subjects answered the question by filling an apposite window in each page. When a question was answered by a participant, the software moved this participant to the next question and the previous questions and answers were no longer accessible to the participant. Subjects had an incentive to perform well in the experiment: they received 30 cents of Euro for each question in which the error of their estimate was lower in absolute value than 10% of the correct answer. This procedure was explained to the subjects at the beginning of the experiment. Subjects were also told that the correct ratios were computed at the first decimal place, so

[7] A subtlety pointed out by Luce (e.g. Luce, 2005, p. 246) is that the property forces W to be a power function with $W(1) = 1$; when however the property fails, W may still be a power function with $W(1) \neq 1$.

[8] Also note that a similar multiplicative consistency is required by various models of multiple decision analysis used in management (like the Analytic Hierarchy Process, Saaty, 1980), which deal with a situation in which an overall objective is evaluated against several alternatives by separating the main objective in levels of sub-objectives and attributes. A ratio evaluation is then used to give a weight to each attribute about the many times it is preferred under one alternative than under another alternative. A final decision is then obtained multiplying and comparing weights through sub-objectives.

[9] The basic property underlying the specification with $\omega = 1$ is called *reduction invariance* (Luce, 2001a). It says the following: suppose that d_3 stands in proportion t to d_1 and stands in proportion q to d_2 when d_2 stands in proportion p to d_1, then for reduction invariance to hold d_5 should stand in proportion t^N to d_1 (for a constant N) and in proportion q^N to d_4 when d_4 stands in proportion p^N to d_1. Aczél and Luce (2007) show the equivalence between the more general specification (2) (where $\omega \neq 1$) and a property of double compound invariance, called *double reduction invariance*.

Table 2
Summaries of the experiments

	Distance experiment	Area experiment
Number of questions	90	90
Number of participants	20	20
Range of stimuli	cm 0.4–29.2	cm^2 0.41–7.22
Ratio range	1.1–13	1.1–13
Ratio average	6.4	6.4
Participants' mean squared errors		
- Average across participants	318.4	133.9
- Mean of standard deviations	7.07	2.77
Average payment (Euros)	11.1	13.3
Average completion time	93′35″	84′30″

they were also encouraged to perform similar rounding in their subjective assessments.

The design of the second experiment was similar. The only difference was in the type of stimuli. In the second experiment subjects faced 90 questions in each of which they were asked to state how many times the area of a section of a disk (Area A) which subjects were viewing on a computer screen was bigger than another section (Area B) of the same disk. Subjects were informed that in all questions Area A was bigger than Area B. The incentive and the procedure for this experiment were the same as in the previous experiment. In the remaining of the paper we will refer to the first experiment as to the 'distance experiment' and to the second experiment as to the 'area experiment'.

Other characteristics summarizing the two experiments are reported in Table 2. 20 subjects participated in each experiment with no subjects participating in both. In the distance experiment stimuli in centimeters on the map of Italy ranged from cm 0.41 for the City B closest to Milan to cm 29.2 for the City A most distant from Milan; in the area experiment the largest Area A on the computer screen measured cm^2 7.2 and the smallest Area B was cm^2 0.41. Given the different stimuli, the ratios asked in the questions of the two experiments were chosen to range over the same interval (1.1–13) and with the same mean (6.4). Subjects seem, however, to have found different difficulties in answering the questions of the two experiments, as they have given much more precise assessments in the area experiment than in the distance experiment: in fact, the participants' average of mean squared errors in the distance experiment is more than twice the average in the area experiment (318.4 over 133.9). This is an interesting feature resulting from the design, confirming a difference in the two experimental treatments, which we consider further when we will discuss the results of estimation of separable representations for the two experiments.

The average payments per subjects were 11.1 Euros in the distance experiment and 13.3 in the area experiment, with an average completion time of about 1 hour and a half in both experiments.

The experiments were run in individual sessions at the laboratory of experimental economics at the University of Insubria, Varese (Italy), between February and March 2007.

4. A statistical framework to test separable representations

In the following we discuss a framework for conducting a statistical analysis of separable representations, which we then apply to the data from our experiments. We proceed in various steps. First, in Section 4.1, we show how to obtain a regression form for the various theoretical representations discussed above. Then, in Section 4.2, we focus on the inference procedure and on various parametric and non-parametric restrictions which characterize the models in their estimable forms. Lastly, in Section 4.3, we specify the strategy chosen to select the best models following the approach of statistical model selection.

We emphasize that the analyses will be conducted for each individual separately. This is standard in psychophysics experiments and particularly important in the present context in which several nonlinearities may affect the functional forms to be estimated. Indeed, it is well known that, except for the linear model, individual differences in the coefficients of a functional form may compromise the use of a group level analysis. For example, in the area of stochastic modeling, it has been proved via hazard analysis that all mixtures of individuals, who are all modeled by an exponential density function but with some variability in their coefficients, cannot be modeled at the group level by an exponential density function. This occurs because the exponential distribution has constant hazard, but all mixtures of exponentials have monotonically decreasing hazard (see Chechile, 2003). But the problem is quite general and is not particular to the exponential distribution. This provides a further rationale for the analysis being conducted on an individual basis in the present context.

4.1. Regression models

As a first step to transform the theoretical representations of Section 2 into models amenable to statistical estimation, we apply a log–log transformation to obtain a better formulation (see Luce, 2002, p. 526). For the basic SEP model, for example we write:

$$\ln W\left[\exp(\ln p)\right] = \ln \psi\left[\exp(\ln d_1)\right] - \ln \psi\left[\exp(\ln d_2)\right].$$

We define:

$$\pi = \ln p$$
$$\delta_i = \ln d_i$$
$$\ln W\left[\exp(\cdot)\right] = w(\cdot)$$
$$\ln \psi\left[\exp(\cdot)\right] = \Psi(\cdot)$$

with the constraint $\psi(1) = 1$ becoming $\Psi(0) = 0$. In this new parameterization, we can write the representation as:

$$w(\pi) = \Psi(\delta_1) - \Psi(\delta_2).$$

The log–log transformation applied to the other models yields the results shown in the fourth column of Table 1. As an example, model UNR gives the relation $\pi = f(\delta_1, \delta_2)$, in which f is left free to vary.

To obtain a regression expression for the various models and subjects of the experiments, we then add an individual error term to the log–log transformations. This is done in the last column of Table 1. As an example, model SEP for a given individual is written in the following regression form:

$$\pi = w^{-1}\left[\Psi(\delta_1) - \Psi(\delta_2)\right] + \varepsilon. \tag{3}$$

There are various reasons which justify the introduction of a stochastic term in the individual representations. Indeed, several mathematical psychologists have also emphasized that axiomatic theories "are about *idealized* situations and do not involve considerations of error" (Narens, 1996, p. 109; on the same point Luce, 1997, p. 81, and 2001b, p. 28; and even Stevens, 1946, p. 680). On the other hand, possibly due to trembling and rounding, people commit errors and this explains why it is sensible to suppose that the theoretical relations hold only approximately. Though this may seem a trivial point, we however remark that the experimental literature does not usually spend much time discussing and studying the nature of the noise and of the randomness in experimental data.[10]

Notice also that the introduction of the stochastic terms leaves unaffected the non-parametric restrictions which characterize the various theoretical models and their relationships. We now show how such restrictions result in models which can be estimated and tested.

4.2. Estimable models and restrictions

The regression models obtained in the previous subsection are nonparametric. We now explain how they can be transformed into estimable forms.

We start describing the structure of the data provided by the experiments. We use the following notations: let J be the number of comparisons for any subject in the experiments. For any subject, we observe a vector of log-ratios $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_J)$ where $J = 90$. For any stated log-ratio π_j, we know also the values of the stimuli, say $d_{j,1}$ and $d_{j,2}$ and of their logarithmic transformations, namely $\delta_{j,1}$ and $\delta_{j,2}$; (note that the stimuli do not depend on the individual). For each individual, we suppose the existence of a relation of the form $\pi_j = f\left(\delta_{j,1}, \delta_{j,2}\right) + \varepsilon_j$, with vector of residuals $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_J)$. The case of interest is the one in which the function f takes the form (3), but as explained above we will consider also a more general model in which f is left unrestricted: this will allow us to test the restrictions through statistical techniques.

The general method we use is to approximate the functions w^{-1}, Ψ and f through polynomials in their arguments. This is generally possible: according to the Weierstrass Approximation Theorem, any continuous function on a compact domain can be approximated to any desired degree of accuracy by a polynomial in its arguments. This use of polynomial approximation is similar to the derivation of flexible functional forms in the theory of production (see Fuss & McFadden, 1978, p. 236). In this case, $\boldsymbol{\varepsilon}$ incorporates the error arising in this approximation. The polynomial approximation applied to the various regression models yields the following individual estimable forms. We start from the unrestricted model.

UNR. In this case, we approximate the function f in Table 1 with a polynomial in the two variables $\delta_{j,1}$ and $\delta_{j,2}$. Therefore, using the notation introduced above, we estimate the following polynomial regression:

$$\pi_j = \boldsymbol{\beta}' \cdot x_j + \varepsilon_j \tag{4}$$

where $x_j \triangleq (1, \delta_{j,1}, \delta_{j,2}, \delta_{j,1}^2, \delta_{j,1} \cdot \delta_{j,2}, \delta_{j,2}^2, \ldots, \delta_{j,1}^M, \delta_{j,1}^{M-1} \cdot \delta_{j,2}, \ldots, \delta_{j,1} \cdot \delta_{j,2}^{M-1}, \delta_{j,2}^M)'$ is a $\left(\frac{(M+1)(M+2)}{2} \times 1\right)$ vector; M is the *order* of the polynomial regression. The selection of the order of the polynomial regression will be performed as described in Section 4.3.

SEP. The same procedure is applied also to the SEP model in order to approximate the functions $w(\cdot)$ and $\Psi(\cdot)$. We represent w^{-1} through the polynomial expression:

$$w^{-1}(x) \simeq w_L^{-1}(x) = \sum_{\ell=0}^{L} \phi_\ell \cdot x^\ell$$

where $w(0) = 0$ implies that $\phi_0 = 0$. The same development can be used for Ψ:

$$\Psi(y) \simeq \Psi_N(y) = \sum_{n=0}^{N} \gamma_n \cdot y^n, \tag{5}$$

with the understanding that γ_0 should be null (since $\Psi(0) = 0$). Moreover, we will also need the normalization $\gamma_1 = 1$ (see the discussion on Stevens' model above). As a result, we get:

$$\begin{aligned} \pi_j &= w^{-1}\left[\Psi\left(\delta_{j,1}\right) - \Psi\left(\delta_{j,2}\right)\right] + \varepsilon_j \\ &\simeq w_L^{-1}\left[\Psi_N\left(\delta_{j,1}\right) - \Psi_N\left(\delta_{j,2}\right)\right] + \varepsilon_j \\ &= \sum_{\ell=0}^{L} \phi_\ell \cdot \left(\sum_{n=0}^{N} \gamma_n \cdot \delta_{j,1}^n - \sum_{n=0}^{N} \gamma_n \cdot \delta_{j,2}^n\right)^\ell + \varepsilon_j. \end{aligned} \tag{6}$$

The parameters are contained in the vectors $\boldsymbol{\gamma}$ and $\boldsymbol{\phi}$. This means that any model of the SEP class is identified by a couple of letters (N, L), corresponding respectively to the orders of the polynomials approximating the (log–log formulations of the) psychophysical function ψ and weighting function W.

LUC. The LUC model lets ψ free to vary and constrains W to take the form given in Table 1:

$$\pi_j = \left[\frac{\sum_{n=0}^{N} \gamma_n \cdot \delta_{j,1}^n - \sum_{n=0}^{N} \gamma_n \cdot \delta_{j,2}^n - \ln \omega}{\rho'}\right]^{\frac{1}{\eta'}} + \varepsilon_j. \tag{7}$$

STG. Model STG arises when W is left free to vary, so that STG with $\boldsymbol{\phi}$ of length L is a subclass of SEP with $N = 1$.

RAT. Model RAT arises when ψ is left free to vary, while W, because of an identification problem, is equal to a power

[10] Here we do not refer necessarily only to experimental psychology; but also in other fields of the behavioral sciences using experimental methods, there is an increasing dissatisfaction with the way the analyses usually deal with the stochastic assumptions (see for example Hey, 2005, for an emphasis of this problem in experimental economics).

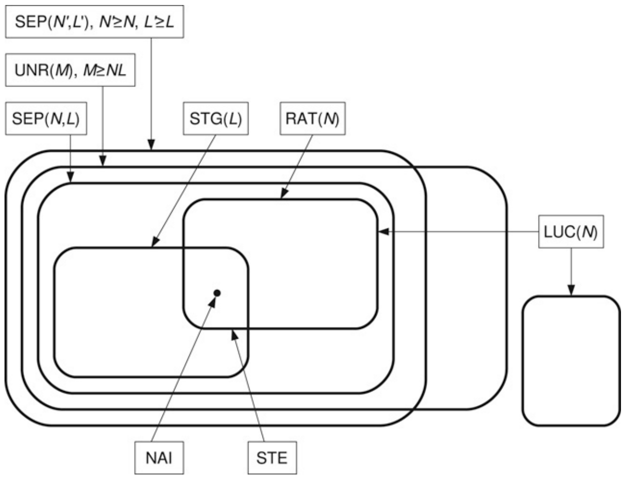

Fig. 2. Respective positions of the econometric models.

function. This means that RAT with $\boldsymbol{\gamma}$ of length N is a subclass of SEP with[11] $L = 1$.

STE. Model STE arises when $N = L = 1$. In this case, some algebra shows that $\psi(d) = d$ and $W(p) = p^{\frac{1}{\phi_1}}$. It is more customary to see Stevens' model in an alternative form with ψ a power function and W the identity. This alternative form of Stevens' model can be obtained as $\psi(d) = d^{\phi_1}$ and $W(p) = p$.

NAI. Model NAI is a special case of the previous ones when $N = L = 1$ and $\phi_1 = 1$ (since γ_1 is already set to 1 for identification).

Fig. 2 shows the respective positions of the econometric models as resulting after the polynomial approximations from the theoretical relations (compare with Fig. 1). Model UNR nests the SEP form, at least as long as its order M is equal or greater than the product $N \cdot L$ of the orders of the two polynomials of the SEP model. When $N = 1$, SEP (N, L) reduces to RAT (N), and when $L = 1$, SEP (N, L) reduces to STG (L). Their intersection is Stevens' original model (STE), containing also the NAI extreme case. The relationships with model LUC are slightly more complex. Indeed, if both $\eta' = 1$ and $\omega = 1$, then LUC (N) reduces to RAT (N) (compare (7) with (8) in footnote 11), but otherwise LUC (N) is not included in any model SEP (N, L) for any value of L. However, when L increases, the function w_L of SEP is able to describe more and more accurately the w appearing in the Luce model (7). Thus, despite the fact that LUC is a special case of SEP (see Section 2 and Fig. 1), LUC (N) is not a special case of SEP (N, L).

4.3. Specification strategy

We propose to analyze the data through the following steps:

- estimating the polynomial regression models for any individual;
- selecting (through tests and information criteria) the best model for any individual.
- performing graphical analyses on the estimated functions.

In the previous Sections, we specified the mean structure of the data. Estimation when only the first one or two moments are known is called *pseudo-* or *quasi-maximum likelihood estimation* (PMLE or QMLE), and was investigated in Gouriéroux, Monfort, and Trognon (1984a). As is customary in the application of PMLE, we perform estimation as if the errors $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_J)$ follow i.i.d. normal distributions with $\mathbb{E}\left(\varepsilon_j\right) = 0$ and $\mathbb{V}\left(\varepsilon_j\right) = \sigma^2$. The estimators obtained in this way enjoy good statistical properties (consistency, asymptotic normality) even if the errors are not homoskedastic or not normal.[12] This estimation method is much more robust than maximum likelihood estimation, since it does not need the complete specification of the density, but only of the first one or two moments; because in our case we have no positive theory concerning the error terms, this characteristic of PMLE turns out to be crucial here. The estimates are therefore obtained through the maximization of the objective function given by:

$$\ell = -\frac{J}{2} \cdot \ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2} \cdot (\boldsymbol{\pi} - \widehat{\boldsymbol{\pi}})' \cdot (\boldsymbol{\pi} - \widehat{\boldsymbol{\pi}}),$$

where $\widehat{\boldsymbol{\pi}}$ is a vector whose elements take one of the forms described in Section 4.2. As an example, for a SEP (N, L) model, the generic element of $\widehat{\boldsymbol{\pi}}$ is

$$\widehat{\pi}_j = \sum_{\ell=0}^{L} \phi_\ell \cdot \left(\sum_{n=0}^{N} \gamma_n \cdot \delta_{j,1}^n - \sum_{n=0}^{N} \gamma_n \cdot \delta_{j,2}^n\right)^\ell$$

for a choice of N and L. The algorithms used to optimize the objective function are described in the Appendix.

Once estimation of the models for several values of the polynomial orders ((N, L) for the separable models, M for the unrestricted ones) has been performed (see Section 5 for the practical method), we have to choose the values of these parameters that give the best models. Two possible interpretations of the procedure we have used can be given. We present a brief account of both of them since they represent two different but interesting approaches to the analyses of psychological data.

In the first approach, the models are considered as parametric and selection of the optimal model is performed using a method that is called BIC (Bayesian Information Criterion) or SC (Schwarz Criterion) (Schwarz, 1978), that weights the likelihood of the model and the number of parameters.[13] If the loglikelihood (not normalized by the

---

[11] In this case, the model is given by:

$$\pi_j = \phi_1 \cdot \left(\sum_{n=0}^{N} \gamma_n \cdot \delta_{j,1}^n - \sum_{n=0}^{N} \gamma_n \cdot \delta_{j,2}^n\right) + \varepsilon_j \tag{8}$$

where $\gamma_0 = 0$ and $\gamma_1 = 1$. The presence of ϕ_1 allows for leaving the function ψ unrestricted (apart from the constraints $\psi(1) = 1$ and, here, also $\phi_0 = 0$).

[12] As an example, in Gouriéroux, Monfort, and Trognon (1984b) this estimation method was applied to discrete Poisson data.

[13] See Pitt et al. (2002), and Wasserman (2000) for general expositions of the BIC as a method for selection between alternative models in psychology; see also Karabatsos (2006). Also note that we decided to use BIC and not AIC (Akaike, 1974) since the latter is not consistent in a statistical sense and we are interested in obtaining good estimates of the parameters representing the order

number of observations) is ℓ, the number of parameters is k and the number of observations is n, we have $\mathsf{BIC} = \frac{\ell - 0.5 \cdot k \cdot \ln(n)}{n}$. We choose the model that has the highest value of BIC. This is the method customarily used in 'autoregressive moving average' (ARMA) models estimation and selection.

In the second approach, the choice of a model using a penalized likelihood function such as BIC can be interpreted as a semiparametric estimation method based on statistical learning theory (see Rissanen, 1978 and Vapnik, 1995, in particular Chapter 4). In this sense, the function to be maximized is:

$$\ell^{\star} = -\frac{\ln\left(2\pi\sigma^2\right)}{2} - \frac{\left(\boldsymbol{\pi}^h - \widehat{\boldsymbol{\pi}}\right)' \cdot \left(\boldsymbol{\pi}^h - \widehat{\boldsymbol{\pi}}\right)}{2J\sigma^2} - \frac{k \cdot \ln(J)}{2J},$$

over all possible values of the order and parameters of the polynomials. This is formally equivalent to the previous procedure, but philosophically different (since the models considered are virtually infinite in number). A discussion on the optimality of this technique is also in Grünwald (2000).

5. Results of the experiments

In the following, we use the above statistical framework and specification strategy to analyze the results of the experiments described in Section 3. The results are presented in two steps. First, we give an account of the best models estimated for each subject in each experiment, namely that on distances and that on areas; then we provide evidence of the actual quality of the best models, supplementing the statistical analyses with some graphical representations of the results.

5.1. Best models

As explained above, in order to select the best models in both experiments for each individual, we use information criteria to firstly choose a model (namely, the order of the polynomial) out of every class, and then compare the best models across classes.

In order to select the best model of every class, we have estimated the models UNR with $M \leq 6$, the models SEP with $N \leq 4$ and $L \leq 4$ and LUC with $N \leq 6$. When the previously described procedure selected a model with M, N or L given by one of these bounds, we have extended the estimation to the nearby models in order to be sure that the selected model was better than the models with similar values of polynomial orders.[14]

The results are summarized in Table 3 for subjects participating in the distance experiment, and in Table 4 for subjects participating in the area experiment. In both tables, BIC represents the Bayesian Information Criterion, ℓ the loglikelihood, k the number of parameters of the model and R^2 the usual coefficient of determination of a regression model. Selection is performed according to BIC, but also the other quantities convey important information; e.g., R^2 gives information about the goodness-of-fit of a model to data.

For the distance data, it turns out that for one subject the best model is a NAI one, for one a UNR model, and for 18 subjects a SEP model: in 12 cases out of 18 a RAT one. No individual model is either a STE or a LUC one. The relevance of separable models in this example is very strong since they turn out to be the best models for the overwhelming majority of cases (19 out of 20 cases, including also the NAI model).

For the area data, the evidence is more balanced across the different classes of models. In 5 cases the best model is a NAI one, in 2 cases a UNR one, in 5 cases a STE one, in 1 case a LUC one, in 7 cases a SEP one. Of these latter, 3 are RAT models and 1 is a STG model. Thus, even in this experiment, the whole class of separable models (including also NAI, STE and LUC) accounts for 18 individuals out of 20. Therefore, also in this experiment separable models appear quite supported by the evidence.

It is also interesting to notice that in both experiments there is some evidence in favor of models consistent with the multiplicative property (see Narens, 1996 and Section 2) underlying some separable forms (see also Fig. 1): namely, in both experiments 13 individuals out of 20 have models consistent with such property (12 RAT models and 1 NAI model in the distance experiment; and 5 STE, 3 RAT and 5 NAI models in the area experiment). This proportion is coherent with a recent result obtained by Augustin and Perner (2007), who found in experiments testing the multiplicative property directly (rather than fitting models) that about 2/3 of their subjects were consistent with the property. It is worth observing that Augustin and Perner (2007) used stimuli very similar to those employed in our experiments: namely, as in our experiments their subjects were asked to compare lengths of lines and areas of circles. On the other hand, as already noted, both Ellermeier and Faulhammer (2000) and Zimmer (2005) reported substantial evidence against the multiplicative property in experiments employing auditory stimuli. This could indicate that the property may hold in some contexts, but not in others.

The comparison between the results of the area experiment and those of the distance experiment provide some further observations. In particular, the evidence shows that the best models and the functions W and ψ selected appear to be simpler for the area data than for the distance data (for the area experiment, in 10 cases out of 20 W and ψ are both power functions, while this happens in only 1 case out of 20 for the distance experiment).[15] This confirms the impression from the comparison

of the polynomials. We recall that AIC is given by the formula $\mathsf{AIC} = \frac{\ell - k \cdot n}{n}$. The use of BIC is also preferred over likelihood ratio tests, since it allows to compare nonnested models as some of those considered in the present analysis (as portrayed by Fig. 2).

[14] This procedure is similar to the one considered in Chambaz (2006). In fact, the method proposed by this author works when selecting a single order parameter (such as M for the UNR class of model), but can reasonably be modified to account for more parameters.

[15] We should note that we also controlled for the effects of a constant, that is $W(1) \neq 1$, in the SEP models estimated in both experiments. For the majority of subjects the constant has been rejected (17 subjects in the distance experiment and 12 in the area experiment). Moreover, for the few cases in which the constant was required by the SEP models, this did not change the best models selected for those individuals.

Table 3
Best models for the distance experiment

		NAI	UNR	STE	STG	RAT	SEP	LUC
Subj. 1	BIC	0.3558392	0.5755687	0.3980556	0.3994690	***0.652608***	***0.652608***	0.602891
	ℓ	34.27543	67.55052	40.32482	42.70193	***67.73434***	***67.73434***	67.75962
	k	1	7 ($M=2$)	2	3 ($L=2$)	***4 ($N=3$)***	***4 ($N=3, L=1$)***	6 ($N=3$)
	R^2	0.9243417	0.9638824	0.9338585	0.9372617	***0.9640296***	***0.9640296***	0.9640498
Subj. 2	BIC	0.159343	***0.2307394***	0.1553487	0.1553487	0.159431	0.159431	0.2122953
	ℓ	16.59077	***36.51588***	18.48119	18.48119	21.09850	21.09850	30.3561
	k	1	***7 ($M=2$)***	2	2 ($L=1$)	3 ($N=2$)	3 ($N=2, L=1$)	5 ($N=2$)
	R^2	0.8853541	***0.9263689***	0.8900705	0.8900705	0.896282	0.896282	0.9155675
Subj. 3	BIC	−0.2775456	0.1099048	−0.07055351	0.02512447	0.1321595	***0.1536641***	0.1368845
	ℓ	−22.72920	25.64077	−1.850006	9.010917	18.64407	***22.82939***	23.56913
	k	1	7 ($M=2$)	2	3 ($L=2$)	3 ($N=2$)	***4 ($N=2, L=2$)***	5 ($N=2$)
	R^2	0.6654476	0.8858053	0.7896419	0.8347503	0.8665953	***0.8784433***	0.8804252
Subj. 4	BIC	−0.1339165	0.2141209	0.08863792	0.1569985	0.1882736	***0.2266557***	0.2081290
	ℓ	−9.802578	28.27050	12.47722	20.87958	23.69434	***31.64854***	29.98114
	k	1	4 ($M=1$)	2	3 ($L=2$)	3 ($N=2$)	***5 ($N=2, L=3$)***	5 ($N=2$)
	R^2	0.7275271	0.8830823	0.8339262	0.8622124	0.870567	***0.8915376***	0.8874434
Subj. 5	BIC	−0.4293568	−0.1301492	−0.2332713	−0.1893004	***−0.0785205***	***−0.0785205***	−0.1037789
	ℓ	−36.39221	−2.713806	−16.49461	−10.28732	***−0.3171311***	***−0.3171311***	1.909425
	k	1	4 ($M=1$)	2	3 ($L=2$)	***3 ($N=2$)***	***3 ($N=2, L=1$)***	5 ($N=2$)
	R^2	0.5943077	0.8080594	0.7392855	0.7728783	***0.8180146***	***0.8180146***	0.8268
Subj. 6	BIC	−0.889972	−0.8215541	−0.806314	−0.806314	***−0.7942027***	***−0.7942027***	−0.8361424
	ℓ	−77.84757	−64.94025	−68.06844	−68.06844	***−64.72853***	***−64.72853***	−66.2532
	k	1	4 ($M=1$)	2	2 ($L=1$)	***3 ($N=2$)***	***3 ($N=2, L=1$)***	4 ($N=1$)
	R^2	0.3060886	0.4791225	0.4416253	0.4416253	***0.4815675***	***0.4815675***	0.4637012
Subj. 7	BIC	0.06301325	0.931199	0.5468022	0.6998209	0.9153664	***0.994172***	0.9663915
	ℓ	7.921098	99.55724	53.712	69.73359	89.13269	***98.4751***	98.22476
	k	1	7 ($M=2$)	2	3 ($L=2$)	3 ($N=2$)	***4 ($N=2, L=2$)***	5 ($N=2$)
	R^2	0.822592	0.9768477	0.9358723	0.9550819	0.9708122	***0.9762842***	0.976152
Subj. 8	BIC	−0.06981755	0.4243652	−0.04736354	0.01573717	***0.5030743***	***0.5030743***	0.4660785
	ℓ	−4.033675	53.9422	0.2370907	10.41596	***52.0264***	***52.0264***	53.19659
	k	1	7 ($M=2$)	2	4 ($L=3$)	***3 ($N=2$)***	***3 ($N=2, L=1$)***	5 ($N=2$)
	R^2	0.818725	0.950018	0.835138	0.8685125	***0.9478442***	***0.9478442***	0.949183
Subj. 9	BIC	−0.5079065	−0.1027510	−0.2383149	−0.2164381	−0.04661070	***−0.02110402***	−0.08283697
	ℓ	−43.46168	6.501747	−16.94853	−10.47981	2.554751	***9.350162***	3.794196
	k	1	7 ($M=2$)	2	4 ($L=3$)	3 ($N=2$)	***5 ($N=2, L=3$)***	5 ($N=2$)
	R^2	0.5728458	0.8592695	0.7630228	0.7947529	0.8463684	***0.8679014***	0.8505421
Subj. 10	BIC	−0.003053316	0.2796558	0.1584040	0.1584040	***0.3505660***	***0.3505660***	0.3135072
	ℓ	1.975106	40.91836	18.75617	18.75617	***38.30066***	***38.30066***	39.46517
	k	1	7 ($M=2$)	2	2 ($L=1$)	***3 ($N=2$)***	***3 ($N=2, L=1$)***	5 ($N=2$)
	R^2	0.8457002	0.9350582	0.8937298	0.8937298	***0.9311684***	***0.9311684***	0.9329268
Subj. 11	BIC	−0.2700566	−0.1871847	−0.2255619	−0.2230210	***−0.1423689***	***−0.1423689***	−0.1699919
	ℓ	−22.05519	−7.847002	−15.80076	−13.32218	***−6.063485***	***−6.063485***	−4.049744
	k	1	4 ($M=1$)	2	3 ($L=2$)	***3 ($N=2$)***	***3 ($N=2, L=1$)***	5 ($N=2$)
	R^2	0.7203672	0.7960776	0.7566526	0.7696937	***0.8040017***	***0.8040017***	0.8125792
Subj. 12	BIC	0.6560692	0.6922386	0.6684267	0.6902837	***0.7099167***	***0.7099167***	0.6881244
	ℓ	61.29613	71.3011	64.65822	68.87525	***72.89212***	***72.89212***	75.43063
	k	1	4 ($M=1$)	2	3 ($L=2$)	***4 ($N=3$)***	***4 ($N=3, L=1$)***	6 ($N=3$)
	R^2	0.9599539	0.9679371	0.9628368	0.9661612	***0.969051***	***0.969051***	0.9707485
Subj. 13	BIC	0.1819191	0.3825154	0.1809258	0.1895696	***0.4669456***	***0.4669456***	0.4180962
	ℓ	18.62263	50.17572	20.78313	23.81098	***51.02472***	***51.02472***	51.12809
	k	1	7 ($M=2$)	2	3 ($L=2$)	***4 ($N=3$)***	***4 ($N=3, L=1$)***	6 ($N=3$)
	R^2	0.8966201	0.9487236	0.9014662	0.907878	***0.949682***	***0.949682***	0.9497974
Subj. 14	BIC	−0.3750792	0.2928246	0.07876666	0.1514760	0.3193653	***0.3370422***	0.3063921
	ℓ	−31.50722	35.35384	11.58881	20.38256	37.7424952	***41.58332***	41.07471
	k	1	4 ($M=1$)	2	3 ($L=2$)	4 ($N=3$)	***5 ($N=3, L=2$)***	6 ($N=3$)
	R^2	0.5405841	0.8960246	0.823686	0.8549832	0.9013998	***0.9094663***	0.9084373
Subj. 15	BIC	−0.1864200	0.5168088	0.3036101	0.3712093	0.6310682	***0.6346396***	0.6142429
	ℓ	−14.52789	71.26174	31.82471	40.15855	68.04566	***70.617***	71.0312
	k	1	11 ($M=3$)	2	3 ($L=2$)	5 ($N=4$)	***6 ($N=4, L=2$)***	7 ($N=4$)
	R^2	0.6907225	0.9540386	0.8895922	0.9082576	0.9506335	***0.9533753***	0.9538025

Table 3 (*continued*)

		NAI	UNR	STE	STG	RAT	SEP	LUC
Subj. 16	BIC	0.3646675	0.373256	0.3695325	0.3695325	***0.4914183***	***0.4914183***	0.4418359
	ℓ	35.06998	49.34237	37.75773	37.75773	***53.22727***	***53.22727***	53.26466
	k	1	7 ($M=2$)	2	2 ($L=1$)	***4 ($N=3$)***	***4 ($N=3, L=1$)***	6 ($N=3$)
	R^2	0.9300895	0.9490904	0.9341428	0.9341428	***0.953301***	***0.953301***	0.9533399
Subj. 17	BIC	0.03272841	0.1845817	0.1188421	0.1753542	***0.2901779***	***0.2901779***	0.2669342
	ℓ	5.195462	25.61197	15.1956	22.53160	***37.36554***	***37.36554***	39.77342
	k	1	4 ($M=1$)	2	3 ($L=2$)	***5 ($N=4$)***	***5 ($N=4, L=1$)***	7 ($N=4$)
	R^2	0.8705023	0.9177336	0.8963067	0.911905	***0.9366438***	***0.9366438***	0.9399448
Subj. 18	BIC	***0.2012491***	0.1447155	0.1838974	0.1838974	0.1838974	0.1838974	0.1601324
	ℓ	***20.36232***	22.02402	21.05057	21.05057	21.05057	21.05057	23.41153
	k	***1***	4 ($M=1$)	2	2 ($L=1$)	2 ($N=1$)	2 ($N=1, L=1$)	4 ($N=1$)
	R^2	***0.9077214***	0.9110668	0.909122	0.909122	0.909122	0.909122	0.913767
Subj. 19	BIC	0.05036855	0.3045986	0.2957756	0.2957756	***0.3119448***	***0.3119448***	0.2834988
	ℓ	6.783075	36.41349	31.11961	31.11961	***34.82475***	***34.82475***	36.76442
	k	1	4 ($M=1$)	2	2 ($L=1$)	***3 ($N=2$)***	***3 ($N=2, L=1$)***	5 ($N=2$)
	R^2	0.8660056	0.9306377	0.9219784	0.9219784	***0.928145***	***0.928145***	0.9311765
Subj. 20	BIC	0.0008843687	0.4981021	0.2850017	0.3068166	***0.5239892***	***0.5239892***	0.4895544
	ℓ	2.329498	53.82881	30.14996	34.36321	***53.90874***	***53.90874***	55.30942
	k	1	4 ($M=1$)	2	3 ($L=2$)	***3 ($N=2$)***	***3 ($N=2, L=1$)***	5 ($N=2$)
	R^2	0.8263914	0.944722	0.906443	0.914805	***0.9448201***	***0.9448201***	0.9465112

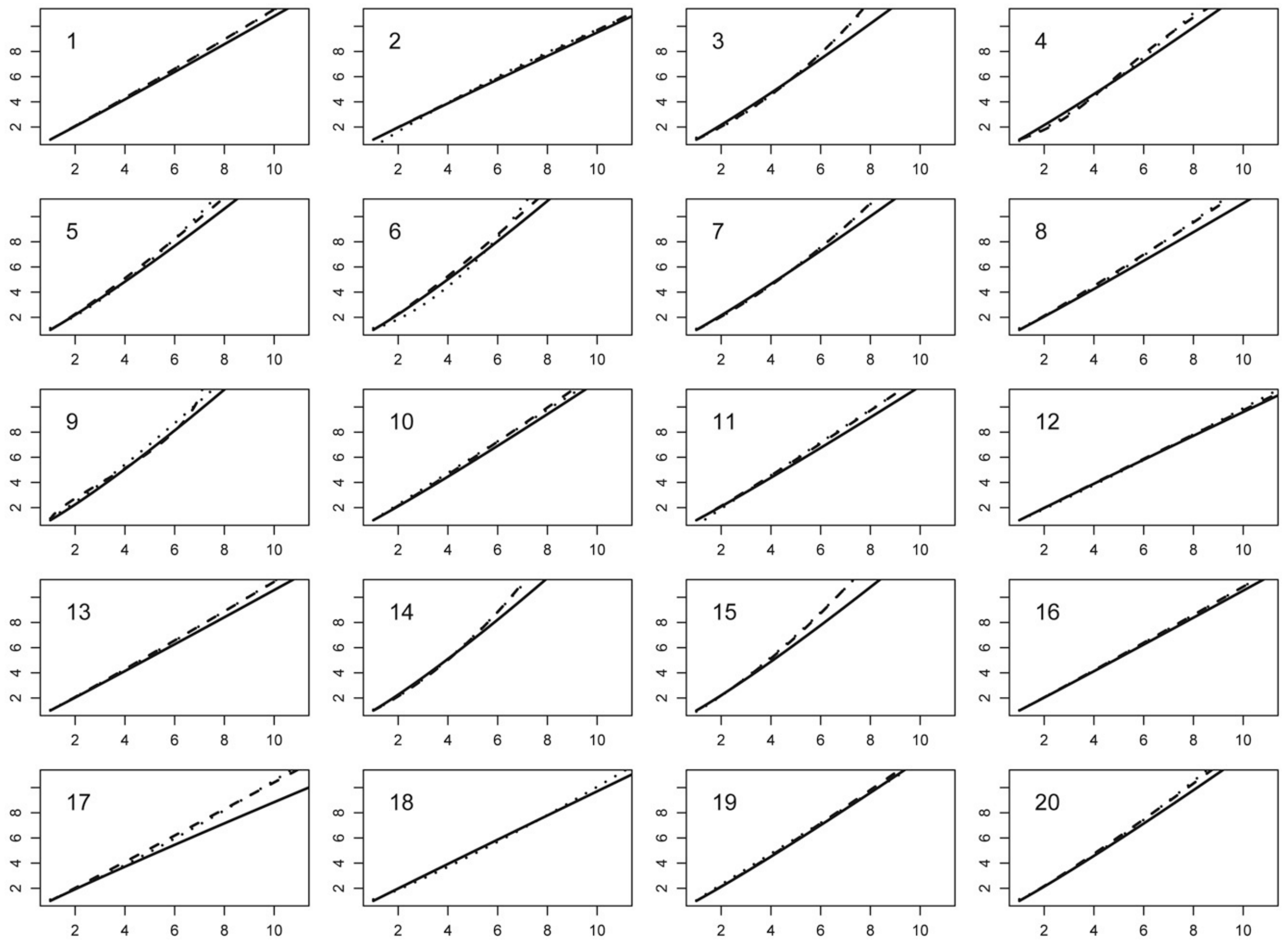

Fig. 3. Subjective weighting functions W for the distance experiment (W_{STE} in solid line, W_{SEP} in dashed line, W_{LUC} in dotted line).

of the average mean squared errors in the two experiments (see Table 2) that subjects may have found the area experiment simpler than the distance experiment. Comparison with evidence from Augustin and Perner (2007) suggests that this could possibly be due to the fact that the estimated distances in the present experiment are evaluated on a map and not along a straight line.

Table 4
Best models for the area experiment

		NAI	UNR	STE	STG	RAT	SEP	LUC
Subj. 1	BIC	0.2621907	***0.3377861***	0.285469	0.3043100	0.3114141	0.3361342	0.3362895
	ℓ	25.84707	***39.40037***	30.19202	38.63742	34.77698	43.75151	41.51558
	k	1	***4 ($M = 1$)***	2	5 ($L = 4$)	3 ($N = 2$)	6 ($N = 2, L = 4$)	5 ($N = 2$)
	R^2	0.8973289	***0.9240294***	0.9067787	0.9227304	0.915809	0.9310312	0.9275177
Subj. 2	BIC	−0.1855750	0.1534926	0.1910611	0.1915993	***0.1957590***	***0.1957590***	0.1621803
	ℓ	−14.45184	22.81396	21.69530	23.99365	***24.36802***	***24.36802***	23.59585
	k	1	4 ($M = 1$)	2	3 ($L = 2$)	***3 ($N = 2$)***	***3 ($N = 2, L = 1$)***	4 ($N = 1$)
	R^2	0.7401156	0.8864654	0.8836077	0.889403	***0.8903193***	***0.8903193***	0.888421
Subj. 3	BIC	0.4255609	0.8964653	0.7960929	0.8216036	0.9007902	0.9007902	***0.9088349***
	ℓ	40.55039	89.6815	76.14817	80.69404	87.82083	87.82083	***93.04466***
	k	1	4 ($M = 1$)	2	3 ($L = 2$)	3 ($N = 2$)	3 ($N = 2, L = 1$)	***5 ($N = 2$)***
	R^2	0.9392501	0.9796117	0.9724582	0.9751046	0.978751	0.978751	***0.9810799***
Subj. 4	BIC	0.1053783	0.4902343	***0.5398454***	***0.5398454***	***0.5398454***	***0.5398454***	0.4930673
	ℓ	11.73395	53.12071	***53.08589***	***53.08589***	***53.08589***	***53.08589***	53.37568
	k	1	4 ($M = 1$)	***2***	***2 ($L = 1$)***	***2 ($N = 1$)***	***2 ($N = 1, L = 1$)***	4 ($N = 1$)
	R^2	0.904865	0.9620757	***0.9620464***	***0.9620464***	***0.9620464***	***0.9620464***	0.96229
Subj. 5	BIC	0.3657834	0.7382387	***0.7868529***	***0.7868529***	***0.7868529***	***0.7868529***	0.7394617
	ℓ	35.17041	75.4411	***75.31657***	***75.31657***	***75.31657***	***75.31657***	75.55117
	k	1	4 ($M = 1$)	***2***	***2 ($L = 1$)***	***2 ($N = 1$)***	***2 ($N = 1, L = 1$)***	4 ($N = 1$)
	R^2	0.9393223	0.9752043	***0.9751355***	***0.9751355***	***0.9751355***	***0.9751355***	0.9752648
Subj. 6	BIC	0.2454235	***0.4994049***	0.3938799	0.3938799	0.3938799	0.3938799	0.3650065
	ℓ	24.33802	***69.69539***	39.949	39.949	39.949	39.949	41.85021
	k	1	***11 ($M = 3$)***	2	2 ($L = 1$)	2 ($N = 1$)	2 ($N = 1, L = 1$)	4 ($N = 1$)
	R^2	0.9176306	***0.9699377***	0.9417757	0.9417757	0.9417757	0.9417757	0.9441844
Subj. 7	BIC	***0.8108012***	0.7747367	0.7875024	0.7875024	0.7875024	0.7875024	0.7768528
	ℓ	***75.22201***	78.72592	75.37502	75.37502	75.37502	75.37502	78.91637
	k	***1***	4 ($M = 1$)	2	2 ($L = 1$)	2 ($N = 1$)	2 ($N = 1, L = 1$)	4 ($N = 1$)
	R^2	***0.9695695***	0.971849	0.9696728	0.9696728	0.9696728	0.9696728	0.971968
Subj. 8	BIC	0.4869596	0.5630296	***0.5887466***	***0.5887466***	***0.5887466***	***0.5887466***	0.5659399
	ℓ	46.07627	59.67228	***57.487***	***57.487***	***57.487***	***57.487***	59.93421
	k	1	4 ($M = 1$)	***2***	***2 ($L = 1$)***	***2 ($N = 1$)***	***2 ($N = 1, L = 1$)***	4 ($N = 1$)
	R^2	0.9462581	0.9602719	***0.958295***	***0.958295***	***0.958295***	***0.958295***	0.9605025
Subj. 9	BIC	0.4338529	0.5106484	0.4673291	0.5935465	0.4673291	***0.6037674***	0.4706943
	ℓ	41.29666	54.95797	46.55943	62.41880	46.55943	***65.58859***	51.36211
	k	1	4 ($M = 1$)	2	4 ($L = 3$)	2 ($N = 1$)	***5 ($N = 2, L = 3$)***	4 ($N = 1$)
	R^2	0.950938	0.963784	0.956353	0.9358723	0.9693171	***0.971404***	0.9607713
Subj. 10	BIC	***0.1754839***	0.1175000	0.1621125	0.1621125	0.1621125	0.1621125	0.1187090
	ℓ	***18.04346***	19.57462	19.08994	19.08994	19.08994	19.08994	19.68343
	k	***1***	4 ($M = 1$)	2	2 ($L = 1$)	2 ($N = 1$)	2 ($N = 1, L = 1$)	4 ($N = 1$)
	R^2	***0.8979018***	0.9013174	0.9002487	0.9002487	0.9002487	0.9002487	0.9015557
Subj. 11	BIC	***0.8062462***	0.7759063	0.7812716	0.787731	0.7812716	0.787731	0.772499
	ℓ	***74.81207***	78.83119	74.81425	77.6455	74.81425	77.6455	78.52453
	k	***1***	4 ($M = 1$)	2	3 ($L = 2$)	2 ($N = 1$)	3 ($N = 1, L = 2$)	4 ($N = 1$)
	R^2	***0.9689395***	0.9715933	0.968941	0.9708349	0.968941	0.9708349	0.971399
Subj. 12	BIC	0.3869065	0.378466	0.362477	***0.4046298***	0.3712455	***0.4046298***	0.3726489
	ℓ	37.07149	43.06156	37.12274	***45.4163***	44.66162	***45.4163***	42.53802
	k	1	4 ($M = 1$)	2	***4 ($L = 3$)***	5 ($N = 4$)	***4 ($N = 1, L = 3$)***	4 ($N = 1$)
	R^2	0.944892	0.9517604	0.9449548	***0.9542197***	0.9534455	***0.9542197***	0.9511959
Subj. 13	BIC	***0.458778***	0.4085165	0.4356576	0.4356576	0.4356576	0.4356576	0.3961295
	ℓ	***43.53992***	45.7661	43.709	43.709	43.709	43.709	44.65127
	k	***1***	4 ($M = 1$)	2	2 ($L = 1$)	2 ($N = 1$)	2 ($N = 1, L = 1$)	4 ($N = 1$)
	R^2	***0.947167***	0.9497171	0.9473652	0.9473652	0.9473652	0.9473652	0.9484558
Subj. 14	BIC	***0.5558471***	0.5220074	0.5412942	0.5412942	0.5412942	0.5412942	0.517969
	ℓ	***52.27614***	55.98029	53.21629	53.21629	53.21629	53.21629	55.61683
	k	***1***	4 ($M = 1$)	2	2 ($L = 1$)	2 ($N = 1$)	2 ($N = 1, L = 1$)	4 ($N = 1$)
	R^2	***0.9505961***	0.9544999	0.9516176	0.9516176	0.9516176	0.9516176	0.954131
Subj. 15	BIC	0.2227264	0.2229523	0.203932	0.2180478	0.2058636	***0.2268932***	0.2022725
	ℓ	22.29528	29.06533	22.85369	28.62392	29.77724	***31.66991***	31.70395
	k	1	4 ($M = 1$)	2	4 ($L = 3$)	5 ($N = 4$)	***5 ($N = 2, L = 3$)***	6 ($N = 3$)
	R^2	0.9258118	0.936174	0.9267267	0.935545	0.9371759	***0.9397634***	0.939809

Table 4 (*continued*)

		NAI	UNR	STE	STG	RAT	SEP	LUC
Subj. 16	BIC	0.6950906	0.6833691	***0.7331783***	***0.7331783***	***0.7331783***	***0.7331783***	0.6832526
	ℓ	64.80806	70.50284	***70.48586***	***70.48586***	***70.48586***	***70.48586***	70.49235
	k	1	4 ($M = 1$)	***2***	***2 ($L = 1$)***	***2 ($N = 1$)***	***2 ($N = 1, L = 1$)***	4 ($N = 1$)
	R^2	0.9661618	0.9701842	***0.970173***	***0.970173***	***0.970173***	***0.970173***	0.9701772
Subj. 17	BIC	0.5736484	***0.7394149***	0.6172378	0.6833293	0.6283050	0.6833293	0.7347015
	ℓ	53.87826	***75.54696***	60.05121	72.74916	67.79698	72.74916	75.12275
	k	1	***4 ($M = 1$)***	2	5 ($L = 4$)	5 ($N = 4$)	5 ($N = 1, L = 4$)	4 ($N = 1$)
	R^2	0.9476398	***0.9676498***	0.9543515	0.9655746	0.9615698	0.9655746	0.9673434
Subj. 18	BIC	0.1667158	0.2979954	0.2988104	0.2988104	***0.3319582***	***0.3319582***	0.2935370
	ℓ	17.25433	35.81921	31.39275	31.39275	***36.62595***	***36.62595***	37.66786
	k	1	4 ($M = 1$)	2	2	***3 ($N = 2$)***	***3 ($N = 2, L = 1$)***	5 ($N = 2$)
	R^2	0.8907787	0.9277	0.9202267	0.9202267	***0.9289847***	***0.9289847***	0.93061
Subj. 19	BIC	0.1533406	0.1529293	***0.1620068***	***0.1620068***	***0.1620068***	***0.1620068***	0.1529397
	ℓ	16.05056	22.76325	***19.08043***	***19.08043***	***19.08043***	***19.08043***	22.76419
	k	1	4 ($M = 1$)	***2***	***2 ($L = 1$)***	***2 ($N = 1$)***	***2 ($N = 1, L = 1$)***	4 ($N = 1$)
	R^2	0.9106638	0.9230439	***0.9164808***	***0.9164808***	***0.9164808***	***0.9164808***	0.9230455
Subj. 20	BIC	0.7959745	0.9065617	0.7904392	0.8301603	***0.932089***	***0.932089***	0.9042947
	ℓ	73.88761	90.59017	75.63934	81.46414	***90.63772***	***90.63772***	92.63605
	k	1	4 ($M = 1$)	2	3 ($L = 2$)	***3 ($N = 2$)***	***3 ($N = 2, L = 1$)***	5 ($N = 2$)
	R^2	0.9666907	0.977019	0.9679624	0.9718522	***0.9770433***	***0.9770433***	0.9780404

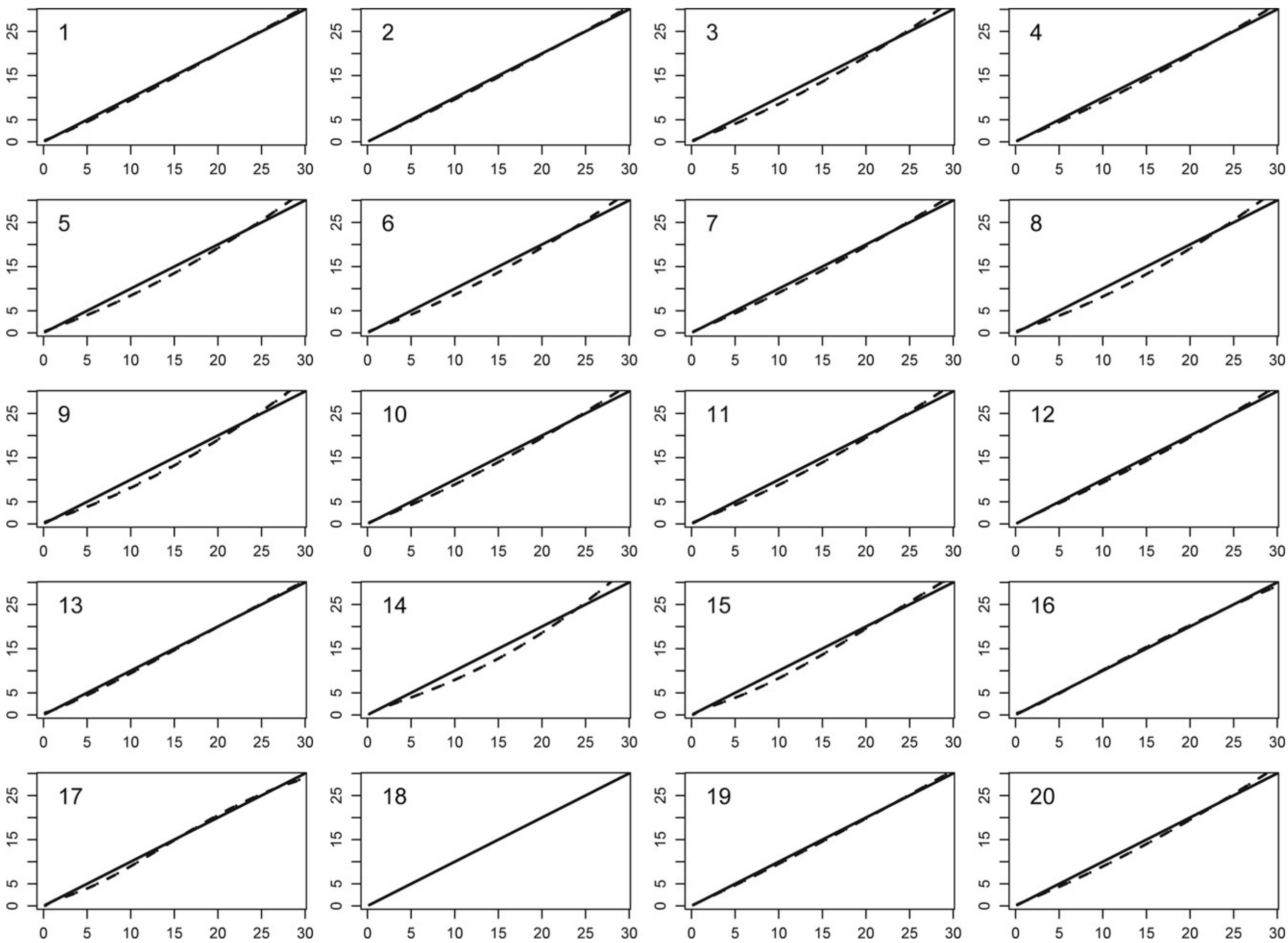

Fig. 4. Psychophysical functions ψ for the distance experiment (ψ_{STE} in solid line, ψ_{SEP} in dashed line, ψ_{LUC} in dotted line).

5.2. Graphs

The statistical analysis we have proposed, based on the method of polynomial approximation, provides us with simultaneous estimates of both the functions ψ and W^{-1} and allows us to measure how much the data conform to the theories. This can be quantified through goodness-of-fit measures (see R^2 in Tables 3 and 4) and checked through

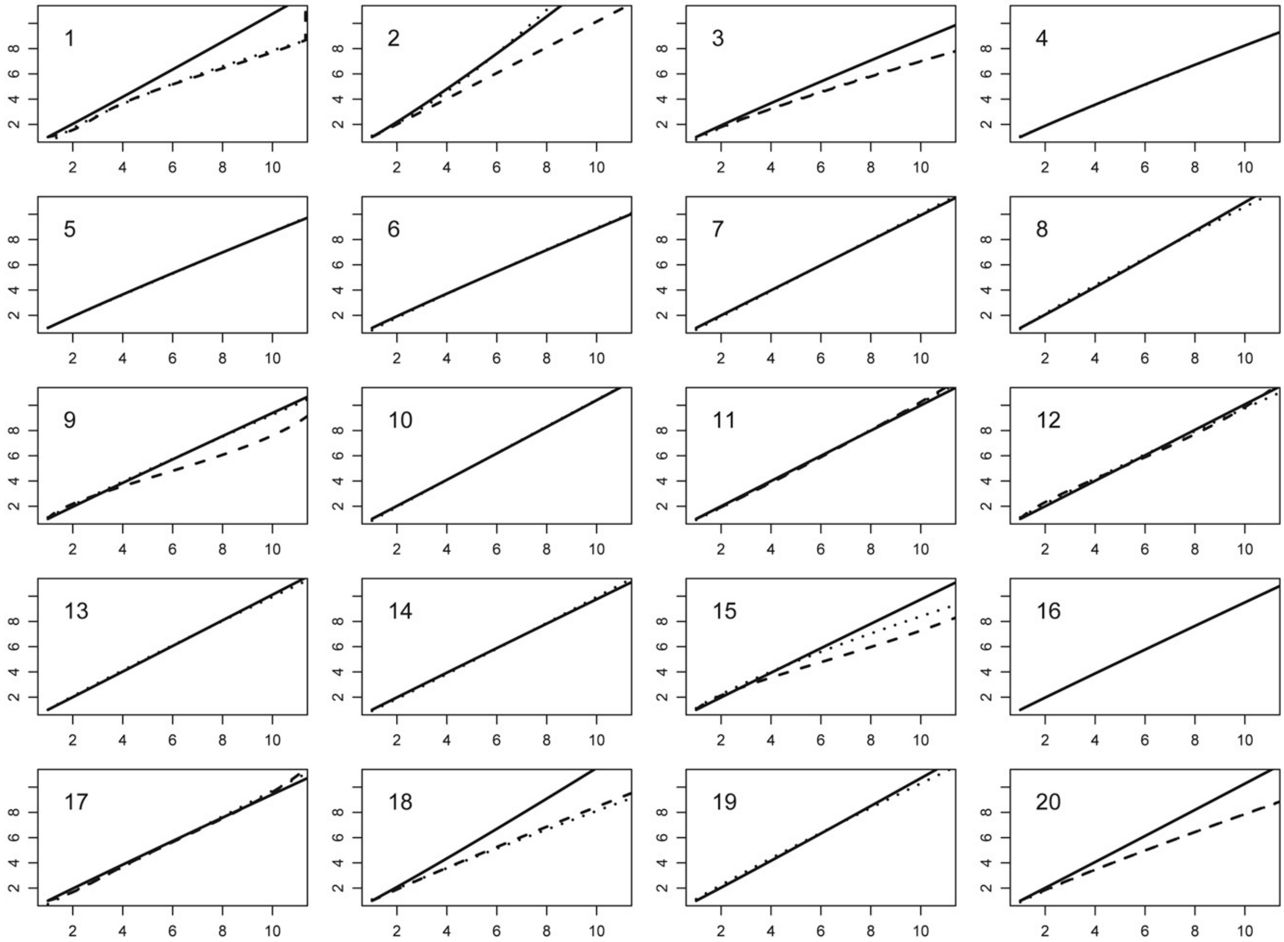

Fig. 5. Subjective weighting functions W for the area experiment (W_{STE} in solid line, W_{SEP} in dashed line, W_{LUC} in dotted line).

graphical methods. In the following we will show how these graphical analyses can be performed.

As explained above, the functions ψ are defined up to a multiplicative transformation (i.e. if a separable representation holds with ψ, then also $k \cdot \psi$ yields a separable representation for any $k \neq 0$), while the couples (ψ, W) are defined up to a power transformation (i.e. if a separable representation holds with (ψ, W), then also (ψ^r, W^r) yields a separable representation for any $r \neq 0$). This implies that, when imposing identification in the estimations as we did above, the functions are not immediately comparable. Therefore, the functions have been rescaled (in particular, we chose values of k and r that minimize a certain distance between the ψ functions; details are available from the authors upon request). The functions W are obtained through numerical inversion of W^{-1}.

Figs. 3 and 4 reproduce respectively the subjective weighting functions W and the psychophysical functions ψ for the distance experiment for every individual (STE in solid line, SEP in dashed line, LUC in dotted line). In both figures we see that for most individuals the functions estimated for SEP and for LUC are nearer to each other than those estimated for STE. This is in line with the fact that while in the distance experiment a separable representation is accepted for most individuals, the selected models are never of the simple Stevens type.

Similarly, Figs. 5 and 6 represent for the area experiment the subjective weighting functions W and the psychophysical functions ψ estimated for every individual (STE in solid line, SEP in dashed line, LUC in dotted line). Two main characteristics of the diagrams are worth noticing. In the diagrams of the psychophysical functions ψ, we see that for most individuals there are little differences in the estimated functions ψ_{STE}, ψ_{SEP} and ψ_{LUC}. In fact, the three functions are almost linear for most individuals. This could be due to the fact that subjects have not found the area experiment to be very difficult (and in any case easier than the distance experiment), with rather precise perceptions of the stimuli (which is what the functions ψ capture). At the same time, we still see some deformations of numerical ratios as portrayed in the subjective weighting functions W_{STE}, W_{SEP}, and W_{LUC}.

After that, we show the fit of the different regression models (see Figs. 7 and 8). For every individual, we plot on logarithmic axes $W(p)$ (where p is the ratio as stated by the subject) against $\psi(d_1)/\psi(d_2)$: the empty circles represent the STE model, the solid circles the SEP model, and the squares the LUC model. The points are expected to be quite near to the diagonal of the diagrams whenever the fit is good. These graphs convey some useful information. First of all, most subjects seem quite precise in their assessments with only few participants displaying more noisy patterns (especially in the distance experiment).

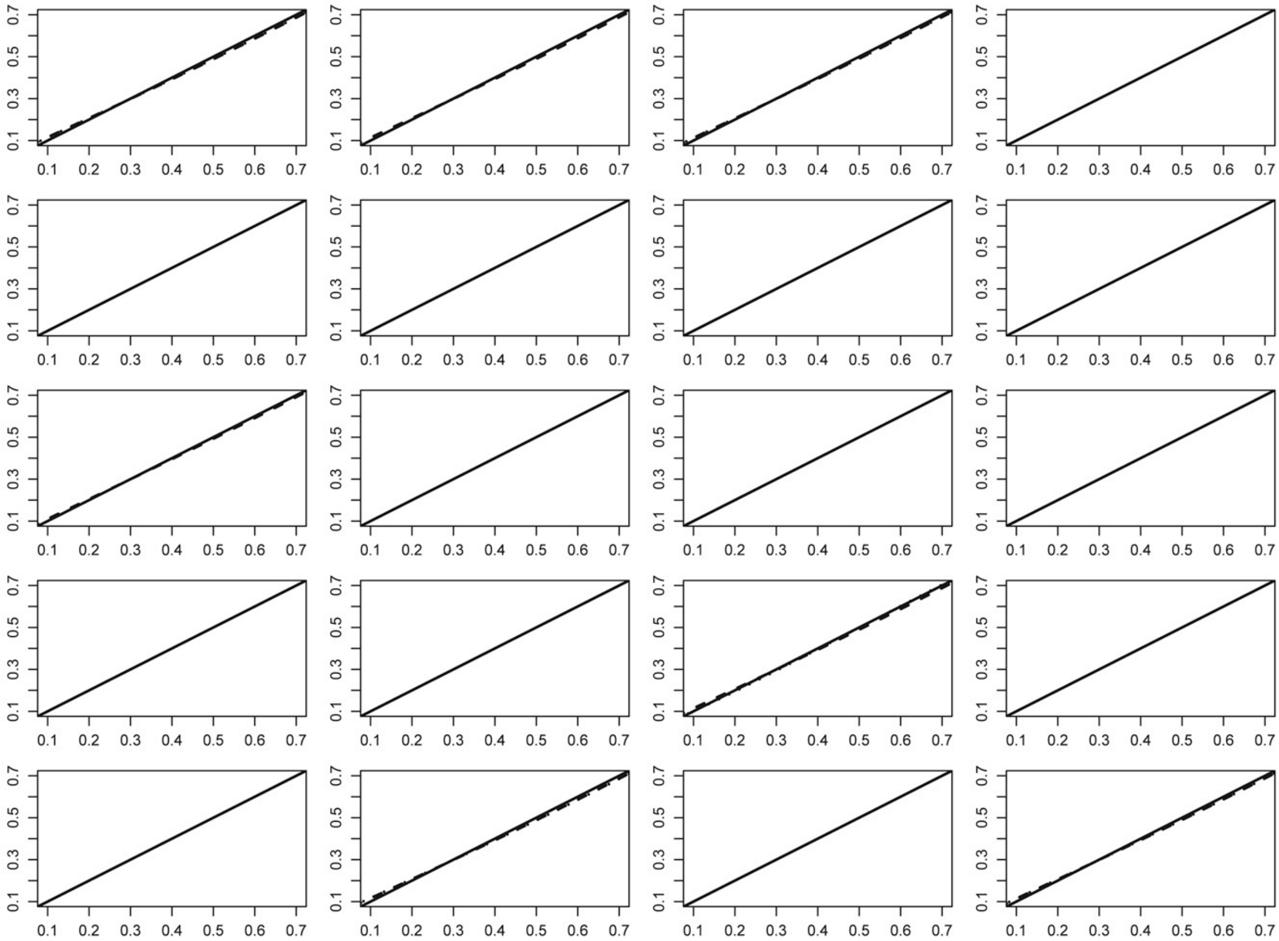

Fig. 6. Psychophysical functions ψ for the area experiment (ψ_{STE} in solid line, ψ_{SEP} in dashed line, ψ_{LUC} in dotted line).

Secondly, all the models seem to fit the data rather well; nevertheless, for various individuals (and more in the distance experiment), there seems to be a better fit for the SEP and LUC models than for the STE ones, as also mirrored through the values of R^2. At last, some heteroskedasticity seems to be present in the data, in the sense that the variances of the elicited ratios are larger when the elicited ratio is large.

Our approach allows us to evaluate the expected value of the elicited ratio p according to the different models as a function of the stimuli d_1 and d_2. In order to stress the difference with respect to the real value of p (namely $\frac{d_1}{d_2}$), Fig. 9 shows the deviations from the NAI model of the values predicted by the main models estimated for three individuals of the distance experiment: subject 1 (first row), subject 3 (second row), and subject 7 (third row).[16] The deviations are as follows: $\left[\left(\frac{d_1}{d_2}\right)^{\kappa} - \frac{d_1}{d_2}\right]$ for STE (the first column), $\left[W_{\mathsf{SEP}}^{-1}\left(\frac{\psi_{\mathsf{SEP}}(d_1)}{\psi_{\mathsf{SEP}}(d_2)}\right) - \frac{d_1}{d_2}\right]$ for SEP (the second column), $\left[W_{\mathsf{LUC}}^{-1}\left(\frac{\psi_{\mathsf{LUC}}(d_1)}{\psi_{\mathsf{LUC}}(d_2)}\right) - \frac{d_1}{d_2}\right]$ for LUC (the third column) and $\left[F(d_1, d_2) - \frac{d_1}{d_2}\right]$ for UNR (the fourth column). In the diagrams the solid lines are the *isolines* or *level sets* of the surface; we represent the level curves corresponding to multiples of 0.05 units. A caveat concerns the graphs of the UNR model, since the function F is meaningfully defined only in the lower diagonal part of the square, the one in which $d_1 \geq d_2$: therefore, we have defined $\left[F(d_1, d_2) - \frac{d_1}{d_2} = 0\right]$ for $d_1 < d_2$. The main conclusion of these graphs is the inadequacy of Stevens' model to take into account the structure of the ratios elicited from the three individuals. In particular, the estimated STE models represent the data as if biases were constant along straight lines; whereas the isolines are often highly nonlinear and similar across all the other three models SEP, LUC and UNR. For all these three models the region with the highest bias is the one with large values of d_1 and intermediate values of d_2, while for the STE model the highest bias is associated with high values of $\frac{d_1}{d_2}$. For example, for subject 3, when $d_1 = 15$ and $d_2 = 10$, the values of p predicted by the best models of the SEP, LUC, UNR and STE classes are in all cases around 1.6 (as can be checked by inspection of the graphs in terms of deviations from the NAI model). On the other hand, when $d_1 = 15$ and $d_2 = 5$, the values of p predicted by the best models of the SEP, LUC and UNR classes are all close to 3.1, while the value is around 3.4 according to the STE model.

[16] The individuals have been chosen in order to illustrate the potential usefulness of this analysis. The graphs of all subjects are available from the authors.

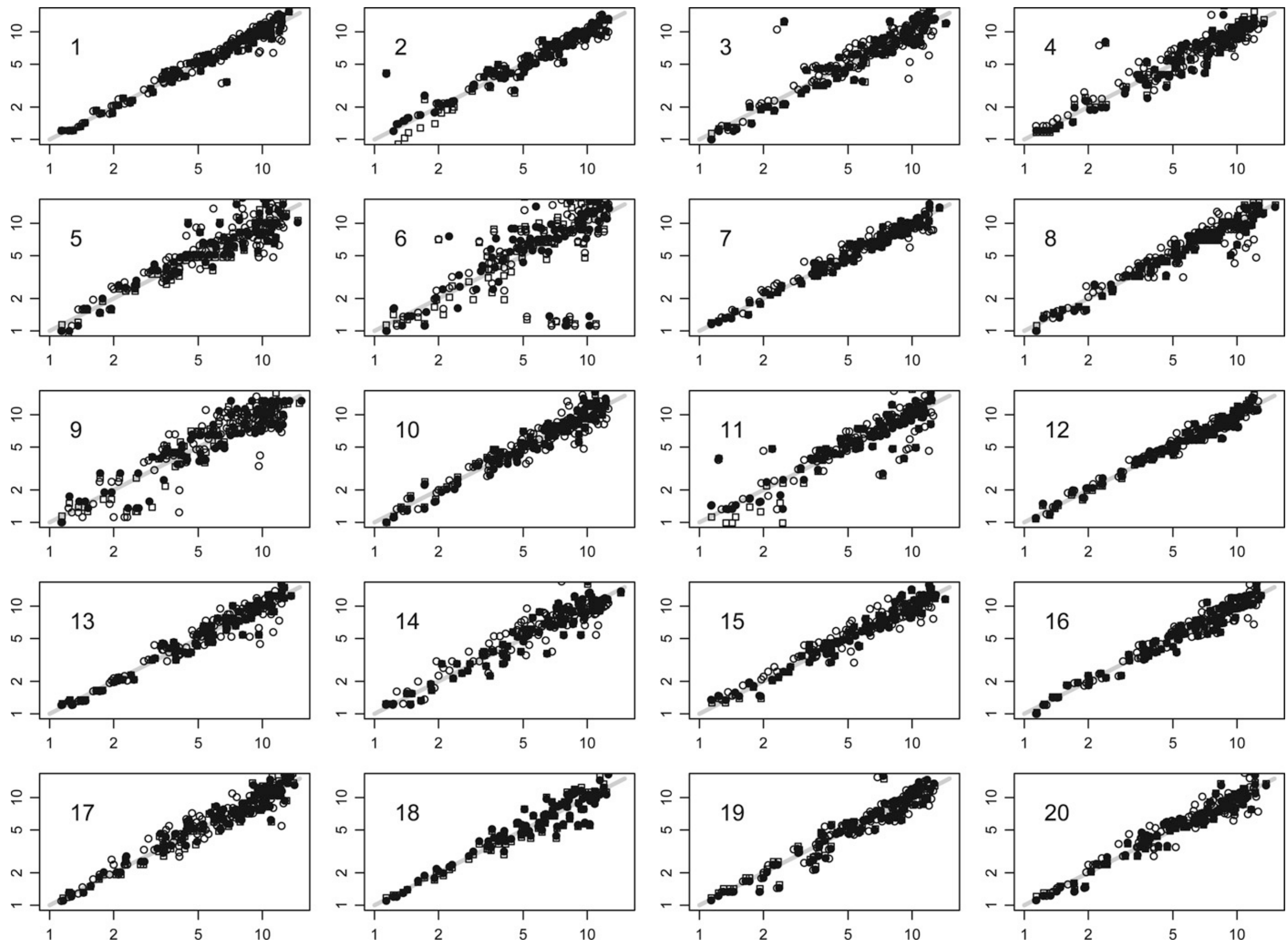

Fig. 7. $W(p)$ and $\psi(d_1)/\psi(d_2)$ on a logarithmic scale for the distance experiment (STE in circles, SEP in solid circles, LUC in squares).

6. Conclusion

We summarize the object of this paper and its relationship to the literature as follows.

We started from a basic question in the behavioral sciences concerning the ability of individuals to perceive and organize stimuli in a meaningful quantitative scale. We have focused on ratio scale measures which, since Stevens' (1946) famous classification, are considered by the scientific community as the very essence of quantitative scientific measurements. We have also recalled how, precisely for this reason, Stevens' theory has originated several controversies and has spawned a large literature (see e.g. Michell, 1999, for a thorough review in a historical perspective).

We have surveyed various theories generalizing Stevens' original power-law model, and have illustrated the importance in the more general theories to clearly distinguish a psychophysical function ψ – which captures the transformation from the stimulus to the perception –, and a subjective weighting function W – which renders the perception of proportions. We have referred to Narens (1996, 2002) and Luce (2002, 2004) as the most lucid treatments of these generalizations.

We have, however, also noted how, despite the great effort to root ratio scale representations on sound scientific bases, less attention has been paid in the past to formally testing separable representations. This has changed in recent years, with a series of experiments which have shown the empirical validity of various behavioral properties underlying models of psychophysical judgments. Important experiments conducted in this wave include Ellermeier and Faulhammer (2000), Zimmer (2005), and the recent series by Steingrimsson and Luce (2005a,b, 2006, 2007).

Our approach has differed from the above since it was designed to evaluate whether individual subjective assessments are overall consistent with ratio scale representations, and to provide a method to coherently and jointly estimate the functions ψ and W. We have classified various theories of ratio scale representations and have adopted the statistical approach of model selection (Pitt et al., 2002; Wasserman, 2000; Zucchini, 2000) to determine the most appropriate models for describing the data which we have obtained from two simple ratio estimation experiments: one in which participants were asked to compare distances between cities on a map, and another one in which subjects were asked to compare the areas of two sections of disks displayed on a computer screen.

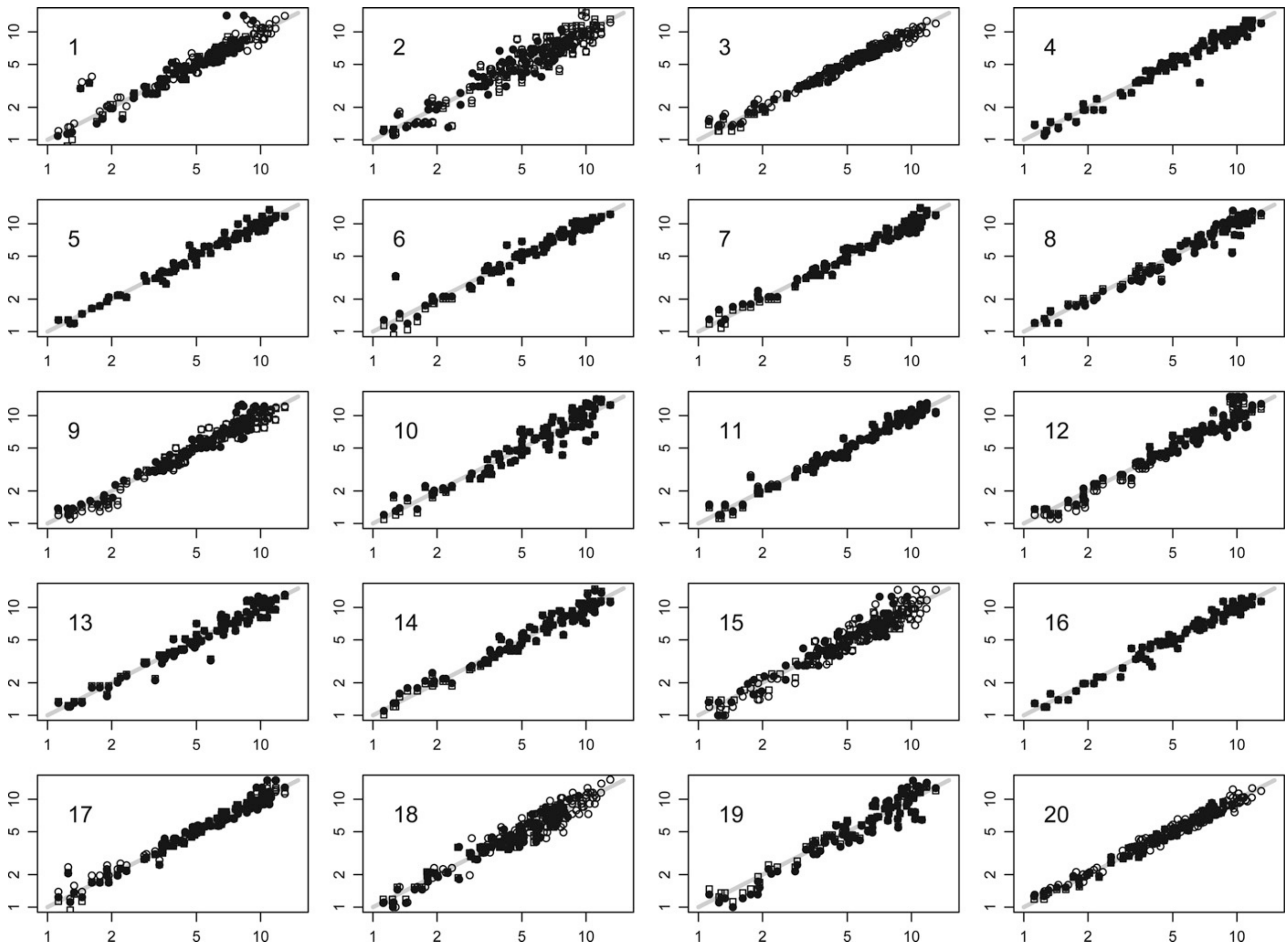

Fig. 8. $W(p)$ and $\psi(d_1)/\psi(d_2)$ on a logarithmic scale for the area experiment (STE in circles, SEP in solid circles, LUC in squares).

We have found substantial support for separable representations. On the one hand, we have seen that in both experiments, and for the great majority of individuals, econometric models estimated under restrictions imposed by different classes of separable models were selected as best models even over specifications left unrestricted. This is important since it shows that separable forms can resolve quite efficiently the delicate trade-off between simplicity and accuracy which, as in any scientific analysis, also emerges in the study of ratio estimation experiments. On the other hand, we have however also seen some heterogeneity among the separable models selected for different individuals and between the two experiments. Both types of evidence are perhaps not surprising, but they confirm just how difficult is the study of subjective measurements to obtain propositions of general validity.

Compared to the previous experiments referred to above, one new piece of evidence is that we found some support for separable representations implying a multiplicative property, described at length by Luce (2005) and Narens (1996), which was instead rejected in direct tests of the hypothesis by Ellermeier and Faulhammer (2000) and Zimmer (2005) in experiments conducted with auditory stimuli. The different stimuli may clearly have had an effect, but also the different statistical approach could be important.

In this respect, we conclude by acknowledging an obvious limitation of the statistical model selection approach itself. In particular, there is no theorem that states that nature is simple. We certainly agree with most scientists in preferring simpler models to more complex ones; but we also believe that statistics, which incorporates a cost function based on the number of parameters, is not alone sufficient to resolve major controversies in science. We more simply believe that model selection statistics are useful measures that need to be considered when models are being fit.

Acknowledgments

We are particularly grateful to Professors R. Duncan Luce, Ragnar Steingrimsson, the editor in charge Richard Chechile and an anonymous referee for several helpful comments on an earlier version of the paper. We also thank Professors Thomas Augustin and Michael Birnbaum for discussions. We are indebted to participants to the 9th and 10th European Congresses of Psychology, the 36th and 38th European Mathematical Psychology Group Meetings, FUR XII 2006 and the Economic Science Association 2007 World Meeting, and to seminar participants at Università degli Studi dell'Insubria, Varese, Università della Svizzera Italiana,

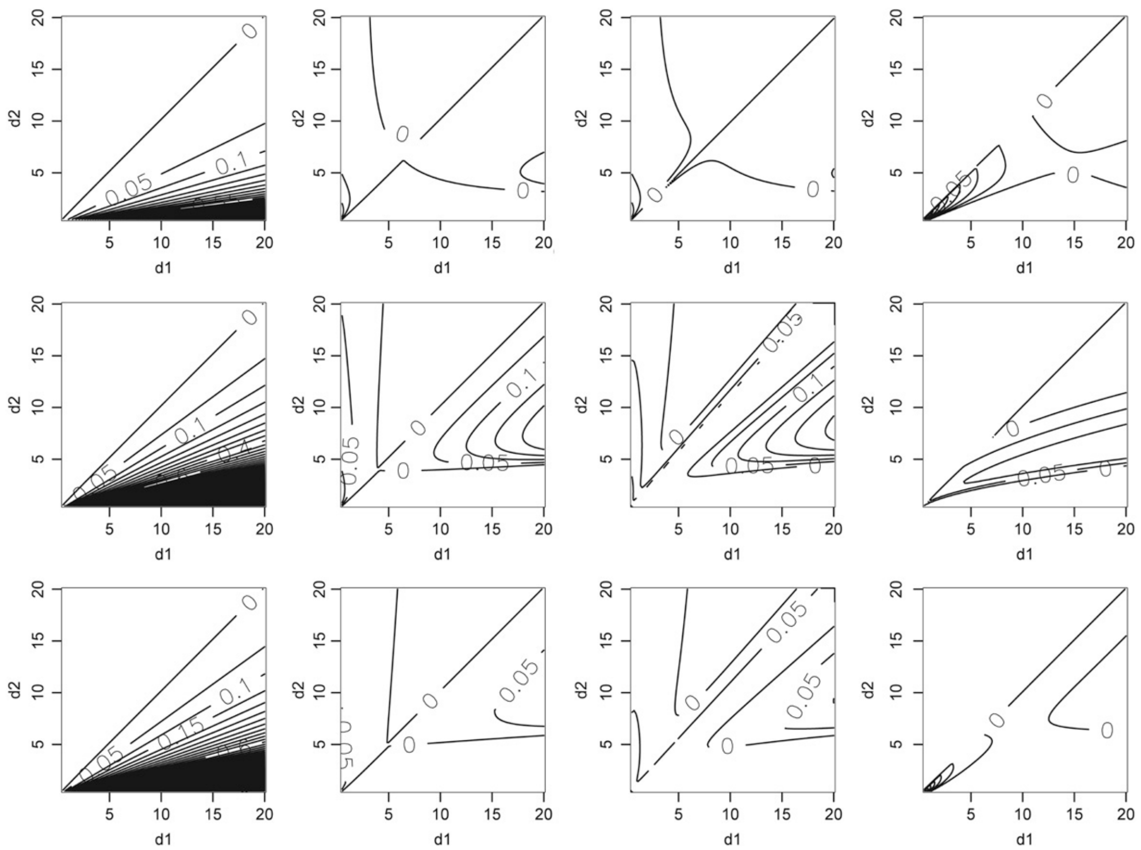

Fig. 9. Deviations from the NAI model predicted by models STE (first column), SEP (second column), LUC (third column) and UNR (fourth column) for three individuals of the distance experiment (subject 1 — first row, subject 3 — second row, subject 7 — third row).

Lugano, Università Ca' Foscari, Venice and Ente Luigi Einaudi, Rome. The work was supported by a grant from MIUR (Ministero dell'Istruzione, dell'Università e della Ricerca).

Appendix. Estimation algorithms

The estimation of models of Section 4.2 can be quite daunting in practice. Therefore, in this Appendix, we propose some algorithms based on iterative procedures. A rationale behind these estimation algorithms can be obtained through the Theorems in Jensen, Johansen, and Lauritzen (1991).

In the case with SEP regression part, if $\left(\phi_{(i)}, \gamma_{(i)}, \sigma^2_{(i)}\right)$ are the parameter values available at step i, we want to compute:

$$\gamma_{(i+1)} = \arg\max_{\gamma} \ell\left(\phi_{(i)}, \gamma, \sigma^2_{(i)}\right),$$

$$\phi_{(i+1)} = \arg\max_{\phi} \ell\left(\phi, \gamma_{(i+1)}, \sigma^2_{(i)}\right),$$

$$\sigma^2_{(i+1)} = \arg\max_{\sigma^2} \ell\left(\phi_{(i+1)}, \gamma_{(i+1)}, \sigma^2\right).$$

The update of γ is performed through numerical optimization; when $\gamma_{(i+1)}$ is fixed, $\phi_{(i+1)}$ can be computed through least squares. Since optimization with respect to γ is a very demanding task, the whole algorithm for the model with $(\phi, \gamma) \in \mathbb{R}^L \times \mathbb{R}^N$ starts at the final estimates for the model with $(\phi, \gamma) \in \mathbb{R}^L \times \mathbb{R}^{N-1}$ (the last element of γ is put equal to 0). As concerns LUC models, the method is analogous to the previous one but in this case the least squares step is replaced by numerical optimization because of the nonlinear nature of the function w^{-1}.

References

Aczél, J., & Luce, R. D. (2007). A behavioral condition for Prelec's weighting function on the positive line without assuming $W(1) = 1$. *Journal of Mathematical Psychology*, *51*, 126–129.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723.

Augustin, T. (2006). Stevens' direct scaling methods and the uniqueness problem. *Psychometrika*, *71*, 469–481.

Augustin, T., & Perner, V. (2007). Stevens' ratio estimation procedure and the uniqueness problem empirical evaluation of axiom fundamental to ratio-, interval, and log-interval scale type. In: *Paper presented at the 38th EMPG colloquium*.

Chambaz, A. (2006). Testing the order of a model. *The Annals of Statistics*, *34*, 1166–1203.

Chechile, R. A. (2003). Mathematical tools for hazard function analysis. *Journal of Mathematical Psychology*, *47*, 478–494.

Cutting, J. E. (2000). Accuracy, scope, and flexibility of models. *Journal of Mathematical Psychology*, *44*, 3–19.

Ellermeier, W., & Faulhammer, G. (2000). Empirical evaluation of axioms fundamental to Stevens' ratio-scaling approach. I. Loudness production. *Perception Psychophysics*, *62*, 1505–1511.

Fuss, M., & McFadden, D. (1978). *Production economics: A dual approach to theory and application*. Amsterdam: North-Holland.

Gouriéroux, C., Monfort, A., & Trognon, A. (1984a). Pseudo maximum likelihood methods: Theory. *Econometrica*, *52*, 681–700.

Gouriéroux, C., Monfort, A., & Trognon, A. (1984b). Pseudo maximum likelihood methods: Applications to Poisson models. *Econometrica*, *52*, 701–720.

Grünwald, P. (2000). Model selection based on minimum description length. *Journal of Mathematical Psychology*, *44*, 133–152.

Hey, J. D. (2005). Why we should not be silent about noise. *Experimental Economics*, *8*, 325–345.

Hollands, J. G., & Dyre, B. P. (2000). Bias in proportion judgements: The cyclical power model. *Psychological Review*, *107*, 500–524.

Jensen, S. T., Johansen, S., & Lauritzen, S. L. (1991). Globally convergent algorithms for maximizing a likelihood function. *Biometrika*, *78*, 867–877.

Karabatsos, G. (2006). Bayesian nonparametric model selection and model testing. *Journal of Mathematical Psychology*, *50*, 123–148.

Luce, R. D. (1997). Several unresolved problems of mathematical psychology. *Journal of Mathematical Psychology*, *41*, 79–97.

Luce, R. D. (2000). *Utility of gains and losses: Measurement-theoretic and experimental approaches*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Luce, R. D. (2001a). Reduction invariance and Prelec's weighting functions. *Journal of Mathematical Psychology*, *45*, 167–179.

Luce, R. D. (2001b). A way to blend Fechner and Stevens. In E. Sommerfeld, R. Kompass, & T. Lachmann (Eds.), *Fechner Day 2001. Proceedings of the seventeenth annual meeting of the international society of psychophysics* (pp. 28–35). Pabst Science Publishers.

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. (2005). Measurement analogies: Comparisons of behavioral and physical measures. *Psychometrika*, *70*, 227–251.

Michell, J. (1999). *Measurement in psychology: A critical history of a methodological concept*. Cambridge, England: Cambridge University Press.

Myung, I. J., Foster, M. R., & Browne, M. W. (2000). Guest editors' introduction: Special issue on model selection. *Journal of Mathematical Psychology*, *44*, 1–2.

Narens, L. (1996). A theory of ratio magnitude estimation. *Journal of Mathematical Psychology*, *40*, 109–788.

Narens, L. (2002). The irony of measurement by subjective estimations. *Journal of Mathematical Psychology*, *46*, 769–129.

Narens, L. (2006). Symmetry, direct measurement, and Torgerson's conjecture. *Journal of Mathematical Psychology*, *50*, 290–301.

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). A method of selecting among computational models of cognition. *Psychological Review*, *109*, 472–491.

Prelec, D. (1998). The probability weighting function. *Econometrica*, *66*, 497–527.

Rissanen, J. (1978). Modelling by shortest data description. *Automatica*, *14*, 465–471.

Saaty, T. L. (1980). *The analytic hierarchy process*. New York: McGraw-Hill.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*, 461–464.

Steingrimsson, R., & Luce, R. D. (2005a). Evaluating a model of global psychophysical judgments-I: Behavioral properties of summations and productions. *Journal of Mathematical Psychology*, *49*, 290–307.

Steingrimsson, R., & Luce, R. D. (2005b). Evaluating a model of global psychophysical judgments-II: Behavioral properties linking summations and productions. *Journal of Mathematical Psychology*, *49*, 308–319.

Steingrimsson, R., & Luce, R. D. (2006). Empirical evaluation of a model of global psychophysical judgments: III. A form for the psychophysical function and perceptual filtering. *Journal of Mathematical Psychology*, *50*, 15–29.

Steingrimsson, R., & Luce, R. D. (2007). Empirical evaluation of a model of global psychophysical judgments: IV. Forms for the weighting function. *Journal of Mathematical Psychology*, *51*, 29–44.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, *103*, 677–680.

Stevens, S. S. (1951). *Handbook of experimental pyschology*. New York: Wiley.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, *64*, 153–181.

Stevens, S. S. (1975). *Psychophysics: Introduction to its perceptual, neural, and social prospects*. New York: Wiley.

Tversky, A., & Fox, C. R. (1995). Weighing risk and uncertainty. *Psychological Review*, *102*, 269–283.

Vapnik, V. N. (1995). *The nature of statistical learning theory*. Berlin: Springer-Verlag.

von Neumann, J., & Morgenstern, O. (1944). *The theory of games and economic behavior*. Princeton: Princeton University Press.

Wagenmakers, E.-J., & Waldorp, L. (2006). Editors' introduction. *Journal of Mathematical Psychology*, *50*, 99–100.

Wasserman, L. (2000). Bayesian model selection and model averaging. *Journal of Mathematical Psychology*, *44*, 43–19.

Zimmer, K., Luce, R. D., & Ellermeier, W. (2001). Testing a new theory of psychophysical scaling: Temporal loudness integration. In E. Sommerfeld, R. Kompass, & T. Lachmann (Eds.), *Fechner Day 2001. Proceedings of the seventeenth annual meeting of the international society of psychophysics*. Pabst Science Publishers.

Zimmer, K. (2005). Examining the validity of numerical ratios in loudness fractionation. *Perception & Psychophysics*, *67*, 569–579.

Zucchini, W. (2000). An introduction to model selection. *Journal of Mathematical Psychology*, *44*, 41–61.