ARTICLE IN PRESS

Journal of Mathematical Psychology ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

What are we estimating when we fit Stevens' power law?

Michele Bernasconi [a], Raffaello Seri [b,c,*]

[a] Dipartimento di Economia, Università "Ca' Foscari", Cannaregio 873, I-30121 Venezia, Italy

[b] Dipartimento di Economia, Università dell'Insubria, Via Monte Generoso 71, I-21100 Varese, Italy

[c] Center for Nonlinear and Complex Systems, Università dell'Insubria, Via Valleggio 11, I-22100 Como, Italy

HIGHLIGHTS

- Estimates in Stevens' power laws often display sensitivities to experimental design.
- These so-called 'contextual effects' concern range, location and averaging.
- This paper links them with the separable representation model of Luce and Narens.
- Theoretical results are illustrated using data from papers of R. Duncan Luce.

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Separable representations
Stevens' model
Psychophysical experiments

ABSTRACT

Estimates of the Stevens' power law model are often based on the averaging over individuals of experiments conducted at the individual level. In this paper we suppose that each individual generates responses to stimuli on the basis of a model proposed by Luce and Narens, sometimes called separable representation model, featuring two distinct perturbations, called psychophysical and subjective weighting function, that may differ across individuals. Exploiting the form of the estimator of the exponent of Stevens' power law, we obtain an expression for this parameter as a function of the original two functions. The results presented in the paper help clarifying several well-known paradoxes arising with Stevens' power laws, including the range effect, i.e. the fact that the estimated exponent seems to depend on the range of the stimuli, the location effect, i.e. the fact that it depends on the position of the standard within the range, and the averaging effect, i.e. the fact that power laws seem to fit better data aggregated over individuals. Theoretical results are illustrated using data from papers of R. Duncan Luce.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

A large part of the success of modern psychophysics is certainly due to the versatility and simplicity of Stevens' psychophysical law (Stevens, 1946, 1951, 1957): namely, the notion that sensation magnitude can be described as a power function of stimulus intensity. This idea has in particular been popularized by Stevens through the application of different direct measuring methods able to reveal the law and to provide estimates of the exponent of the power model in several sensory domains (see the posthumous book by Stevens, 1975, for a comprehensive survey).

In possibly the simplest of Stevens' direct measuring techniques, known as ratio magnitude estimation, a subject is asked to compare two stimuli, a comparison stimulus d_1 and a reference standard d_2, and to state in what proportion p the stimuli are with respect to each other. According to the power law, the following ratio scale of subjective intensities holds:

$$\left(\frac{d_1}{d_2}\right)^{\beta} = p. \tag{1}$$

Hence, the exponent β of the law can be easily estimated from a series of trials, in which d_1 varies between trials while the standard d_2 can both be kept constant or let vary as well (see below). Magnitude estimation and the complementary approach of magnitude production, in which the standard d_2 is given and the subject is asked to adjust d_1 to a prescribed ratio p, are still widely used (recent surveys, examples, discussions in several fields of psychophysics in Fagot, 2011; Lawless & Heymann, 2010; Lim, 2011; Masin, 2014; Shofner & Selas, 2002).

* Corresponding author at: Dipartimento di Economia, Università dell'Insubria, Via Monte Generoso 71, I-21100 Varese, Italy.

E-mail addresses: bernasconi@unive.it (M. Bernasconi), raffaello.seri@uninsubria.it (R. Seri).

http://dx.doi.org/10.1016/j.jmp.2016.04.007
0022-2496/© 2016 Elsevier Inc. All rights reserved.

Please cite this article in press as: Bernasconi, M., & Seri, R.What are we estimating when we fit Stevens' power law?. Journal of Mathematical Psychology (2016), http://dx.doi.org/10.1016/j.jmp.2016.04.007

Notwithstanding the ample and enduring success, it is however also well-known that Stevens' power law suffers from both empirical and theoretical shortcomings.

Under the empirical perspective, an ever growing experimental literature has documented several inconsistencies and puzzles of the law (Luce & Krumhansl, 1988, provide a classical survey; Zwislocki, 2009, Chapter 2, a recent one). Among them, a list of studies conducted since the early development of the approach documented that the exponent β of Stevens' law is sensitive to the experimental parameters and design of the investigations. In Section 2 we provide an account of the above earlier evidence.

Under the theoretical perspective, Stevens never provided a formalized theory of measurement. Successive scholars have worked in the tradition of the representational theory of measurement (in the three classic volumes of *Foundations of Measurement*, Vol. I by Krantz, Luce, Suppes, & Tversky, 1971, Vol. II by Suppes, Krantz, Luce, & Tversky, 1989, and Vol. III by Luce, Krantz, Suppes, & Tversky, 1990) to give more proper mathematical and philosophical foundations to the notion of psychophysical measurement (earlier works in Krantz, 1972, Luce, 1959, 1990, Shepard, 1981, and several others quoted in Luce, 1996, and Luce & Krumhansl, 1988).

Mainly, a recurrent criticism of mathematical psychologists was that neither the power law nor Stevens' method of direct estimation were derived from primitive behavioral conditions, or axioms, which could be independently expressed and tested. The development of the representational theory of measurement drawn in the three volumes of the *Foundations* represented in such a respect a revolution (see, e.g., Steingrimsson, this issue, in particular the Introduction).

An important more recent achievement developed in this stream of literature includes the axiomatization of various novel theories which comprehend Stevens' model as a special case (Augustin, 2006, 2010; Luce, 2002, 2004, 2008; Narens, 1996, 2002, 2006). Following a terminology introduced by Luce (2002), we say formally that a psychological scale of subjective intensities can be represented in a *separable form* if there exist a *psychophysical function* ψ and a *subjective weighting function* W such that p is in the following relation with d_1 and d_2:

$$\frac{\psi\left(d_1\right)}{\psi\left(d_2\right)} = W\left(p\right). \tag{2}$$

Eq. (2) incorporates the notion that various and independent distortions may occur both in the assessment of subjective intensities and in the determination of subjective ratios. Stevens' power model in Eq. (1) is obviously a particular case of separable representation, holding when W can be represented as the identity function and ψ is a power function.

Several experiments have given substantial support to separable representations, but not to the restrictions implied by the power law (Augustin & Maier, 2008; Bernasconi, Choirat, & Seri, 2008; Ellermeier & Faulhammer, 2000; Steingrimsson, 2009; Steingrimsson and Luce, 2005a, 2007; Zimmer, 2005).

Narens (1996) has obtained Eq. (2) in an article in which he formalized Stevens' magnitude methods in terms of axiomatic measurement theory.[1] Luce (2002, 2004, 2008) has axiomatized Eq. (2) as a special case of a global psychophysical theory of intensity perception. The theory has been shown to be general enough to be extended theoretically in many directions (Luce, 2012a, 2012b, 2013; Luce, Steingrimsson, & Narens, 2010). In the course of the paper we will give some accounts of the properties predicted and of the results obtained. We will also refer to some earlier nonaxiomatic approaches that have considered forms similar to separable representations providing experimental results in their support (as in, e.g., Birnbaum, 1980; Birnbaum & Elmasian, 1977 and Birnbaum & Mellers, 1978).

Given the inconsistencies against Stevens' power law anticipated above and described in detail below, a natural question arises about what are the many people that keep fitting the power law actually estimating.[2] This is what we clarify in this paper. In order to conduct the analysis we use an empirical working model developed from Bernasconi et al. (2008). In that previous paper, we estimated several variants of separable representation models and we saw which one performed best. In order to do so, we rewrote the model in Eq. (2) using a log–log transformation, we added a Fechnerian error term, and then expanded the log-transformed functions in polynomials of the various separable models. In this paper we develop a similar model, but do not use polynomial expansions. This allows for greater flexibility, which we use to reinterpret the parameters associated with the power law. In particular, by exploiting the form of the estimator of β, we obtain an expression for this parameter that we use to predict several facts, documented in the earlier literature, on the sensitivity of β to the experimental design, including the so called 'range effect', 'location effect' and 'averaging effect', which we illustrate with data digitized from two classical experiments of Green and Luce (1974) and Luce and Mo (1965).

We start in Section 2 with a review of the earlier evidence on Stevens' power law. In Section 3 we present the empirical working model and its relations to the literature. Results are in Sections 4 and 5. In Section 4 we apply the model to the study of ratio magnitude estimation with a standard for a single individual and we provide a theoretical account of range and location effects. In Section 5 we show how conducting the analysis with data averaged across individuals, rather than at the individual level, leads to the averaging effect. Section 6 concludes with a summary of the main findings and a discussion of extensions and implications of the approach. The proofs are gathered in the Appendix.

2. Contextual effects

In the following we are going to present a series of phenomena arising in magnitude estimation, as well as in related forms of scaling, that describe some deviations with respect to Stevens' power law that are often observed in the data. These are sometimes called 'context' or 'contextual effects' in the literature.

One of the most commonly observed contextual effects is the so-called 'range effect', i.e. the fact that for larger ranges of δ_1 (here and in the following $\delta_i = \ln d_i$) we expect β to be smaller. Early evidence on the effect was observed by Engen (1956), Engen and Levy (1958) and Künnapas (1960, 1961). Various experiments conducted in the following years confirmed the same conclusion (surveys and examples in, e.g., Bonnet, 1969a, 1969b; Teghtsoonian, 1971, 1973; Vincent, Brown, Markley, & Arnoult, 1968). Poulton (1968) reviews the literature up to that date and states that the range of stimuli "alone accounts for about $\frac{1}{3}$ of the variance in S. S. Stevens' table of exponents" (p. 1). It should also be noted that most of the previous evidence is based on data grouped over individuals, while for individuals the situation is less clear. The individual-level results in Pradhan and Hoffman (1963) do not seem to support this contextual effect (see, however, below for more discussion). On the other hand, always at the individual

[1] Therein he also asserts something that is slightly more general than our Eq. (1), namely $W(p) = p^k$ with $k > 0$, and $W(1) = 1$ (more detail in Section 3 below and in Steingrimsson & Luce, 2007, in particular their Section 2.1.1).

[2] For example, in addition to the surveys quoted above, in a systematic literature search Kornbrot (2014) identifies 193 items with "magnitude estimation" in the title published between 2000 and 2013 and remarks that just two studies have estimated psychophysical functions more general than Stevens' power law.

Please cite this article in press as: Bernasconi, M., & Seri, R.What are we estimating when we fit Stevens' power law?. Journal of Mathematical Psychology (2016), http://dx.doi.org/10.1016/j.jmp.2016.04.007

level, Green, Luce, and Smith (1980, p. 485) observe "a slight, but systematic, decrease of the slope with range". In any case, the fact that individual-level studies yield no overwhelming support for the range effect should not be taken as evidence that this is stronger for averaged than for individual data, but only as a clue that the literature reporting individual data is quite limited.

Another contextual effect is what is sometimes called 'location effect'. The exponent tends to be larger when the standard is in the center of the range of δ_1, near to or immediately above the expectation $\mathbb{E}\delta_1$,[3] than when it moves toward either extreme of the range. Engen and Levy (1955) considered the case of ratio magnitude estimation with a standard (and also without a standard, but we will not address that case here) and reported median values across a sample of 60 individuals for three choices of the standard (low, middle and high) and two different continua (weight and brightness). They found that the exponent is higher for a middle value of the standard. The individual-level results in Pradhan and Hoffman (1963) also confirm the effect. Further evidence on grouped data is reported by Macmillan, Moschetto, Bialostozky, and Engel (1974), which show that when $\mathbb{E}\delta_1 = \delta_2$ (their case MS) the exponent is larger than when either $\mathbb{E}\delta_1 \ll \delta_2$ (their case LS) or $\mathbb{E}\delta_1 \gg \delta_2$ (their case SS). The effect has also been observed in Marks (1988, p. 522), Ahlström and Baird (1989), Fagot and Pokorny (1989) and Kowal (1993, p. 558). The last reference draws an unprecedented connection between the location and the range effects, and suggests that the former may be more robust than the latter (Kowal, 1993, p. 561) and may even have a role in determining it (p. 555).

Another contextual effect concerns the behavior of the responses for extreme values of the stimulus δ_1, i.e. values near the threshold of detection or in a saturation area. We will not consider this effect in the following because, on the one hand, Stevens widely acknowledged its importance, in some cases incorporating it in the law itself, and, on the other hand, this phenomenon generally becomes evident only for stimuli outside the ranges considered in most applications. Nevertheless, we discuss it as it represents yet another case of lack of fit of Stevens' power law that separable representations automatically take into account. The distortion in the lower portion of the power law due to the presence of a threshold of detectability (see Laming, 1997, Section 3.1) was acknowledged by Stevens himself as a relevant phenomenon (see Stevens, 1958; Stevens & Stevens, 1960). If the threshold is d_{10}, the law is described by the equation $p = k\,(d_1 - d_{10})^{\theta}$, where k depends on the standard d_2. The formula was proposed in Ekman (1958, p. 288) (see also Bonnet, 1969b, p. 248, MacKay, 1963, p. 1213). Piéron (1963, p. 46) traces back the origin of this formula to Helmholtz. Poulton (1968) reviews the literature up to that date and remarks that d_{10} can be larger than the threshold if the power law needs to be corrected for nonlinearity. This shows that, while d_{10} was introduced to take into account the threshold of detectability, it has since been used to correct a general lack of fit of Stevens' power law model on the lower part of the range of δ_1.

We now move to the discussion of the so-called 'averaging effect', namely the tendency of data averaged over individuals to provide a better fit to Stevens' power laws. Despite the paucity of individual data, this is a quite recurrent trait of the evidence on the power law. Indeed, averaging over individuals has always been a distinctive feature of the original Stevens' method. Stevens (1975, pp. 23–24) describes the procedure through which he arrived, in the summer of 1953, to consider averaging across individuals: "I had no assurance that it would be proper to average the data from different observers. The general agreement among the responses of the first few observers persuaded me that I had probably hit upon a promising method". As a result of this finding, he started to devote his attention to measures of central tendency (Stevens, 1955, 1956, 1957).

After these early years, several authors tried to apply Stevens' model to individual data, with mixed results. Ekman, Hosman, Lindman, Ljungberg, and Akesson (1968) attribute to Künnapas (1958a,1958b) the first confirmation of the power law at the level of the individual, but acknowledge that the power law had been refuted at the same level by others, including Luce and Mo (1965) and Pradhan and Hoffman (1963). Gulliksen (1959, p. 189) proposed a test of the hypothesis that data collected from a sample of individuals obey the power law. Jones and Marcus (1961) tested through ANOVA the hypothesis that the same power law holds for a sample of individuals and rejected it. Analyses fitting individual data were also conducted for models obtained as a two-stage version of Stevens' power law, where the subjective number system is itself treated as a psychophysical power function of arithmetic numbers (Curtis, Attneave, & Harrington, 1968; Ramsay, 1979; Rule and Curtis, 1973, 1977). The studies provided further evidence of large differences between the functions fitted for the individuals. Nevertheless, Teghtsoonian (1973, p. 3) affirmed that "there is now considerable evidence that [the power relation] is not an artifact of pooling data over many observers but is evident in the behavior of individual observers". More recently, Steingrimsson and Luce (2006) have shown that the power form may not be the best description for all individuals.

Stevens limited his analysis to the average of the group and did not devote the same attention to measures of variability around this central tendency (see Piéron, 1963, p. 46). Moreover, he did not consider whether averaging across individuals is an appropriate operation when individuals are heterogeneous. One of the first papers to consider the latter question explicitly is Sidman (1952). This line of research was pursued by Estes (1956) (see, more recently, Estes & Maddox, 2005) who, after analyzing several categories of functions, ended up justifying aggregation even for functions that are modified in form by the operation of averaging. What seems to lack from this stream of research is the acknowledgment that the individual functions could be affected by deviations, with respect to power laws, of such an intra-individual complexity and an inter-individual variety that averaging hides both of them.

The latter point is explicitly recognized in Green and Luce (1974, p. 291), who argue against aggregation (see also Luce, 1995, p. 4): "one should not average over observers unless one is quite sure of the functional form of the data, so that the true form will not be distorted in the averaging process".[4] As we will show below theoretically and using the experimental evidence from Green and Luce (1974) and Luce and Mo (1965), whether individual data conform or not to Stevens' model, there are good mathematical reasons for which the fit improves after averaging across individuals.

3. The working model

Stevens' power model is a special case of a so-called separable representation shown in Eq. (2). As indicated above, the separable representation has arisen in several modeling efforts. Narens (1996) presented an axiomatization of Stevens's ratio magnitude

[3] Here and in the following, $\mathbb{E}$ denotes the expectation of a random variable. Even though δ_1 is not a random variable, it is still possible to take its expectation through the argument explained at the beginning of the Appendix.

[4] In reference to learning curves, Skinner (1958, p. 99) makes a similar point: "The curves we get cannot be averaged or otherwise smoothed without destroying properties which we know to be of first importance". Further discussions which criticize aggregation are in, e.g., Laming (1997, p. viii) and Yost (1981, p. 212).

production. He showed that a strong property, called multiplicativity, is implied by Stevens' implicit assumptions that the method delivers directly the psychophysical function of interest. He also developed a commutative property and showed that if neither it nor the multiplicative property held, then magnitude production does not provide ratio-scale measurements. However, if multiplicativity failed, but the commutativity property held, then the measurement was on a sub-scale of a ratio scale. Commutativity asserts that the result of, say, first doubling ($p = 2$) a stimulus and then tripling ($p = 3$) should be equivalent to first tripling and then doubling. Multiplicativity asserts the stronger requirement that the result should be equivalent to a single magnitude production instruction of six times ($2 \times 3 = 6$) the reference stimulus.

Luce (2002, 2004, 2008) derived separable representations as a special case of a theory of global psychophysics originally developed from 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 intervals being related as some proportion. Steingrimsson (this issue) provides a thorough summary of Luce's model of global psychophysics, including the more recent topics covered by the theory.

A part of the research concentrated on the form of the psychophysical function ψ and the subjective weighting function W. Clearly, Stevens' model in Eq. (1) holds when ψ is a power function and W is the identity or it takes the slightly more general power specification $W(p) = p^k$, with $k > 0$ and $W(1) = 1$. In particular, for Eq. (1) to hold, multiplicativity must hold and this assumption implies $W(1) = 1$ because $W(1) \cdot W(q) = W(1 \cdot q) = W(q)$ iff $W(1) = 1$. However, Luce (2005) pointed out, initially as a subtlety, that when multiplicativity fails, $W(p)$ may still be a power function with $W(1) \neq 1$. In fact, removing the condition $W(1) = 1$ leads to show that commutativity is all that is needed to obtain ratio scalability. Steingrimsson and Luce (2007) presented a careful analysis of the various conditions, providing behavioral equivalents for $W(p)$ being a power function or a so-called Prelec's (1998) exponential representation, in both cases with and without the restriction $W(1) = 1$. The experimental evidence reported in the same paper rejected the latter restriction, showing that most respondents either satisfied the general power form (with $W(1) \neq 1$) or the general Prelec's form (also with $W(1) \neq 1$). Therefore, as multiplicativity has been rejected in every domain in which it has been evaluated (loudness, brightness, contrast, perception of time, size of circles, and more), also $W(1) = 1$ has been rejected (as a general rule) empirically. Augustin (2006, 2010) presented theoretical extensions of Narens (2006) giving support to the fact that $W(p)$ is either a power or a logarithmic function.

All the above approaches were developed for stimuli varying on a single attribute. Luce et al. (2010) further expanded the theory to stimuli varying in more than one attribute and called the underlying axiom cross-dimensional commutativity. In the same paper they tested the property with favorable results for perception of loudness–pitch pairs. A subsequent paper of the same authors also found support for the property in the case of luminance–hue pairs (Steingrimsson, Luce, & Narens, 2012).

Experiments have also been conducted to evaluate the form of ψ. Steingrimsson and Luce (2006) develop an analysis of Luce's global psychophysical theory for inherently binary stimuli (e.g., to the two ears or to the two eyes). They evaluated the theory for binaural loudness providing evidence supporting the notion that ψ is a power function. Luce (2012a) expands the theory to unary sensory intensities, that is for sensations which are inherently one dimensional, like taste, vibration, shock, force, preference for money, etc. He in particular observed that his original theory, formulated in terms of joint presentation of two sensory organs, had to be reformulated in the case of unary sensations because the matching operation does not work in the same way in the two situations. So, in some sense, his original theory did not cover the unary case. He also found that an exponential form could offer an improvement over the power law as a description of psychophysical data for some unary domains, including some of the data reported in Stevens (1975).

Separable forms have also similarities with earlier approaches developed without axiomatic foundations, but nevertheless based on the similar idea that two mathematical transformations affect subjective measurements. One approach is the two-stage version of Stevens' power law quoted above (Attneave, 1962; Curtis et al., 1968; Rule and Curtis, 1973, 1977). A more general model is a theory of judgment functions (e.g., Birnbaum, 1980; Birnbaum & Elmasian, 1977 and Birnbaum & Mellers, 1978). There, W (or better its inverse W^{-1}, see below) is replaced by a function J_R introduced directly as a monotonic judgmental transformation. The model has been used to conduct several analyses, including one of the first tests of a well-known implication of Stevens' power law, stronger than the power law itself, that if the physical ratios are constant then the ratio judgments should be constant.[5] The property was systematically violated (Birnbaum & Elmasian, 1977).[6]

A characteristic of the approach is that it considers the bias due to the judgment functions in the perspective of various scaling methods and related psychophysical scales. Indeed, in the approach the ratio model was often connected to a subtractive model for category scaling, relating overt rated differences to subjective differences. Various analyses comparing the two models demonstrated that while magnitude estimations and category ratings of the same stimuli are not linearly related, nevertheless they were monotonically related for a number of continua (e.g., Birnbaum, 1980, 1982; Birnbaum, Anderson, & Hynan, 1989; Birnbaum & Elmasian, 1977 and Birnbaum & Veit, 1974). This is quite interesting also from the perspective of axiomatic separable representations: on the one hand, it further shows the relevance of the behavioral properties underlying separable representations to address issues which go beyond Stevens' power law;[7] on the other hand, it suggests that the machinery used below for the empirical analysis of a ratio model may in principle be also applied to a subtractive model. This is left for future research.

Given the uncertainty about the forms of ψ and W, we now propose an analysis of Eq. (2) which does not impose any specific form for the two functions. We only assume that W is monotonic, so that it is invertible and we can write:

$$p = W^{-1}\left(\frac{\psi(d_1)}{\psi(d_2)}\right). \tag{3}$$

We allow the function W to differ for $p > 1$ and $p < 1$; and we do not restrict $W(1) = 1$. More importantly, our approach includes

---

5 This follows because, under the power law, $\frac{d_1'}{d_2'} = \frac{d_1}{d_2} \iff p' = p$. The latter equivalence holds more generally, applying also to any model like, e.g., $W(p) = \left(\frac{d_1}{d_2}\right)^{\beta}$. This is called Stevens' generalized model or STG in Bernasconi et al. (2008).

6 Analyses also showed that J_R is not invariant, but ratio judgments (in addition to the contextual effects described in Section 2) may change depending on responding procedures and other details of the experiments (including the range and distribution of the examples used in the instructions of ratio judgment tasks; Hardin & Birnbaum, 1990).

7 Indeed, undertaking from the axiomatic point of view a question somehow similar to the earlier literature, Luce (2012b) shows that the separable form for magnitude production, when applied to fractionations and equisection scaling, is inconsistent with Torgerson's (1961) conjecture, namely the ides that respondents fail to distinguish subjective differences from subjective ratios. This because the conjecture implies that W is the identity function (which, as indicated above, is however firmly rejected by data).

Please cite this article in press as: Bernasconi, M., & Seri, R.What are we estimating when we fit Stevens' power law?. Journal of Mathematical Psychology (2016), http://dx.doi.org/10.1016/j.jmp.2016.04.007

an explicit error term in a stochastic version of Eq. (2). In order to illustrate the approach, we define:

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

The constraints $W(1) = 1$ and $\psi(1) = 1$,[8] when satisfied, become respectively $w(0) = 0$ and $\Psi(0) = 0$. In this new parameterization, we can write the representation as:

$$\pi = w^{-1}\left[\Psi\left(\delta_1\right) - \Psi\left(\delta_2\right)\right]. \quad (4)$$

Following Bernasconi et al. (2008) and Bernasconi, Choirat, and Seri (2010, 2011), we now add a Fechnerian error term ε to the model:

$$\pi = w^{-1}\left[\Psi\left(\delta_1\right) - \Psi\left(\delta_2\right)\right] + \varepsilon. \quad (5)$$

Even if $\delta_1 = \ln d_1$ is not a random variable, we will nevertheless take expectations and variances with respect to δ_1. A justification for this is at the beginning of the Appendix. On the other hand, the error term ε is supposed to be a random variable and we will repeatedly need its expectation $\mathbb{E}(\varepsilon)$ and variance $\mathbb{V}(\varepsilon)$ as well as its conditional expectation $\mathbb{E}(\varepsilon|\cdot)$ and conditional variance $\mathbb{V}(\varepsilon|\cdot)$. Therefore, from a mathematical point of view, we require that $\mathbb{E}(\varepsilon|\delta_1) = 0$. This is much weaker than, but implied by, the requirement that the error term ε is independent of δ_1. As an example, $\mathbb{E}(\varepsilon|\delta_1) = 0$ is still compatible with heteroskedasticity, i.e. the situation in which $\mathbb{V}(\varepsilon|\delta_1)$ is not constant, while independence is not.

The error term is justified in various ways. In fact, we recall that Eq. (2) "[is] about idealized situations and [does] not involve considerations of error" (Narens, 1996, p. 109). Rarely, however, have there been discussions about how to properly integrate a theory of errors in measurement theory (Luce, 1997, p. 81).[9] The Fechnerian model, while perhaps very direct and simple,[10] is nevertheless able to account for several sources of random noise that may affect subjective measurement data, including those due to lapses of reason or concentration, states of mind, trembling, rounding effects, and computational mistakes. Moreover it is somehow consistent with the notion that there is a well-defined structure underlying separable representations, which however people apply in actual experimental tasks with errors. This introduces naturally some variability in the responses even when the same stimuli are given (an issue about which we will say something below).

Conversely, it is important to remark that while the exact form of Eq. (2) (or Eq. (4)) applies equally to experiments of magnitude estimation or magnitude production, the stochastic version in Eq. (5) implies that $p = \exp(\pi)$ is in fact the subject's answer of a ratio magnitude estimation task to which a multiplicative error term (additive in the log–log transformation) is appended.[11]

4. Ratio magnitude estimation with a standard for a single individual

Suppose that the stimulus δ_1 and the response π are linked by the following model:

$$\pi = f\left(\delta_1\right) + \varepsilon,$$

where $f(\delta_1) := \mathbb{E}(\pi|\delta_1)$. This is equivalent to $\mathbb{E}(\varepsilon|\delta_1) = 0$. We do not yet suppose that f comes from a separable representation, but if this is the case, $f(\cdot) := w^{-1}[\Psi(\cdot) - \Psi(\delta_2)]$. The function f can be estimated in a quite precise way using the data digitized from Luce's papers (see below). Moreover, the exponent estimated in a Stevens' power law, β, can be linked to the function f through the formulas contained in our theorems below. As a result, we create a correspondence between the contextual effects in β and the form of the function f as estimated on data from Luce's papers. The support for separable representations comes from the fact that the function f has features that are coherent with separable representations (the inflection point around the standard, the approximate linearity over short ranges, etc.).

We suppose that Stevens' power law model is fitted instead:

$$\pi = \alpha + \beta\delta_1 + u.$$

This model is obtained from Eq. (1) by taking a log–log transformation as in Eq. (4), letting $\alpha = -\beta\delta_2$ and adding an error term u.

Then the following proposition holds.

Proposition 1. *The parameters of Stevens' power law are defined by the moment equalities $\alpha = \mathbb{E}\pi - \beta \cdot \mathbb{E}\delta_1$ and:*

$$\beta \simeq \operatorname{cor}\left(f\left(\delta_1\right), \delta_1\right) \cdot f'\left(\mathbb{E}\delta_1\right) \cdot \left\{1 + \frac{f''(\mathbb{E}\delta_1)}{f'(\mathbb{E}\delta_1)} \cdot \frac{\mathbb{E}(\delta_1 - \mathbb{E}\delta_1)^3}{\mathbb{V}(\delta_1)} + \frac{f'''(\mathbb{E}\delta_1)}{3f'(\mathbb{E}\delta_1)} \cdot \frac{\mathbb{E}(\delta_1 - \mathbb{E}\delta_1)^4}{\mathbb{V}(\delta_1)} + \frac{f''^2(\mathbb{E}\delta_1)}{4f'^2(\mathbb{E}\delta_1)} \cdot \frac{\mathbb{E}(\delta_1 - \mathbb{E}\delta_1)^4 - \mathbb{V}(\delta_1)^2}{\mathbb{V}(\delta_1)}\right\}^{\frac{1}{2}} \quad (6)$$

where $f(\cdot) := \mathbb{E}(\pi|\cdot)$. Moreover, the following equality holds:

$$\operatorname{cor}\left(\pi, \delta_1\right) = \operatorname{cor}\left(f\left(\delta_1\right), \delta_1\right) \cdot \operatorname{cor}\left(\pi, f\left(\delta_1\right)\right). \quad (7)$$

Eq. (6) allows one to say something about the value of β. In order to interpret it correctly, it is wise to use as a guideline Figs. 1 and 2 in Luce and Mo (1965) and Fig. 1 in Green and Luce (1974).[12]

Figs. 1 and 2 in Luce and Mo (1965) display the results of two experiments, respectively concerning weights and sound intensities, in which subjects are asked to compare a stimulus with a fixed standard. Fig. 1 in Green and Luce (1974) displays the results of their Experiment 1, qualified in their paper as a "magnitude estimation with a standard (MES)" task.[13] These figures show a strong variability across subjects, but the picture that comes out is quite clear. Below the standard δ_2, the function $\delta_1 \mapsto w^{-1}[\Psi(\delta_1) - \Psi(\delta_2)]$ appears generally concave. Above the standard δ_2, there seems to be more variability in the form of the curve: for most subjects the curve is first convex for values of δ_1 not too far from δ_2, and for some of them it becomes eventually concave for large δ_1. Around the standard δ_2, the curves often display an inflection point whose salience depends on the strength of the convexity immediately above the standard.

[8] Eq. (2) suffers from an identification problem, namely if ψ satisfies it, so does $k \cdot \psi$. Therefore, if $W(1) = 1$, it is always possible to choose ψ in such a way that $\psi(1) = 1$. However, we will not impose these constraints in the following.

[9] A general approach, alternative to the introduction of an error term, which mathematically represents judgments and preferences with randomness is that of probabilizing directly deterministic measurement structures (Regenwetter & Marley, 2001, and references therein). Marley (1972) is an early model in such a spirit within the magnitude estimation literature.

[10] In particular, the Fechnerian model of errors assumes that people have well-defined systems of values, preferences, judgments, that they apply in actual choices with errors. The model has for example been recently criticized by an approach which treat people systems as inherently stochastic and leads to models in which noise and inconsistencies arise because of the imprecision of people to use the same specification of the theory every time it is used (discussions and references in, e.g., Loomes, 2005 and Myung, Karabatsos, & Iverson, 2005).

[11] Obviously, a similar approach could nevertheless be used to add a Fechnerian error term to the respondent's chosen stimulus and obtain a stochastic version of the model for magnitude production. The analysis of such an alternative model is left for future work.

[12] We signal here that Teghtsoonian, Teghtsoonian, and Baird (1995) have provided an alternative explanation of the data in both papers.

[13] Note that the same paper also contains a "ratio estimation (RE)" experiment in which the standard varies across different administrations of each pair of stimuli.

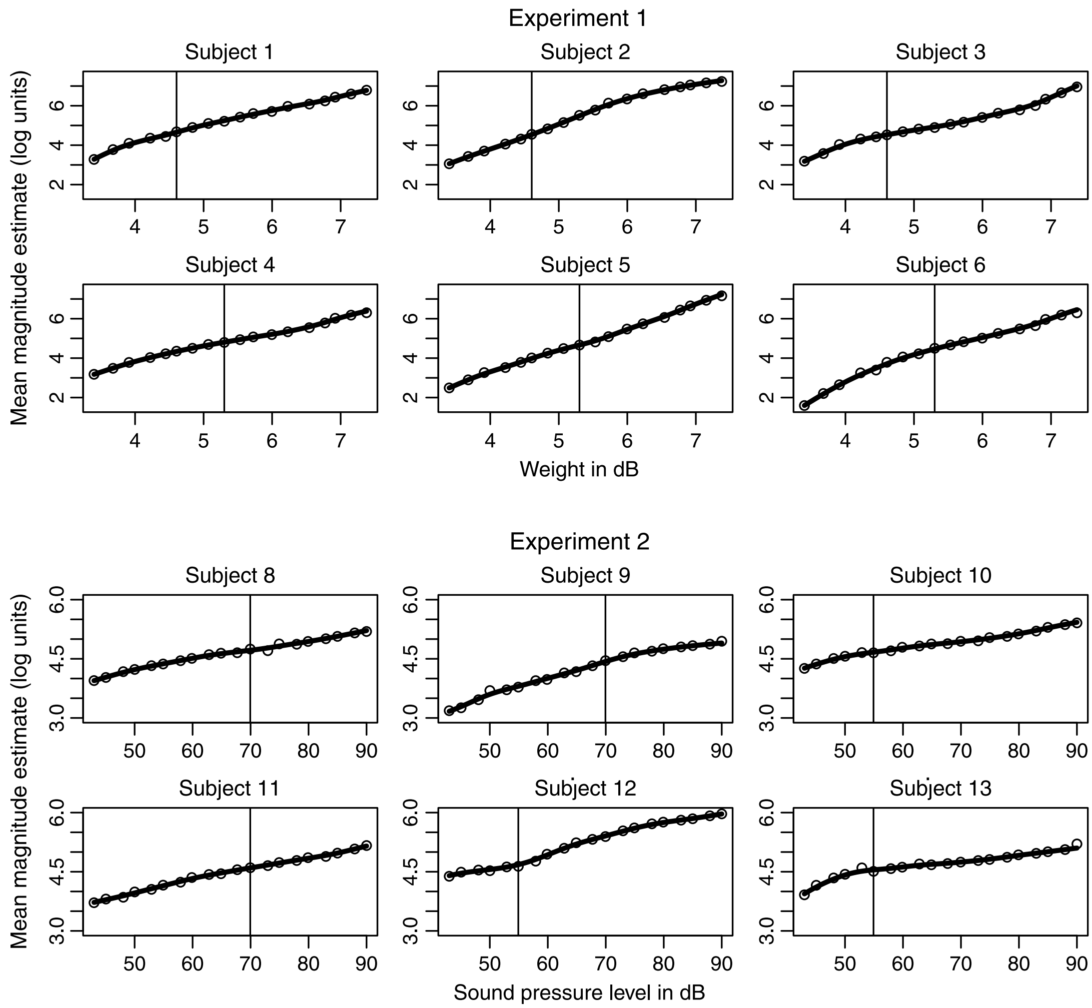

Fig. 1. Function f estimated through local polynomial regression for Experiments 1 and 2 in Luce and Mo (1965); the dots are the points in the original paper, the vertical line is the standard.

In order to ease the interpretation we have digitized Figs. 1 and 2 in Luce and Mo (1965) and, on the basis of the data, we have estimated the functions f, f' and f'' through local polynomial regression (see, e.g., Wand & Jones, 1995, Sections 5.2 and 5.7), using a normal kernel and a bandwidth selected by visual inspection. When looking at Figs. 1–3, the reader should take into account that the estimated functions suffer from bias near the extremes of the range. Despite the huge inter-individual variability, the graphs show that f' is generally decreasing below δ_2 and concave upwards above it, while f'' is increasing below δ_2 and concave downwards above it.

Now, let us return to Eq. (6), starting from the first term, i.e. the correlation between δ_1 and $w^{-1}[\Psi(\delta_1) - \Psi(\delta_2)]$. This measures the linearity of $w^{-1}[\Psi(\delta_1) - \Psi(\delta_2)]$. Therefore, we expect β to be higher when δ_1 varies in an interval in which the function $w^{-1}[\Psi(\delta_1) - \Psi(\delta_2)]$ is roughly linear in δ_1. This holds when both w^{-1} and Ψ are almost linear, but may be true also when w^{-1} and Ψ offset each other, thus yielding an almost linear f. As to the second term, it shows that the exponent depends on the slope of the true curve in a neighborhood of $\mathbb{E}\delta_1$; this implies that if the mean of δ_1 changes, also β may change. The third term depends on the higher moments of δ_1:

- If the distribution of δ_1 is not symmetric, then $\mathbb{E}(\delta_1 - \mathbb{E}\delta_1)^3 \neq 0$ and the sign of $f''(\mathbb{E}\delta_1)$ determines whether the fourth term decreases or not.
- If the distribution of δ_1 is symmetric, then $\mathbb{E}(\delta_1 - \mathbb{E}\delta_1)^3 = 0$. This happens when, say, the values of δ_1 are in arithmetic progression. Then, the fourth term may decrease or increase: note that $\frac{f''^2(\mathbb{E}\delta_1)}{4f'^2(\mathbb{E}\delta_1)} \cdot \frac{\mathbb{E}(\delta_1-\mathbb{E}\delta_1)^4-\mathbb{V}(\delta_1)^2}{\mathbb{V}(\delta_1)} \geq 0$, while $\frac{f'''(\mathbb{E}\delta_1)}{3f'(\mathbb{E}\delta_1)} \cdot \frac{\mathbb{E}(\delta_1-\mathbb{E}\delta_1)^4}{\mathbb{V}(\delta_1)} \gtreqless 0$ according to the behavior of $f'''(\mathbb{E}\delta_1) \gtreqless 0$.

This formula can help explain the range and location effects. First, as concerns the range effect, the linear term δ_1 may be able to better adapt to the function f, i.e. $\operatorname{cor}(f(\delta_1), \delta_1)$ may be higher, when the range is small. The nonlinearity tests in Pradhan and Hoffman (1963, Table 1) show that linearity is never rejected for shorter ranges (condition VI) while it is often at longer ranges (conditions I–V). Using the data digitized from Figs. 1 and 2 in Luce and Mo (1965), we have performed regression analyses at the individual level selecting pairs of stimuli and responses corresponding to subsets of the whole range of stimuli. This confirms the previous conclusion (i.e. that increasing the range leads to a decrease in the correlation coefficient), but the effect is far from being equal for all the individuals.

Second, as concerns the location effect, the formula predicts that β should be higher when $f'(\mathbb{E}\delta_1)$ is higher. This leads us to investigate the behavior of f'. As an example, Pradhan and Hoffman (1963, pp. 536–537) state that for three subjects out of six the lower section is steeper than the upper one; this is also apparent from their Fig. 2. This is confirmed by Figs. 1 and 2 in Luce and Mo (1965), whose estimated derivatives are reported in our Fig. 3, and Fig. 1 in Green and Luce (1974), from which it is apparent that the slope is higher in a range under the standard. This implies that,

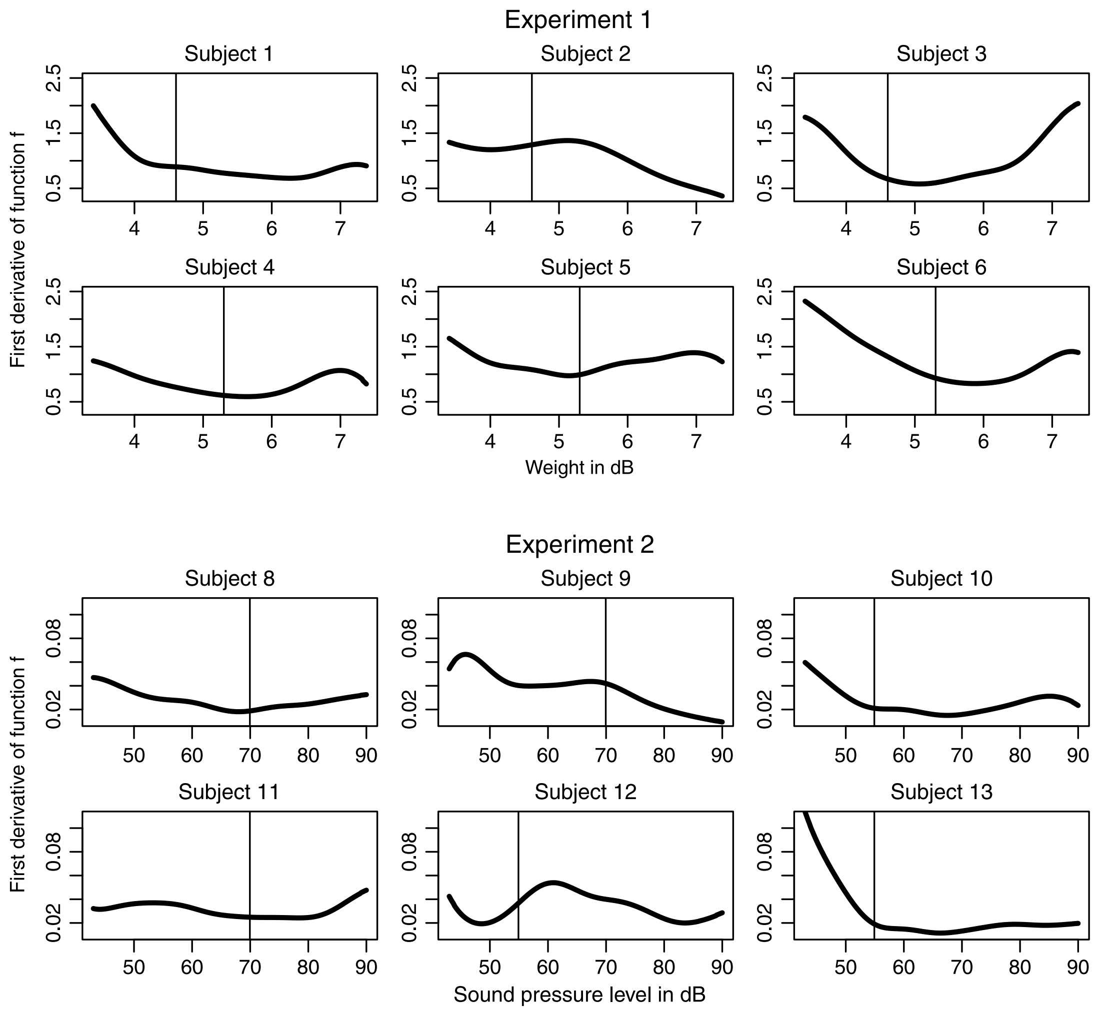

Fig. 2. Function f' estimated through local polynomial regression for Experiments 1 and 2 in Luce and Mo (1965); the vertical line is the standard.

when the expectation $\mathbb{E}\delta_1$ is under the standard δ_2, we may expect $f'(\mathbb{E}\delta_1)$ and β to be higher.

Third, the formula accounts for what Bonnet (1969b, pp. 251–252) calls "[l]a variabilité de la valeur de l'exposant pour les échelles moyennes en fonction des modifications de la situation expérimentale". This happens because there seems to be an inflection point near the standard and the curve appears to be less regular in that neighborhood: Pradhan and Hoffman (1963, p. 537) state something similar for the case of several individuals. When the range is not too large, changing the range of δ_1 or the standard δ_2 may lead to haphazard results. It is interesting to remark, indeed, that the figures in Luce and Mo (1965) show a great variability in linearity around δ_2: in Fig. 1, the curve $\delta_1 \mapsto w^{-1}[\Psi(\delta_1) - \Psi(\delta_2)]$ is reasonably linear for most subjects in a neighborhood of the standard δ_2, while the opposite is true in Fig. 2 of the same source (see also Figs. 1, 2 and 3 in Bernyer, 1962, Fig. 1 in Green & Luce, 1974).

As to Eq. (7), it is probably better to consider the squares of its terms. Here, $\mathrm{cor}^2(\pi, f(\delta_1)) = 1 - \frac{\mathbb{V}(\varepsilon)}{\mathbb{V}(\pi)}$ is the R^2 associated with a separable representation, namely the percentage of the variance of π explained by the variation in $f(\delta_1)$ (see Eq. (5)). Instead, $\mathrm{cor}^2(\pi, \delta_1) = \beta^2 \cdot \frac{\mathbb{V}(\delta_1)}{\mathbb{V}(\pi)}$ is the R^2 associated with a Stevens' power law, namely the part of the variance of π explained by the linear variation in δ_1. The relation:

$$\mathrm{cor}^2(\pi, \delta_1) = \mathrm{cor}^2(f(\delta_1), \delta_1) \cdot \mathrm{cor}^2(\pi, f(\delta_1))$$

means that the fit of Stevens' power law is better (i.e. $\mathrm{cor}^2(\pi, \delta_1)$ is higher) whenever the fit of the separable representation is better (i.e. $\mathrm{cor}^2(\pi, f(\delta_1))$ is higher) and whenever the function f is more easily approximated by a straight line (i.e. $\mathrm{cor}^2(f(\delta_1), \delta_1)$ is higher). In line with the statistical literature quoted in Weiss (1981, pp. 432–433), asserting that power laws are quite good at fitting monotonic (Weiss has "monotone") relations (see, e.g., Good, 1972, 1987), $\mathrm{cor}^2(f(\delta_1), \delta_1)$ is expected to be quite near to one but, as also the other terms are quite near to one, it can make the difference. An order of magnitude of these correlations can be obtained from the analysis of Figs. 1 and 2 in Luce and Mo (1965); remark that in this case each value of π is obtained as the logarithm of the arithmetic mean of 100 measurements,[14] so that this correlation more accurately measures $\mathrm{cor}(f(\delta_1), \delta_1)$ (and not $\mathrm{cor}(\pi, \delta_1)$). The correlation coefficient varies between 0.95 and 1.00 with 10 values out of 12 between 0.985 and 0.995.

5. Ratio magnitude estimation with a standard for several individuals

We suppose that the same stimuli are proposed to J different individuals.[15] We will introduce an index j, as in f_j, w_j, Ψ_j, f_j, π_j, to indicate that the corresponding quantity is relative to the jth individual. Then the following proposition holds.

[14] The most natural choice would have been to take the logarithm of the geometric mean or the arithmetic mean of the logarithm of the values, but this should not alter the results.

[15] A related but different problem, namely aggregation of individual judgments in the context of pairwise comparison matrices, has been considered in Bernasconi, Choirat, and Seri (2014).

Please cite this article in press as: Bernasconi, M., & Seri, R.What are we estimating when we fit Stevens' power law?. Journal of Mathematical Psychology (2016), http://dx.doi.org/10.1016/j.jmp.2016.04.007

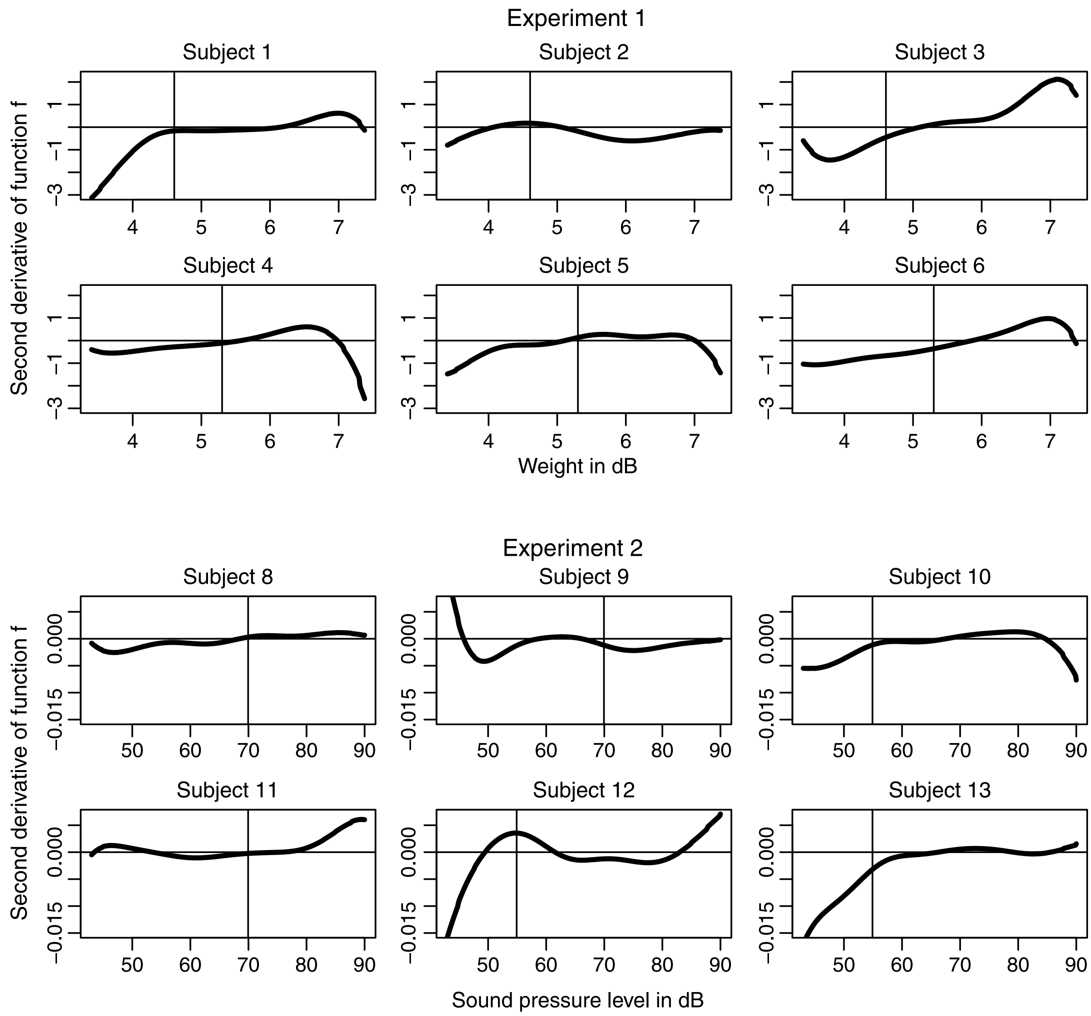

Fig. 3. Function f'' estimated through local polynomial regression for Experiments 1 and 2 in Luce and Mo (1965); the vertical line is the standard.

Proposition 2. *In the case of several individuals and a unique standard, β is given by (6) where $f(x) := \frac{1}{J}\sum_{j=1}^{J} f_j(x)$ and $f_j(x) := \mathbb{E}\left(\pi_j \mid \delta_1\right)$.*

If we define $\overline{\pi} := \frac{1}{J}\sum_{j=1}^{J}\pi_j$, the following formulas hold:

$$\operatorname{cor}\left(f\left(\delta_1\right), \delta_1\right) = \frac{1}{J}\sum_{j=1}^{J}\operatorname{cor}\left(f_j\left(\delta_1\right), \delta_1\right) \cdot \left(\frac{J^2\mathbb{V}\left(f_j(\delta_1)\right)}{\sum_{i=1}^{J}\mathbb{V}(f_i(\delta_1)) + 2\sum_{1\leq i<\ell\leq J}\operatorname{Cov}(f_i(\delta_1), f_\ell(\delta_1))}\right)^{\frac{1}{2}},$$

$$\operatorname{cor}\left(\overline{\pi}, \delta_1\right) = \frac{1}{J}\sum_{j=1}^{J}\operatorname{cor}\left(\pi_j, \delta_1\right) \cdot \left(\frac{J^2\mathbb{V}\left(f_j(\delta_1)\right) + J^2\mathbb{V}(\varepsilon_j)}{\sum_{i=1}^{J}\mathbb{V}(f_i(\delta_1)) + \sum_{i=1}^{J}\mathbb{V}(\varepsilon_i) + 2\sum_{1\leq i<\ell\leq J}\operatorname{Cov}(f_i(\delta_1), f_\ell(\delta_1))}\right)^{\frac{1}{2}},$$

and:

$$\operatorname{cor}\left(\overline{\pi}, \delta_1\right) = \operatorname{cor}\left(f\left(\delta_1\right), \delta_1\right) \cdot \left(1 + \frac{\sum_{j=1}^{J}\mathbb{V}(\varepsilon_j)}{J^2\mathbb{V}(f(\delta_1))}\right)^{-\frac{1}{2}}.$$

Suppose to estimate the model:

$\pi = \alpha + \beta\delta_1 + \beta^{(2)}\delta_1^2 + u.$

Then $\beta = \frac{1}{J}\sum_{j=1}^{J}\beta_j$ and $\beta^{(2)} = \frac{1}{J}\sum_{j=1}^{J}\beta_j^{(2)}$, where β_j and $\beta_j^{(2)}$ are the corresponding coefficients on the data obtained for individual j.

Remark 1. Consider the simplified situation in which $\mathbb{V}\left(\varepsilon_j\right) \equiv \sigma_\varepsilon^2$, $\mathbb{V}\left(f_j\left(\delta_1\right)\right) \equiv \sigma_f^2$ for any j and $\operatorname{Cov}\left(f_i\left(\delta_1\right), f_\ell\left(\delta_1\right)\right) \equiv \rho_f\sigma_f^2$ for any i and ℓ. In this case:

$$\operatorname{cor}\left(f\left(\delta_1\right), \delta_1\right) = \left(\frac{J}{1+(J-1)\rho_f}\right)^{\frac{1}{2}} \cdot \frac{1}{J}\sum_{j=1}^{J}\operatorname{cor}\left(f_j\left(\delta_1\right), \delta_1\right)$$
$$\geq \frac{1}{J}\sum_{j=1}^{J}\operatorname{cor}\left(f_j\left(\delta_1\right), \delta_1\right),$$

$$\operatorname{cor}\left(\overline{\pi}, \delta_1\right) = \left(\frac{J}{1+(J-1)\rho_f} \cdot \frac{\sigma_f^2+\sigma_\varepsilon^2}{\sigma_f^2+\frac{\sigma_\varepsilon^2}{1+(J-1)\rho_f}}\right)^{\frac{1}{2}} \cdot \frac{1}{J}\sum_{j=1}^{J}\operatorname{cor}\left(\pi_j, \delta_1\right)$$
$$\geq \frac{1}{J}\sum_{j=1}^{J}\operatorname{cor}\left(\pi_j, \delta_1\right),$$

as both $\frac{J}{1+(J-1)\rho_f} \geq 1$ and $\frac{\sigma_f^2+\sigma_\varepsilon^2}{\sigma_f^2+\frac{\sigma_\varepsilon^2}{1+(J-1)\rho_f}} \geq 1$. As concerns $\operatorname{cor}\left(f\left(\delta_1\right), \delta_1\right)$, note that ρ_f, being the correlation between two

Please cite this article in press as: Bernasconi, M., & Seri, R.What are we estimating when we fit Stevens' power law?. Journal of Mathematical Psychology (2016), http://dx.doi.org/10.1016/j.jmp.2016.04.007

similar functions, will generally be quite near to 1. Moreover, note that $\mathrm{cor}(\overline{\pi}, \delta_1) \le \mathrm{cor}(f(\delta_1), \delta_1)$ but, if J is large enough, $\mathrm{cor}(\overline{\pi}, \delta_1) \simeq \mathrm{cor}(f(\delta_1), \delta_1)$. The result just stated shows that in general one can use $\mathrm{cor}(\overline{\pi}, \delta_1)$ to approximate $\mathrm{cor}(f(\delta_1), \delta_1)$ and that, if the variances are not too different, $\mathrm{cor}(f(\delta_1), \delta_1)$ is higher than the average of the correlations $\frac{1}{J}\sum_{j=1}^{J} \mathrm{cor}\left(f_j(\delta_1), \delta_1\right)$.

The previous proposition shows that the same result in Proposition 1 still holds true in the present case when the function f is intended as an average over several individuals. As the new function is an average of individual functions, it is interesting to stress the effects of averaging.

Two different phenomena are at work here. The first one, highlighted in Remark 1, is the fact that, if the variances of the answers are not too different between individuals (i.e. in a neighborhood of $\mathbb{V}\left(f_j(\delta_1)\right) \equiv \sigma_f^2$), the correlation of the mean function f with δ_1 is higher than the mean correlation of the individual functions f_j with δ_1. Therefore grouping increases linearity: this holds even if all individuals are sufficiently similar. The second phenomenon is the fact that, if one considers $\beta^{(2)}$ as a measure of nonlinearity, the fact that this coefficient is obtained as an average of the individual coefficients $\beta_j^{(2)}$ implies that coefficients with different signs tend to average out. This requires a certain degree of heterogeneity across individuals. As a result, even if the individuals functions $f_j(x) := w_j^{-1}\left[\Psi_j(x) - \Psi_j(\delta_2)\right]$ are sufficiently nonlinear, the aggregate f is often linear. Pradhan and Hoffman (1963, pp. 537–538) lucidly state that the curve based on data aggregated over individuals "is straighter than every significantly nonlinear individual [...] curve and is not significantly different from linearity. This tendency toward linearity [...] suggests that the power function is an adequate psychophysical function on the data averaged over a group of individuals and, hence, is an artifact of grouping". A similar fact is remarked in Macmillan et al. (1974, p. 344) when they say that at the individual level "[t]he correlation between log stimulus and log response [...] averaged 0.972; all correlations based on grouped data exceeded 0.99".

Now, we turn to evaluate the accuracy of the main formula of Proposition 2 through the data in Green and Luce (1974). In order to do so, we digitized the data from their Fig. 1. The 20 values of the stimulus δ_1 submitted to each of the six individuals are given by the equispaced sequence from 32.5 to 80 dB with separation between adjacent stimuli equal to 2.5 dB. The standard δ_2 is equal to 50 dB. This implies that we have 120 observations, i.e. 20 responses to 20 stimuli for 6 individuals.[16] Fig. 4 represents the 120 observations on the same δ_1 scale, with different symbols representing different individuals. Values of π can be read on the vertical axis, that appears different from the one in Green and Luce (1974) because the latter represents the original value p in logarithmic coordinates. The 20 black solid points are the averages over the responses that the 6 subjects gave to each stimulus. The solid black line is the function f estimated through local polynomial regression. The correlations between π and δ_1 for each of the subjects (in our notation $\mathrm{cor}\left(\pi_j, \delta_1\right)$, for $j \in \{1, 2, \dots, 6\}$) are 0.993, 0.997, 0.993, 0.981, 0.981, 0.989; the correlation between all the responses and the stimuli is 0.844, the average correlation (in our notation $\frac{1}{J}\sum_{j=1}^{J} \mathrm{cor}\left(\pi_j, \delta_1\right)$) is 0.989. The correlation between the average responses $\overline{\pi}$ and the stimuli (in our notation $\mathrm{cor}(\overline{\pi}, \delta_1)$) is 0.998, that is higher than the previous one and than any individual correlation. This shows that averaging increases linearity.

[16] Strangely enough, Fig. 1 in Green and Luce (1974) contains 124 observations. We have removed the observations that do not fit with the description of the text, i.e. that seem to correspond with a value of δ_1 equal to 82.5 dB. Even the standard varies between the caption of Fig. 1 (in which it is 50 dB) and the text (in which it is 55 dB). However, this will not affect our analysis.

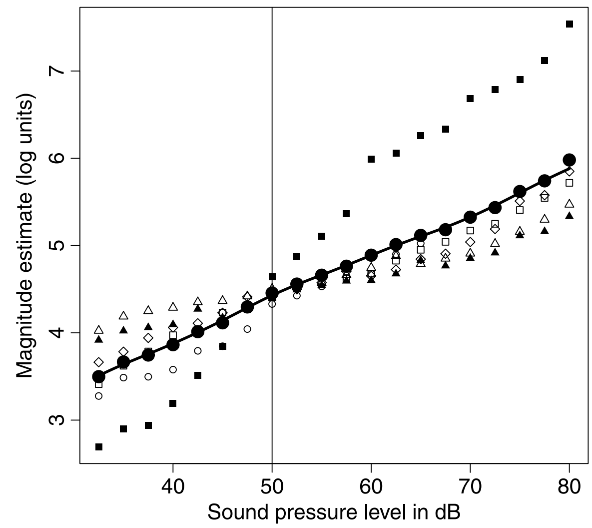

Fig. 4. Individual responses (smaller points; each symbol corresponds to a different individual), averages of the responses across individuals (larger black points), standard δ_2 (vertical black line) and function f estimated through local polynomial regression (black solid curve) for Experiment 1 in Green and Luce (1974).

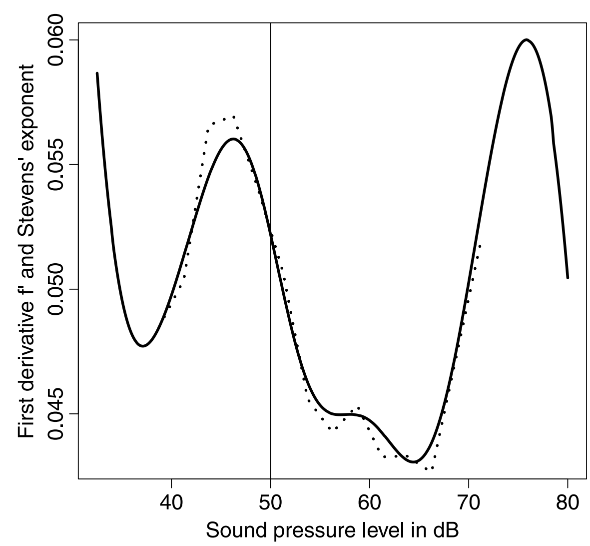

Fig. 5. Function f' estimated through local polynomial regression (black solid curve) and value of the Stevens' exponent in an interval of length 12.5 around the point (black dotted curve) for Experiment 1 in Green and Luce (1974).

Now we come to the illustration of Proposition 2. In Fig. 5, we represent as a solid black line the value of the function f' estimated through local polynomial regression from the observations of Fig. 4. Then for each group of 6 adjacent stimuli in the interval [32.5, 80], we estimate the exponent of Stevens' power law based on the 36 observations (6 responses from 6 individuals) and we plot the value of the exponent against the average value of the 6 stimuli as a black dotted curve. In practice, the latter represents the couples $(\mathbb{E}\delta_1, \beta)$, where each value of β is computed on a restricted range of length 12.5 centered in $\mathbb{E}\delta_1$, while the former represents the curve $\left(\mathbb{E}\delta_1, f'(\mathbb{E}\delta_1)\right)$. The very good agreement between β and $f'(\mathbb{E}\delta_1)$ is due to the fact that, on each subinterval of these data, $\mathrm{cor}(f(\delta_1), \delta_1) \simeq \mathrm{cor}(\overline{\pi}, \delta_1)$ is quite near to 1 (as $\mathrm{cor}(\overline{\pi}, \delta_1)$ varies between 0.990 and 0.999) and $\mathbb{E}(\delta_1 - \mathbb{E}\delta_1)^3 = 0$. This provides a graphical check of our formula. From a statistical point of view, this result is reasonable, because local polynomial regression estimates the derivative of the function in a given point using the slope of a linear regression fitted in a neighborhood of the point.

Please cite this article in press as: Bernasconi, M., & Seri, R.What are we estimating when we fit Stevens' power law?. Journal of Mathematical Psychology (2016), http://dx.doi.org/10.1016/j.jmp.2016.04.007

Despite averaging increases linearity, some salient facts concerning the individual functions f_j can still be found in f. As an example, the fact that most individuals have a decreasing f_j' over most of the range implies that also f' will have this property. This explains why the evidence on the location effect is so overwhelming in grouped data, namely because averaging across individuals reduces the impact of those that do not conform to the more common behavior in the sample. This can be seen in Fig. 4 where the only individual showing a different curvature of f_j does not succeed in changing the fact that f' is decreasing as the other f_j' are.

6. Conclusion

It is by now well-known in mathematical psychology that the application of Stevens' methods to measure sensation requires the fulfillment of specific axioms. In particular, the axiomatization of magnitude methods is about what scale of measurement they produce. The axiomatization of Narens (1996) included the multiplicative axiom, largely found to fail empirically. However, Luce's (2002, 2004, 2008) model of global psychophysics (see also Steingrimsson & Luce, 2007) later showed that, by abandoning the requirement $W(1) = 1$, the commutativity axiom is equivalent to obtaining ratio-scale measures. This axiom has been favorably evaluated in many experiments. The conclusion, currently, is that magnitude methods do produce ratio-scale measures (e.g., Steingrimsson, this issue, in particular Section 3.1).

The recent literature has devoted more attention to testing the axioms than to linking precisely the well-known empirical inconsistencies of Stevens' power law to the newer theories of psychological measurement. This article has used separable representations in the spirit of Narens and Luce to examine several contextual effects involving Stevens' power law.

We have focussed on ratio estimation and have developed an approach to analyze how the parameter β of Stevens' power law in Eq. (1) depends on the distribution of the stimuli and on a function f which links δ_1 to π and involves the subjective weighting function W and the psychophysical function ψ of separable representations. We have obtained several results which explain and predict the sensitivity of Stevens' power law to the experimental parameters and conditions. We have illustrated the results referring to various classical papers originating in the sixties and seventies around Stevens' model, including earlier experiments performed by Green and Luce (1974) and Luce and Mo (1965). We have mainly focused on the contextual effects known as 'range effect', attributable to the varying linearity of f over different ranges, 'location effect', depending on the derivative of f at the standard, and 'averaging', or 'grouping effect', explained by the linearization properties of aggregation over individuals.

Despite the references to the earlier literature, the main interest of the paper is not historical. Stevens' power law is still applied in several fields of psychophysics. Separable representations represent a special case of the more global psychophysical theory of intensity perception developed by Luce (2002, 2004, 2008) which is receiving substantial empirical support. So it is quite important to establish precisely how the power exponent of Stevens' law depends on the actual psychophysical functions. Our approach could give even more precise predictions on the behavior of Stevens' exponent with more knowledge about the shape and forms of both ψ and W. This could come pursuing the line of research drawn in several recent experiments by Luce and Steingrimsson (Steingrimsson and Luce, 2005a, 2005b, 2006, 2007, 2012).

Finally, we remark that while our approach has focussed on ratio magnitude estimation, the method can be extended to analyze also magnitude production. Among other things, this would offer the opportunity to consider also the so-called Stevens' 'regression' effect, which Luce (2013) considered in his last single-author paper, with a related but distinct approach: in it, following a suggestion by Ragnar Steingrimsson, he made an explicit reference to the role of the estimation method in the regression effect.

Acknowledgments

We are particularly grateful to three anonymous referees and the guest editor in charge, Ragnar Steingrimsson, for several very helpful comments which helped us to revise the paper. Raffaello Seri gratefully acknowledges financial support of the COFIN Project 2010J3LZEN_006.

Appendix. Proofs

The values of δ_1 are generally chosen in a deterministic way but, nevertheless, we will take expectations with respect to the distributions of the stimuli (unless they are fixed, as is the case of δ_2 for ratio magnitude estimation with a standard). This can be justified by the following reasoning. Consider, as an example, the case in which δ_1 takes the values $\{\delta_{1i},\ i = 1, \ldots, I\}$. It is possible to define a design probability $\mathbb{P}^{(I)}(A) \triangleq \frac{1}{I} \cdot \sum_{i=1}^{I} 1\left\{\delta_{1,i} \in A\right\}$. Whether I is supposed to diverge or not, $\mathbb{P}^{(I)}$ can be considered "near" to an *asymptotic design measure* $\mathbb{P}$ (see Cox, 1988). Therefore, in the following expectations and other moments should be considered as computed according to the design or the asymptotic design measure.

Proof of Proposition 1. In the following we will repeatedly use the facts that $\operatorname{Cov}(\varepsilon, f(\delta_1)) = \mathbb{E}(\varepsilon \cdot f(\delta_1)) = 0$ and $\operatorname{Cov}(\varepsilon, \delta_1) = \mathbb{E}(\varepsilon \cdot \delta_1) = 0$. Both derive, through the Law of Iterated Expectations, from $\mathbb{E}(\varepsilon|\delta_1) = 0$, and lead, respectively, to $\operatorname{Cov}(\pi, f(\delta_1)) = \mathbb{V}(f(\delta_1))$ and $\operatorname{Cov}(\pi, \delta_1) = \operatorname{Cov}(f(\delta_1), \delta_1)$.

The formula of the OLS slope parameter in a linear regression yields:

$$\beta = \frac{\operatorname{Cov}(\pi, \delta_1)}{\mathbb{V}(\delta_1)} = \frac{\operatorname{Cov}(f(\delta_1), \delta_1)}{\mathbb{V}(\delta_1)}$$
$$= \operatorname{cor}(f(\delta_1), \delta_1) \cdot \left(\frac{\mathbb{V}(f(\delta_1))}{\mathbb{V}(\delta_1)}\right)^{\frac{1}{2}}.$$

Then, we use the delta method for moments to get:

$$\begin{aligned}\mathbb{V}(f(\delta_1)) \simeq{}& f'^2(\mathbb{E}\delta_1) \cdot \mathbb{V}(\delta_1) + f'(\mathbb{E}\delta_1) f''(\mathbb{E}\delta_1) \cdot \mathbb{E}(\delta_1 - \mathbb{E}\delta_1)^3 \\ &+ \frac{f'(\mathbb{E}\delta_1) f'''(\mathbb{E}\delta_1)}{3} \cdot \mathbb{E}(\delta_1 - \mathbb{E}\delta_1)^4 \\ &+ \frac{\left[f''(\mathbb{E}\delta_1)\right]^2}{4} \cdot \left(\mathbb{E}(\delta_1 - \mathbb{E}\delta_1)^4 - \mathbb{V}(\delta_1)^2\right).\end{aligned}$$

Now we turn to the correlation. We have:

$$\begin{aligned}\frac{\operatorname{cor}(\pi, \delta_1)}{\operatorname{cor}(\pi, f(\delta_1))} &= \frac{\operatorname{Cov}(\pi, \delta_1)\sqrt{\mathbb{V}(f(\delta_1))}}{\operatorname{Cov}(\pi, f(\delta_1))\sqrt{\mathbb{V}(\delta_1)}} \\ &= \frac{\operatorname{Cov}(f(\delta_1), \delta_1)}{\sqrt{\mathbb{V}(f(\delta_1))\,\mathbb{V}(\delta_1)}} = \operatorname{cor}(f(\delta_1), \delta_1).\end{aligned}$$

Proof of Proposition 2. The two models considered in the statement, namely $\pi = \alpha + \beta\delta_1 + u$ and $\pi = \alpha + \beta\delta_1 + \beta^{(2)}\delta_1^2 + u$, must also hold averaging over individuals. Therefore, they respectively become $\overline{\pi} = \alpha + \beta\delta_1 + \overline{u}$ and $\overline{\pi} = \alpha + \beta\delta_1 + \beta^{(2)}\delta_1^2 + \overline{u}$. We will also use repeatedly the facts that $\operatorname{Cov}(\overline{\varepsilon}, \delta_1) = \frac{1}{J}\sum_{j=1}^{J} \operatorname{Cov}\left(\varepsilon_j, \delta_1\right) = 0$, because $\operatorname{Cov}\left(\varepsilon_j, \delta_1\right) = \mathbb{E}\left(\varepsilon_j|\delta_1\right) = 0$, and that $\operatorname{Cov}(\overline{\pi}, \delta_1) = \operatorname{Cov}(f(\delta_1), \delta_1)$.

The formula of β in the first model is $\beta = \frac{\operatorname{Cov}(\overline{\pi}, \delta_1)}{\mathbb{V}(\delta_1)} = \frac{\operatorname{Cov}(f(\delta_1), \delta_1)}{\mathbb{V}(\delta_1)}$. Therefore, the proof of Proposition 1 still holds with the new definition of f.

As concerns the correlation $\mathrm{cor}\,(f\,(\delta_1)\,,\delta_1)$, we have:

$$\begin{aligned}\mathrm{cor}\,(f\,(\delta_1)\,,\delta_1) &= \frac{\mathrm{Cov}(f\,(\delta_1),\delta_1)}{\sqrt{\mathbb{V}(f\,(\delta_1))\mathbb{V}(\delta_1)}}\\ &= \frac{1}{J}\sum_{j=1}^{J}\frac{\mathrm{Cov}\left(f_j(\delta_1),\delta_1\right)}{\sqrt{\mathbb{V}\left(f_j(\delta_1)\right)\mathbb{V}(\delta_1)}}\cdot\left(\frac{\mathbb{V}\left(f_j(\delta_1)\right)}{\mathbb{V}(f\,(\delta_1))}\right)^{\frac{1}{2}}\\ &= \frac{1}{J}\sum_{j=1}^{J}\mathrm{cor}\left(f_j\,(\delta_1)\,,\delta_1\right)\\ &\quad\cdot\left(\frac{J^2\mathbb{V}\left(f_j(\delta_1)\right)}{\sum_{i=1}^{J}\mathbb{V}(f_i(\delta_1))+2\sum_{1\le i<\ell\le J}\mathrm{Cov}(f_i(\delta_1),f_\ell(\delta_1))}\right)^{\frac{1}{2}}.\end{aligned}$$

A similar equality holds for $\mathrm{cor}\,(\overline{\pi},\delta_1)$:

$$\begin{aligned}\mathrm{cor}\,(\overline{\pi},\delta_1) &= \frac{\mathrm{Cov}(\overline{\pi},\delta_1)}{\sqrt{\mathbb{V}(\overline{\pi})\mathbb{V}(\delta_1)}} = \frac{\mathrm{Cov}(f\,(\delta_1),\delta_1)}{\sqrt{\mathbb{V}\left(f\,(\delta_1)+\frac{1}{J}\sum_{j=1}^{J}\varepsilon_j\right)\mathbb{V}(\delta_1)}}\\ &= \frac{1}{J}\sum_{j=1}^{J}\mathrm{cor}\left(\pi_j,\delta_1\right)\cdot\left(\frac{\mathbb{V}\left(f_j(\delta_1)\right)+\mathbb{V}\left(\varepsilon_j\right)}{\mathbb{V}(f(\delta_1))+\mathbb{V}\left(\frac{1}{J}\sum_{i=1}^{J}\varepsilon_i\right)}\right)^{\frac{1}{2}}\\ &= \frac{1}{J}\sum_{j=1}^{J}\mathrm{cor}\left(\pi_j,\delta_1\right)\\ &\quad\cdot\left(\frac{J^2\mathbb{V}\left(f_j(\delta_1)\right)+J^2\mathbb{V}\left(\varepsilon_j\right)}{\sum_{i=1}^{J}\mathbb{V}(f_i(\delta_1))+\sum_{i=1}^{J}\mathbb{V}(\varepsilon_i)+2\sum_{1\le i<\ell\le J}\mathrm{Cov}(f_i(\delta_1),f_\ell(\delta_1))}\right)^{\frac{1}{2}}.\end{aligned}$$

Remark that:

$$\mathrm{cor}\,(\overline{\pi},\delta_1) = \frac{\mathrm{Cov}(\overline{\pi},\delta_1)}{\sqrt{\mathbb{V}(\overline{\pi})\mathbb{V}(\delta_1)}} = \mathrm{cor}\,(f\,(\delta_1)\,,\delta_1)\cdot\left(1+\frac{\sum_{j=1}^{J}\mathbb{V}\left(\varepsilon_j\right)}{J^2\mathbb{V}(f\,(\delta_1))}\right)^{-\frac{1}{2}}.$$

As concerns the coefficient $\beta^{(2)}$, from $\overline{\pi}=\alpha+\beta\delta_1+\beta^{(2)}\delta_1^2+\overline{u}$, it is easy to see that $\beta^{(2)}$ is defined by the formula:

$$\begin{aligned}\beta^{(2)} &= \frac{\mathbb{E}\delta_1^2\cdot\mathbb{E}\left(\delta_1^2\overline{\pi}\right)-\mathbb{E}\delta_1^3\cdot\mathbb{E}(\delta_1\overline{\pi})}{\mathbb{E}\delta_1^2\cdot\mathbb{E}\delta_1^4-\left(\mathbb{E}\delta_1^3\right)^2} = \frac{1}{J}\sum_{j=1}^{J}\frac{\mathbb{E}\delta_1^2\cdot\mathbb{E}\left(\delta_1^2\pi_j\right)-\mathbb{E}\delta_1^3\cdot\mathbb{E}\left(\delta_1\pi_j\right)}{\mathbb{E}\delta_1^2\cdot\mathbb{E}\delta_1^4-\left(\mathbb{E}\delta_1^3\right)^2}\\ &= \frac{1}{J}\sum_{j=1}^{J}\beta_j^{(2)}.\end{aligned}$$

References

Ahlström, R., & Baird, J. C. (1989). Shift in stimulus range and the exponent of the power function for loudness. *Perception & Psychophysics*, *46*, 603–607. http://dx.doi.org/10.3758/BF03208158.

Attneave, F. (1962). Perception and related areas. In *Biologically oriented fields: their place in psychology and in biological science*: *Vol. 4*. *Psychology: A study of a science* (pp. 619–659). New York: McGraw-Hill.

Augustin, T. (2006). Stevens' direct scaling methods and the uniqueness problem. *Psychometrika*, *71*, 469–481. http://dx.doi.org/10.1007/s11336-004-1200-y.

Augustin, T. (2010). Extensions of Narens' theory of ratio magnitude estimation. *The British Journal of Mathematical and Statistical Psychology*, *63*, 647–663. http://dx.doi.org/10.1348/000711010X486075.

Augustin, T., & Maier, K. (2008). Empirical evaluation of the axioms of multiplicativity, commutativity, and monotonicity in ratio production of area. *Acta Psychologica*, *129*, 208–216. http://dx.doi.org/10.1016/j.actpsy.2008.06.004.

Bernasconi, M., Choirat, C., & Seri, R. (2008). Measurement by subjective estimation: Testing for separable representations. *Journal of Mathematical Psychology*, *52*, 184–201. http://dx.doi.org/10.1016/j.jmp.2008.01.002.

Bernasconi, M., Choirat, C., & Seri, R. (2010). The Analytic Hierarchy Process and the theory of measurement. *Management Science*, *56*, 699–711. http://dx.doi.org/10.1287/mnsc.1090.1123.

Bernasconi, M., Choirat, C., & Seri, R. (2011). A re-examination of the algebraic properties of the AHP as a ratio-scaling technique. *Journal of Mathematical Psychology*, *55*, 152–165. http://dx.doi.org/10.1016/j.jmp.2010.11.002.

Bernasconi, M., Choirat, C., & Seri, R. (2014). Empirical properties of group preference aggregation methods employed in AHP: Theory and evidence. *European Journal of Operational Research*, *232*, 584–592. http://dx.doi.org/10.1016/j.ejor.2013.06.014.

Bernyer, G. (1962). Etude sur la validité d'une échelle de sensation d'effort musculaire. *L'année psychologique*, *62*, 1–15. http://dx.doi.org/10.3406/psy.1962.7153.

Birnbaum, M. H. (1980). Comparison of two theories of "ratio" and "difference" judgments. *Journal of Experimental Psychology: General*, *109*, 304–319. http://dx.doi.org/10.1037/0096-3445.109.3.304.

Birnbaum, M. H. (1982). Controversies in psychological measurement. In B. Wegener (Ed.), *Social attitudes and psychophysical measurement* (pp. 401–485). Hillsdale, NJ: Erlbaum.

Birnbaum, M. H., Anderson, C. J., & Hynan, L. G. (1989). Two operations for "ratios" and "differences" of distances on the mental map. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 785–796. http://dx.doi.org/10.1037/0096-1523.15.4.785.

Birnbaum, M. H., & Elmasian, R. (1977). Loudness "ratios" and "differences" involve the same psychophysical operation. *Perception & Psychophysics*, *22*, 383–391. http://dx.doi.org/10.3758/BF03199704.

Birnbaum, M. H., & Mellers, B. A. (1978). Measurement and the mental map. *Perception & Psychophysics*, *23*, 403–408. http://dx.doi.org/10.3758/BF03204143.

Birnbaum, M. H., & Veit, C. T. (1974). Scale convergence as a criterion for rescaling: Information integration with difference, ratio, and averaging tasks. *Perception & Psychophysics*, *15*, 7–15. http://dx.doi.org/10.3758/BF03205820.

Bonnet, C. (1969a). La construction d'échelles psychophysiques de rapport. *Bulletin de Psychologie*, *22*, 624–630.

Bonnet, C. (1969b). Les échelles subjectives directes en psychophysique. *L'année psychologique*, *69*, 247–264. http://dx.doi.org/10.3406/psy.1969.27659.

Cox, D. D. (1988). Approximation of least squares regression on nested subspaces. *The Annals of Statistics*, *16*, 713–732. http://dx.doi.org/10.1214/aos/1176350830.

Curtis, D. W., Attneave, F., & Harrington, T. L. (1968). A test of a two-stage model of magnitude judgment. *Perception & Psychophysics*, *3*, 25–31. http://dx.doi.org/10.3758/BF03212707.

Ekman, G. (1958). Two generalized ratio scaling methods. *Journal of Psychology*, *45*, 287–295. http://dx.doi.org/10.1080/00223980.1958.9916259.

Ekman, G., Hosman, B., Lindman, R., Ljungberg, L., & Akesson, C. A. (1968). Interindividual differences in scaling performance. *Perceptual and Motor Skills*, *26*, 815–823. http://dx.doi.org/10.2466/pms.1968.26.3.815.

Ellermeier, W., & Faulhammer, G. (2000). Empirical evaluation of axioms fundamental to Stevens's ratio-scaling approach: I. Loudness production. *Perception & Psychophysics*, *62*, 1505–1511. http://dx.doi.org/10.3758/BF03212151.

Engen, T. (1956). An evaluation of a method for developing ratio-scales. *The American Journal of Psychology*, *69*, 92–95. http://dx.doi.org/10.2307/1418120.

Engen, T., & Levy, N. (1955). The influence of standards on psychophysical judgments. *Perceptual and Motor Skills*, *5*, 193–197. http://dx.doi.org/10.2466/pms.1955.5.3.193.

Engen, T., & Levy, N. (1958). The influence of context on constant-sum loudness-judgments. *The American Journal of Psychology*, *71*, 731–736. http://dx.doi.org/10.2307/1420331.

Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin*, *53*, 134–140. http://dx.doi.org/10.1037/h0045156.

Estes, W. K., & Maddox, W. T. (2005). Risks of drawing inferences about cognitive processes from model fits to individual versus average performance. *Psychonomic Bulletin & Review*, *12*, 403–408. http://dx.doi.org/10.3758/BF03193784.

Fagot, R. F., & Pokorny, R. (1989). Bias effects on magnitude and ratio estimation power function exponents. *Perception & Psychophysics*, *45*, 221–230. http://dx.doi.org/10.3758/BF03210701.

Florentine, M., Popper, A. N., & Fay, R. R. (Eds.) (2011). *Springer handbook of auditory research*: *Vol. 37*. *Loudness*. New York, NY: Springer.

Good, I. J. (1972). 342. Note: Correlation for power functions. *Biometrics*, *28*, 1127–1129. http://dx.doi.org/10.2307/2528645.

Good, I. J. (1987). C286. Correlation and linearity: some lapses in exposition. *Journal of Statistical Computation and Simulation*, *28*, 93–94. http://dx.doi.org/10.1080/00949658708811017.

Green, D. M., & Luce, R. D. (1974). Variability of magnitude estimates: A timing theory analysis. *Perception & Psychophysics*, *15*, 291–300. http://dx.doi.org/10.3758/BF03213947.

Green, D. M., Luce, R. D., & Smith, A. F. (1980). Individual magnitude estimates for various distributions of signal intensity. *Perception & Psychophysics*, *27*, 483–488. http://dx.doi.org/10.3758/BF03198675.

Gulliksen, H. (1959). Mathematical solutions for psychological problems. *American Scientist*, *47*, 178–201.

Hardin, C., & Birnbaum, M. H. (1990). Malleability of "ratio" judgments of occupational prestige. *The American Journal of Psychology*, *103*, 1–20. http://dx.doi.org/10.2307/1423256.

Jones, F. N., & Marcus, M. J. (1961). The subject effect in judgments of subjective magnitude. *Journal of Experimental Psychology*, *61*, 40–44. http://dx.doi.org/10.1037/h0046074.

Kornbrot, D. E. (2014). Human psychophysical functions, an update: methods for identifying their form; estimating their parameters; and evaluating the effects of important predictors. *Psychometrika*, 1–16. http://dx.doi.org/10.1007/s11336-014-9418-9.

Kowal, K. H. (1993). The range effect as a function of stimulus set, presence of a standard, and modulus. *Perception & Psychophysics*, *54*, 555–561. http://dx.doi.org/10.3758/BF03211777.

Krantz, D. H. (1972). A theory of magnitude estimation and cross-modality matching. *Journal of Mathematical Psychology*, *9*, 168–199. http://dx.doi.org/10.1016/0022-2496(72)90025-9.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement, Vol. I: Additive and polynomial representations*. New York, NY: Academic Press.

Künnapas, T. M. (1958a). Measurements of subjective length in the vertical-horizontal illusion. *Nordisk Psykologi*, *10*, 203–206. http://dx.doi.org/10.1080/00291463.1958.10780388.

Künnapas, T. M. (1958b). Measurements of subjective length in the vertical-horizontal illusion. *Acta Psychologica*, *14*, 371–374. http://dx.doi.org/10.1016/0001-6918(58)90029-5.

Künnapas, T. M. (1960). Scales for subjective distance. *Scandinavian Journal of Psychology*, *1*, 187–192. http://dx.doi.org/10.1111/j.1467-9450.1960.tb01299.x.

Künnapas, T. M. (1961). Scales for subjective distance. *Acta Psychologica*, *19*, 214–216. http://dx.doi.org/10.1016/S0001-6918(61)80080-2.

Laming, D. (1997). *The measurement of sensation*. Oxford, United Kingdom: Oxford University Press.

Lawless, H. T., & Heymann, H. (2010). *Sensory evaluation of food—principles and practices*. New York, NY: Springer.

Lim, J. (2011). Hedonic scaling: A review of methods and theory. *Food Quality and Preference*, *22*, 733–747. http://dx.doi.org/10.1016/j.foodqual.2011.05.008.

Loomes, G. (2005). Modelling the stochastic component of behaviour in experiments: Some issues for the interpretation of data. *Experimental Economics*, *8*, 301–323. http://dx.doi.org/10.1007/s10683-005-5372-9.

Luce, R. D. (1959). On the possible psychophysical laws. *Psychological Review*, *66*, 81–95. http://dx.doi.org/10.1037/h0043178.

Luce, R. D. (1990). ''On the possible psychophysical laws'' revisited: Remarks on cross-modal matching. *Psychological Review*, *97*, 66–77. http://dx.doi.org/10.1037/0033-295X.97.1.66.

Luce, R. D. (1995). Four tensions concerning mathematical modeling in psychology. *Annual Review of Psychology*, *46*, 1–27. http://dx.doi.org/10.1146/annurev.ps.46.020195.000245.

Luce, R. D. (1996). The ongoing dialog between empirical science and measurement theory. *Journal of Mathematical Psychology*, *40*, 78–98. http://dx.doi.org/10.1006/jmps.1996.0005.

Luce, R. D. (1997). Several unresolved conceptual problems of mathematical psychology. *Journal of Mathematical Psychology*, *41*, 79–87. http://dx.doi.org/10.1006/jmps.1997.1150.

Luce, R. D. (2002). A psychophysical theory of intensity proportions, joint presentations, and matches. *Psychological Review*, *109*, 520–532. http://dx.doi.org/10.1037/0033-295X.109.3.520.

Luce, R. D. (2004). Symmetric and asymmetric matching of joint presentations. *Psychological Review*, *111*, 446–454. http://dx.doi.org/10.1037/0033-295X.111.2.446.

Luce, R. D. (2005). Measurement analogies: comparisons of behavioral and physical measures. *Psychometrika*, *70*, 227–251. http://dx.doi.org/10.1007/s11336-004-1248-8.

Luce, R. D. (2008). ''Symmetric and asymmetric matching of joint presentations'': Correction to Luce (2004). *Psychological Review*, *115*, 601. http://dx.doi.org/10.1037/0033-295X.115.3.601.

Luce, R. D. (2012a). Predictions about bisymmetry and cross-modal matches from global theories of subjective intensities. *Psychological Review*, *119*, 373–387. http://dx.doi.org/10.1037/a0027122.

Luce, R. D. (2012b). Torgerson's conjecture and Luce's magnitude production representation imply an empirically false property. *Journal of Mathematical Psychology*, *56*, 176–178. http://dx.doi.org/10.1016/j.jmp.2012.02.002.

Luce, R. D. (2013). Analogs in Luce's global psychophysical theory of Stevens's psychophysical regression effect. *The American Journal of Psychology*, *126*, 47–52. http://dx.doi.org/10.5406/amerjpsyc.126.1.0045.

Luce, R. D., Krantz, D. H., Suppes, P., & Tversky, A. (1990). *Foundations of measurement, vol. III: Representation, axiomatization, and invariance*. New York, NY: Academic Press.

Luce, R. D., & Krumhansl, C. L. (1988). Measurement, scaling, and psychophysics. In *Stevens' handbook of experimental psychology* (pp. 3–73). New York, NY: Wiley.

Luce, R. D., & Mo, S. S. (1965). Magnitude estimation of heaviness and loudness by individual subjects: A test of a probabilistic response theory. *The British Journal of Mathematical and Statistical Psychology*, *18*, 159–174. http://dx.doi.org/10.1111/j.2044-8317.1965.tb00339.x.

Luce, R. D., Steingrimsson, R., & Narens, L. (2010). Are psychophysical scales of intensities the same or different when stimuli vary on other dimensions? Theory with experiments varying loudness and pitch. *Psychological Review*, *117*, 1247–1258. http://dx.doi.org/10.1037/a0020174.

MacKay, D. M. (1963). Psychophysics of perceived intensity: A theoretical basis for Fechner's and Stevens' laws. *Science*, *139*, 1213–1216. http://dx.doi.org/10.1126/science.139.3560.1213-a.

Macmillan, N. A., Moschetto, C. F., Bialostozky, F. M., & Engel, L. (1974). Size judgment: The presence of a standard increases the exponent of the power law. *Perception & Psychophysics*, *16*, 340–346. http://dx.doi.org/10.3758/BF03203953.

Marks, L. E. (1988). Magnitude estimation and sensory matching. *Perception & Psychophysics*, *43*, 511–525. http://dx.doi.org/10.3758/BF03207739.

Marley, A. A. J. (1972). Internal state models for magnitude estimation and related experiments. *Journal of Mathematical Psychology*, *9*, 306–319. http://dx.doi.org/10.1016/0022-2496(72)90021-1.

Masin, S. C. (2014). Test of the ratio judgment hypothesis. *The Journal of General Psychology*, *141*, 130–150. http://dx.doi.org/10.1080/00221309.2014.883355.

Myung, J. I., Karabatsos, G., & Iverson, G. J. (2005). A Bayesian approach to testing decision making axioms. *Journal of Mathematical Psychology*, *49*, 205–225. http://dx.doi.org/10.1016/j.jmp.2005.02.004.

Narens, L. (1996). A theory of ratio magnitude estimation. *Journal of Mathematical Psychology*, *40*, 109–129. http://dx.doi.org/10.1006/jmps.1996.0011.

Narens, L. (2002). The irony of measurement by subjective estimations. *Journal of Mathematical Psychology*, *46*, 769–788. http://dx.doi.org/10.1006/jmps.2002.1429.

Narens, L. (2006). Symmetry, direct measurement, and Torgerson's conjecture. *Journal of Mathematical Psychology*, *50*, 290–301. http://dx.doi.org/10.1016/j.jmp.2005.12.007.

Piéron, H. (1963). La psychophysique. In *Traité de psychologie expérimentale*. Paris, France: Presses Universitaires de France.

Poulton, E. C. (1968). The new psychophysics: Six models for magnitude estimation. *Psychological Bulletin*, *69*, 1–19. http://dx.doi.org/10.1037/h0025267.

Pradhan, P. L., & Hoffman, P. J. (1963). Effect of spacing and range of stimuli on magnitude estimation judgements. *Journal of Experimental Psychology*, *66*, 533–541. http://dx.doi.org/10.1037/h0041492.

Prelec, D. (1998). The probability weighting function. *Econometrica*, *66*, 497–528. http://dx.doi.org/10.2307/2998573.

Ramsay, J. O. (1979). Intra- and interindividual variation in the power law exponent for area summation. *Perception & Psychophysics*, *26*, 495–500. http://dx.doi.org/10.3758/BF03204290.

Regenwetter, M., & Marley, A. A. J. (2001). Random relations, random utilities, and random functions. *Journal of Mathematical Psychology*, *45*, 864–912. http://dx.doi.org/10.1006/jmps.2000.1357.

Rule, S. J., & Curtis, D. W. (1973). Reevaluation of two models for judgments of perceptual intervals. *Perception & Psychophysics*, *14*, 433–436. http://dx.doi.org/10.3758/BF03211178.

Rule, S. J., & Curtis, D. W. (1977). Subject differences in input and output transformations from magnitude estimation of differences. *Acta Psychologica*, *41*, 61–65. http://dx.doi.org/10.1016/0001-6918(77)90011-7.

Shepard, R. N. (1981). Psychological relations and psychophysical scales: On the status of ''direct'' psychophysical measurement. *Journal of Mathematical Psychology*, *24*, 21–57. http://dx.doi.org/10.1016/0022-2496(81)90034-1.

Shofner, W. P., & Selas, G. (2002). Pitch strength and Stevens's power law. *Perception & Psychophysics*, *64*, 437–450. http://dx.doi.org/10.3758/BF03194716.

Sidman, M. (1952). A note on functional relations obtained from group data. *Psychological Bulletin*, *49*, 263–269. http://dx.doi.org/10.1037/h0063643.

Skinner, F. B. (1958). Reinforcement today. *American Psychologist*, *13*, 94–99. http://dx.doi.org/10.1037/h0049039.

Steingrimsson, R. (2009). Evaluating a model of global psychophysical judgments for brightness: I. Behavioral properties of summations and productions. *Attention, Perception, and Psychophysics*, *71*, 1916–1930. http://dx.doi.org/10.3758/APP.71.8.1916.

Steingrimsson, R. (2016). Subjective intensity: Behavioral laws, numerical representations, and behavioral predictions in Luce's model of global psychophysics. *Journal of Mathematical Psychology*. http://dx.doi.org/10.1016/j.jmp.2016.06.004 (this issue).

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. http://dx.doi.org/10.1016/j.jmp.2005.03.003.

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. http://dx.doi.org/10.1016/j.jmp.2005.03.001.

Steingrimsson, R., & Luce, R. D. (2006). Empirical evaluation of a model of global psychophysical judgments: III. A form for the psychophysical function and intensity filtering. *Journal of Mathematical Psychology*, *50*, 15–29. http://dx.doi.org/10.1016/j.jmp.2005.11.005.

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. http://dx.doi.org/10.1016/j.jmp.2006.08.001.

Steingrimsson, R., & Luce, R. D. (2012). Predictions from a model of global psychophysics about differences between perceptual and physical matches. *Attention, Perception, and Psychophysics*, *74*, 1668–1680. http://dx.doi.org/10.3758/s13414-012-0334-8.

Steingrimsson, R., Luce, R. D., & Narens, L. (2012). Brightness of different hues Is a single psychophysical ratio scale of intensity. *American Journal of Psychology*, *125*, 321–333. http://dx.doi.org/10.5406/amerjpsyc.125.3.0321.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, *103*, 677–680. http://dx.doi.org/10.1126/science.103.2684.677.

Stevens, S. S. (1951). Mathematics, measurement, and psychophysics. In *Handbook of experimental psychology* (pp. 1–49). New York, NY: Wiley.

Stevens, S. S. (1955). On the averaging of data. *Science*, *121*, 113–116. http://dx.doi.org/10.1126/science.121.3135.113.

Stevens, S. S. (1956). The direct estimation of sensory magnitudes–loudness. *The American Journal of Psychology*, *100*, 664–689. http://dx.doi.org/10.2307/1418112.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, *64*, 153–181. http://dx.doi.org/10.1037/h0046162.

Stevens, S. S. (1958). Problems and methods of psychophysics. *Psychological Bulletin*, *55*, 177–196. http://dx.doi.org/10.1037/h0044251.

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

Stevens, J. C., & Stevens, S. S. (1960). Warmth and cold: dynamics of sensory intensity. *Journal of Experimental Psychology*, *60*, 183–192. http://dx.doi.org/10.1037/h0040840.

Suppes, P., Krantz, D. H., Luce, R. D., & Tversky, A. (1989). *Foundations of measurement, Vol. II: Geometrical, threshold, and probabilistic representations*. New York, NY: Academic Press.

Teghtsoonian, R. (1971). On the exponents in Stevens' law and the constant in Ekman's law. *Psychological Review*, *78*, 71–80. http://dx.doi.org/10.1037/h0030300.

Teghtsoonian, R. (1973). Range effects in psychophysical scaling and a revision of Stevens' law. *The American Journal of Psychology*, *86*, 3–27. http://dx.doi.org/10.2307/1421845.

Teghtsoonian, R., Teghtsoonian, M., & Baird, J. C. (1995). On the nature and meaning of sinuosity in magnitude-estimation functions. *Psychological Research*, *57*, 63–69. http://dx.doi.org/10.1007/BF00447076.

Torgerson, W. S. (1961). Distances and ratios in psychophysical scaling. *Acta Psychologica*, *19*, 201–205. http://dx.doi.org/10.1016/S0001-6918(61)80072-3.

Vincent, R. J., Brown, B. R., Markley, R. P., & Arnoult, M. D. (1968). Magnitude estimation of perceived distance over various distance ranges. *Psychonomic Science*, *13*, 303–304. http://dx.doi.org/10.3758/BF03342597.

Wand, M. P., & Jones, M. C. (1995). *Monographs on statistics and applied probability*: *Vol. 60*. *Kernel smoothing*. London, United Kingdom: Chapman and Hall. http://dx.doi.org/10.1007/978-1-4899-4493-1.

Weiss, D. J. (1981). The impossible dream of Fechner and Stevens. *Perception*, *10*, 431–434. http://dx.doi.org/10.1068/p100431.

Yost, W. A. (1981). Variability in the measurement of sensory intensity. *Behavioral and Brain Sciences*, *4*, 211–212. http://dx.doi.org/10.1017/S0140525X00008530.

Zimmer, K. (2005). Examining the validity of numerical ratios in loudness fractionation. *Perception & Psychophysics*, *67*, 569–579. http://dx.doi.org/10.3758/BF03193515.

Zwislocki, J. J. (2009). *Sensory neuroscience: Four laws of psychophysics*. New York, NY: Springer. http://dx.doi.org/10.1007/978-0-387-84849-5.