* Manuscript

Empirical Evaluation of a Model of Global Psychophysical Judgments: IV. Forms for the Weighting Function

Ragnar Steingrimsson [a,1]

[a] *New York University, Department of Psychology, 6 Washington Place, New York, NY 10003*

R. Duncan Luce [b]

[b] *University of California, Irvine, Department of Cognitive Sciences, Social Science Plaza, Irvine, CA 92697-5100*

Abstract

Understanding the psychological interpretation of numbers is of both practical and theoretical interest. In classical magnitude estimation, respondents match numbers to sensations and in magnitude production they select tones that stand in a prescribed numerical ratio to a given tone. The present work focusses on evaluating several possible, and related, forms for this numerical distortion function, or weighting function, W. The main form, of which a power function is a special case, is the Prelec exponential/power representation. Behavioral equivalents to power and to Prelec functions are described, tested, and rejected. Thus, either the mathematical form or the assumption of $W(1) = 1$ is wrong. Whereas, the axiomatic literature has focussed exclusively on the former inference, we explore the possibility that $W(1) \neq 1$. Behavioral axioms are formulated in each case and experimentally tested. We conclude that most respondents satisfy a general power function and that those who do not, appear to satisfy the more general Prelec function.

Key words: audition; power functions; multiplicative form; numerical distortion; psychophysical regression effect; Prelec functions; reduction invariance; ratio production and ratio estimation; weighting function; sequential effects.

Email addresses: ragnar@nyu.edu (Ragnar Steingrimsson), rdluce@uci.edu (R. Duncan Luce).

[1] Now: University of California, Irvine, Institute for Mathematical Behavioral Sciences, Social Science Plaza, Irvine, CA 92697-5100

Any theory that attempts to deal with either magnitude estimation or magnitude production involves numbers or ratios that are produced, respectively, by the respondent or by the experimenter (see Stevens, 1975, for overview). One cannot safely assume that either one involves a veridical interpretation of numbers: some degree of distortion on the part of the respondent is to be expected. This article, the fourth in a series evaluating Luce's (2002, 2004) theory of global psychophysics in the auditory context, aims at greater understanding of the mathematical form of such a numerical distortion or weighting function, which we call W. It has also appeared as an unknown function in several other global psychophysical theories as well (see Section 1.2).

The article is structured as follows. Section 1 provides an abbreviated account of the context of the present article. Section 2 presents the theoretical issues that are evaluated in the article. There we focus on a class of possible functional forms for W, the Prelec ones (Section 2.3) of which power ones (Section 2.2) are a special case. Both were originally formulated and tested under the assumption of $W(1) = 1$, and under that assumption, both forms were rejected. After a bit, it dawned on us that the rejection might mean, instead of the power function or the more general Prelec being wrong, that $W(1) \neq 1$. For both classes of representation and within the context of a fairly general, empirically sustained psychophysical model, we then formulate behavioral equivalents of these two classes. These are tested experimentally.

The experimental portions of the article (Section 3) concerns three things. First, in Experiment 1, we analyze portions of the data collected in Experiments 2 in order to evaluate Naren's (1996) multiplicative form, (10), replicating and adding to previously published results, which concern, in part, the question of whether $W(1) = 1$—that property is soundly rejected. Second, Experiment 2 focuses on the power function form with $W(1) \neq 1$, which we find is well sustained for all but two respondents. Third, for those two respondents, Experiment 3 evaluated whether or not the generalized Prelec function gives a good fit; we conclude that it does.

Section 4 summarizes the main conclusion of the paper both separately and in the context of the previous three articles in the series.

There are a number of Appendices mostly concerned with various technical issues, but the one on sequential effects, Appendix D, is more substantive; however, those forms are not tested in this article.

1 Background psychophysical theory

1.1 Directly relevant background

Two theoretical articles (Luce, 2002, 2004) and three empirical ones (Steingrimsson and Luce, 2005a,b, 2006) underlie the present development. The two Luce articles arrived at a global psychophysical theory of intensities based upon "summations," such as presenting signals to the two ears, and on judgments of "productions," which generalize the standard ratio production method. In the general situation, pairs of signal intensities [2] (x, u), such as x to the left ear and u to the right of the same frequency and phase, are ordered by subjective intensity, e.g., loudness. The other primitive is a form of magnitude production where two signal pairs, (x, x) and (y, y), $y < x$, and a number p are presented and the respondent is asked to produce the signal, denoted $(z, z) = (x, x) \circ_p (y, y)$, such that the subjective difference between (z, z) and (y, y) is perceived to be p times the given "interval" from (y, y) to (x, x).

Luce (2002, 2004) gave behaviorally testable axioms that lead to the following numerical representation: An order-preserving psychophysical function $\Psi(x, u)$, i.e.,

$$(x, u) \succsim (y, v) \Leftrightarrow \Psi(x, u) \geq \Psi(y, v), \tag{1}$$

$$\Psi(0, 0) = 0, \tag{2}$$

which has the *p-additive* form

$$\Psi(x, u) = \Psi(x, 0) + \Psi(0, u) + \delta\Psi(x, 0)\Psi(0, u) \quad (\delta \geq 0). \tag{3}$$

In addition there is a numerical distortion function $W(p)$ satisfying

$$W(p) = \frac{\Psi\left[(x, x) \circ_p (y, y)\right] - \Psi(y, y)}{\Psi(x, x) - \Psi(y, y)} \quad (x > y \geq 0). \tag{4}$$

Moreover, under some assumptions, but not those for the symmetric case evaluated here, if $\delta > 0$, then for some constant $\gamma > 0$,

$$\Psi(x, 0) = \gamma\Psi(0, x). \tag{5}$$

Steingrimsson and Luce (2005a,b) experimentally evaluated the underlying behavioral axioms and found good support for them and so for (3) and (4).

Steingrimsson and Luce (2006) examined one possible mathematical form for Ψ, namely a sum of power functions, and although there was considerable

[2] The stimuli are measured, for each ear, in terms of the physical intensity, not dB, presented less the threshold intensity.

support for that, clear exceptions existed. We lack a behavioral equivalent for $\delta \neq 0$ and $\Psi(x, 0)$ and $\Psi(0, u)$ are power functions, and so that case, which we suspect may hold, has not been evaluated.

For this article, we only need the production representation, (4), and that only for $y = 0$. We focus on two possible forms for W with results both about their theoretical and empirical features. The major, relevant, feature of the general theory is (4) for $y = 0$, i.e.,

$$\Psi\left[(x, x) \circ_p (0, 0)\right] = \Psi(x, x)W(p), \tag{6}$$

which property is called *separability.* It is a multiplicative conjoint representation whose underlying qualitative axioms are well understood (Krantz, Luce, Suppes, & Tversky, 1971, Ch. 6), and so it can be evaluated by well developed methods of conjoint measurement. So far, that has not been done.

1.2 Historically relevant background

The earliest explicit theoretical effort that we know of to tease apart the psychophysical function and a separate weighting function was Attneave (1962). He proposed a two-stage model of magnitude estimation, where in the first stage, stimuli are mapped into psychological magnitudes or sensations and in the second stage, the psychological magnitudes are mapped into the number continuum. Attneave (1962) assumed that both transformation functions, stimuli to sensations (the psychophysical function) and sensations to numbers (the weighting function), were power functions. Indeed, Steingrimsson and Luce (2006) derived from the model described in Section 1.1 expressions for magnitude and production estimation that are just such a compounding of functions (see Eqs. 16 and 17 of that paper).

The Attneave (1962) model was supported empirically by Rule and Curtis (1973) using nonmetric conjoint scaling of judgments of the subjective magnitude of weights with the integers 1-9. Similar support was provided by Schneider et al. (1974), using multidimensional scaling on similarity judgments. Other efforts include Baird, (1975), who suggested that W, although well approximated overall by a power function, is not really a more-or-less continuous map of natural numbers (or even just of integers) into the real numbers; rather, it is onto a quite discrete subset of the real numbers, usually rounded to 5s and 10s. Banks and Hill (1974) used the apparent magnitude of number scaled by random production and Banks and Coleman, (1981) used data where the respondent rated the perceived randomness of a set of numbers. Although these theoretical and experimental efforts employed very different methods, they all had in common that in one way or another they attempted to fit data to functions and they were unanimous (with some caveats) in supporting a power function $W(p) = \alpha p^{\beta}$, rather than, say, a logarithmic or a linear transformation, as the weighting function. Indeed Schneider et al. (1974) suggested how exponents obtained using magnitude estimation must be "corrected" based on the exponent of the number continuum. Additionally, they suggested the possibility that differences in exponents obtained in magnitude estimation experiments for different respondents were "actually due to different psychological number scales rather than differences in the underlying representation of signals" (p. 46).

This was also the approach taken by Narens (1996) in his formulation of what he believed Stevens (e.g., 1975) might have meant theoretically in his magnitude estimation methods. Narens (1996) states explicitly that Stevens underlying assumption must have been that respondents treated numbers as if $W(p) = p$. Narens' axioms yield the slightly weaker power function $W(p) = p^{\beta}$, an assumption he showed justifiable under what amounted to the assumptions of $W(1) = 1$ and the multiplicative property, (10) below, both holding. It should be noted that those working in the Attneave (1962) tradition and indeed anyone using function fitting techniques did not assume $W(1) = 1$.

2 Functional forms for W, behavioral equivalents, and some implications

2.1 Power function with $W(1) = 1$

The key behavioral property that we work with is that for either $p > 1, q > 1$ or $p < 1, q < 1$, which together are equivalent to $(p-1)(q-1) > 0$,

$$[(x, x) \circ_p (0, 0)] \circ_q (0, 0) \sim (x, x) \circ_t (0, 0). \tag{7}$$

The left side is a compound production of the proportion p applied to stimulus (x, x) followed by the proportion q applied to that, and, the right side the proportion $t = t(p, q, x)$ needed to match the compound one.

The following Proposition establishes that, under separability, (6), $t(p, q, x)$ does not depend upon x.

Proposition 1 *If separability, (6), holds, then (7) is equivalent to*

$$W(t) = W(p)W(q). \tag{8}$$

The nearly trivial proof is in Appendix A.

2.1.1 Production commutativity and the multiplicative form

Narens (1996), in his formulation of what he thought underlay the theoretical thinking of S. S. Stevens (e.g.,1975), arrived at two predictions: The first is a form of *threshold production commutativity* also derived by Luce [2004, Eq. (21)]

$$[(x, x) \circ_p (0, 0)] \circ_q (0, 0) \sim [(x, x) \circ_q (0, 0)] \circ_p (0, 0), \tag{9}$$

or in terms of (9)

$$t(p, q) = t(q, p).$$

The second prediction in Narens (1996) is the *multiplicative form* [3]

$$t = t(p, q) = pq. \tag{10}$$

It is easy to show (Aczél, 1966, p. 41) that (10) is equivalent to the existence of a constant $\omega > 0$ such that

$$W(p) = p^{\omega}. \tag{11}$$

Because (10) is a special case of the more general *k-multiplicative property* (14), below, with $k = 1$, we may call (10) the *1-multiplicative property* or *1-MP* for short.

2.1.2 Published empirical tests of the 1-multiplicative property

Ellermeier and Faulhammer (2000) for $p > 1, q > 1$ and Zimmer (2005) for $p < 1, q < 1$ have empirically examined both predictions (9) and (10). Each article sustained [4] production commutativity, (9), whereas both unambiguously rejected the multiplicative form, (10), 1-MP. In Section 3.2 we replicate, using a different methodology, both previously explored conditions.

Note carefully that if $t(p, q) = pq$, fails in general, then specifically it cannot be assumed to hold for particular cases, e.g., $t(1, q) = 1 \times q$, which is tantamount to

$$W(1) = 1. \tag{12}$$

So the data imply that if a power function is correct at all, it is with $W(1) \neq 1$.

As mentioned, although the earlier authors in the function fitting tradition did not assume (12), most authors in the axiomatic school, including us, Narens (1996), and more recently DeCarlo (in press), have implicitly assumed (12). Therefore empirical failure of the multiplicative form was rather unexpected

[3] Luce (2002), p. 525, termed it the probability-reduction property.

[4] The generalization of (7) where 0 is replaced everywhere by y, $0 < y < x$, was also sustained by Steingrimsson and Luce (2005a).

and presents some theoretical problems. Thus, we first consider modifications of the power function theory so as to eliminate (12). Our arguments rest on assuming that W is strictly increasing and onto the positive real numbers, and so must be continuous. There is, of course, the possibility that W is simply discontinuous at 1 in the sense that either $\lim_{p \nearrow 1} W(p) < 1$ or $\lim_{p \searrow 1} W(p) > 1$. We do not have data that suggest this. So, we turn to the more general case.

Another source of doubt about (12) is the phenomenon of time-order error. In standard matching experiments, where one asks the respondent to state the intensity z such that z stands in the ratio 1 to a given signal x, then z is usually different from x (see Hellström, 1985 and 2003, for a survey and recent data). Thus, if (6) is a correct model for matching, the existence of the time-order error means that $W(1) \neq 1$. One question is whether the general power function model of the next section leads to results comparable to the data. We take that up in Steingrimsson & Luce (in preparation). We do not go into it further here.

2.2 General power function without $W(1) = 1$

2.2.1 General multiplicative form

Consider the general power function form:

$$W(p) = W(1) \begin{cases} p^{\omega}, & 0 < p \leq 1 \\ p^{\omega'}, & p > 1 \end{cases} \qquad (\omega > 0, \omega' > 0), \tag{13}$$

An argument for this form can be based on some magnitude estimation ideas: see Appendix B. A better argument is given next.

The generalization of Proposition 1 to $W(1) \neq 1$ replaces (10) with the a more general form that we call the *general multiplicative form*.

Proposition 2 *Suppose that (7) and (8) are satisfied. Then, the following are equivalent:*

1. *The power function form (13) obtains.*
2. *The general multiplicative form,*

$$t(p,q) = pq \begin{cases} k, & p < 1, q < 1 \\ k', & p > 1, q > 1 \end{cases} \qquad (k > 0, k' > 0) \tag{14}$$

is satisfied.

Corollary 1 *Under the conditions of the Proposition, either*

(i) $\omega' = \omega$ *and* $k' = k$,

(ii) *or,*

  (a) *when* $p < 1, q < 1$, *then* $kpq < 1$,

  (b) *when* $p > 1, q > 1$, *then* $k'pq > 1$.

  In all cases,

$$k = W(1)^{1/\omega}, \; k' = W(1)^{1/\omega^*}. \tag{15}$$

The proof is in Appendix C.

We refer to the condition (14) as the $k-$*multiplicative property* or k-*MP* for short.[5] We empirically evaluate k-MP in Experiment 2.

Assuming the $k-$MP property, we see that (7) holds in the mixed cases $p > 1 > q$ or $p < 1 < q$ iff $\omega = \omega'$. In general, however, nothing simple is predicted for these cases. In practice, this is not problematic: Ham, Biggs, and Cathey (1962) in an analysis of multiple and large data sets of magnitude estimation with fractionation and multiples consistently found, in our notation, $\omega \neq \omega'$, a result also easily inferred from Fig. 9 of Hellman & Zwislocki (1961). In addition, data that we collected and present in Appendix E, lead to the conclusion that $\omega \neq \omega'$. In the light of an unclear theoretical path and substantial evidence against $\omega = \omega'$, it seems pointless to pursue the mixed case very deeply.

Although knowing the value of k alone does not allow us to estimate $W(1)$, it is useful to note that from (14) and (15) that it is easily inferred that since k and ω are positive constants, $k < (>)1 \iff W(1) < (>)1$ and similarly, $k' < (>)1 \iff W(1) < (>)1$.

2.3 Prelec's weighting function with $W(1) = 1$

Within the context of utility theory for risky gambles, Prelec (1998) proposed and axiomatized an interesting weighting function of the form of an exponential of a power of $-\ln p$, (16), and Luce (2001) offered a simpler axiomatization, for $0 < p \leq 1$. Both authors assumed, without comment, that $W(1) = 1$, which in the utility context may be justified, whereas in the psychophysical context, it appears empirically wrong (see Section 2.1.2, Experiment 3.2). Recently, Aczél and Luce (submitted) extended these results to the case where $W(1) \neq 1$. We present each result in turn.

The original Prelec form of the weighting function, generalized from the unit

[5] Obviously, 1-MP, (10), is just the special case of (14) were $k = 1$.

interval to all positive numbers, is

$$W(p) = \begin{cases} \exp\left[-\omega\left(-\ln p\right)^{\mu}\right] & (0 < p \leq 1) \\ \\ \exp\left[\omega'\left(\ln p\right)^{\mu'}\right] & (1 < p) \end{cases}, \tag{16}$$

where the constants satisfy $\omega > 0, \omega' > 0, \mu > 0, \mu' > 0$.

Equivalence to reduction invariance (RI)

Luce (2001) proved the following.

Proposition 3 *Suppose that separability, (6), is satisfied and that W is a strictly increasing function. The following two statements are equivalent:*

1. *The weighting function is given by the Prelec form, (16).*
2. Reduction Invariance *Suppose that positive p, q, t are such that (7) is satisfied for all positive x. Then for any natural number*[6] *N*

$$[(x, x) \circ_{p^N} (0, 0)] \circ_{q^N} (0, 0) = (x, x) \circ_{t^N} (0, 0). \tag{17}$$

Note that reduction invariance—or RI in shorthand—is a behavioral condition. In words, if the compounding of p and q in magnitude productions is the same as the single production of $t = t(p, q)$, then the compounding of p^N and q^N is the same as the single production of t^N.

2.4 General Prelec weighting function without $W(1) = 1$

Without constraints on $W(1)$, (16) becomes (18):

$$W(p) = W(1) \begin{cases} \exp\left[-\omega\left(-\ln p\right)^{\mu}\right] & (0 < p \leq 1) \\ \\ \exp\left[\omega'\left(\ln p\right)^{\mu'}\right] & (1 < p) \end{cases}, \tag{18}$$

where $\omega > 0, \omega' > 0, \mu > 0, \mu' > 0$. This is simply $W(1)$ times the Prelec function (16).

[6] Actually, it is only necessary for it to hold for integer pairs, such as 2 and 3, that do not have a common integer factor. From this apparently much weaker assumption and the strictly increasing character of W, one proves that the condition holds for all positive, real exponents in place of N.

Note that when $\mu = \mu' = 1$, (18) reduces to the power function (13). For either $\mu \neq 1$ or $\mu' \neq 1$, the form of W can be far more complex, including S- and inverse S-shaped. Our interest will be in finding behavioral equivalents and testing them.

Equivalence to double reduction invariance (D-RI)

Aczél and Luce (submitted) proved the following

Proposition 4 *Under the same assumptions as for Proposition 3, the following two statements are equivalent:*

1. *The weighting function is given by the general Prelec form, (18).*
2. Double Reduction Invariance *Suppose that positive p, q, t are such that*

$$[(x,x) \circ_p (0,0)] \circ_q (0,0) \sim [(x,x) \circ_t (0,0)] \circ_t (0,0) \tag{19}$$

is satisfied for all positive x. Then for any natural number N

$$[(x,x) \circ_{p^N} (0,0)] \circ_{q^N} (0,0) \sim [(x,x) \circ_{t^N} (0,0)] \circ_{t^N} (0,0). \tag{20}$$

Corollary 2 *If the Proposition holds, then if $(1-p)(1-q) < 0$, then $\omega' = \omega$ and $\mu' = \mu$.*

Double reduction invariance differs from RI (17) only in that the right side is compounded twice, just like the left side but with the same t, hence the term double reduction invariance, or D-RI in shorthand. In words, if the compounding of p and q in magnitude productions is the same as double production of $t = t(p,q)$, then the compounding of p^N and q^N is the same as the double compound of t^N.

3 Experiments

We present three experiments, evaluating, in turn, 1: 1-MP (10), 2: k-MP (14), and 3: D-RI (20).

3.1 Experimental methods

The 3 experiments reported share a number of testing strategies that are now outlined. Specifics to each experiment are described later as relevant. Most of the methods employed here are identical to those used in Steingrimsson and

Luce (2005a), hence only an abbreviated account of those methods is provided here.

3.1.1 Signal presentations

The experiments were carried out in the auditory domain using 1,000 Hz sinusoidal tones presented for 100 ms, which included 10 ms on and off ramps.

The theory is cast in terms of intensities less threshold, i.e., with a threshold of x_τ, a signal intensity of x', and a stimulus (x, u), the intensity for the left ear is $x = x' - x_\tau$. Similarly, for the right ear we have $u = u' - u_\tau$. In experimental descriptions we report x' in dB SPL, denoted x'_{dB}, rather than $x_{\text{dB}} = (x' - x_\tau)_{\text{dB}}$, and likewise for u'. Because all signals were well above threshold and our respondents normal hearing, for which they were selected, the resulting errors are negligible.

3.1.2 Notational convention

Because we only used symmetric stimuli, we simplify in experimental descriptions our notation by simply writing x for (x, x). In that notation, our goal is to obtain an estimate $v = x \circ_p 0$, which is the special case of the general ratio production $v = x \circ_p y$, where $y = 0$.

3.1.3 Respondents

A total of 12 students—graduate and undergraduate, 3 males and 9 females—from New York University participated in the experiments reported in this article. The first author (R22) was one of them [7]. All respondents were within 20 dB of normal hearing thresholds (ANSI, 1996) in the range 250–8000 Hz, assessed by an audiometric test (Micro Audiometric EarScan ES-AM).

All respondents, except for the first author, were compensated $10 per session. Each person provided written consent and was treated in accordance with the "Ethical Principles of Psychologists and Code of Conduct" (American Psychological Association, 2002). Consent forms and procedures were approved by NYU's and UC Irvine's Institutional Review Boards.

[7] We judged this acceptable because the behavioral measures of ratio production are not determined by the experimental design.

3.1.4 Equipment

Stimuli were generated digitally using a personal computer and played through a 24-bit digital-to-analog converter (RP2.1 Real-time processor, Tucker-Davis Technology). Presentation level was controlled by built-in features of the RP2.1 and stimuli were presented over Sennheiser HD265L headphones to listeners seated in individual, double-walled, IAC sound booths.

An 85 dB SPL safety limit was imposed in all experiments.

3.1.5 Procedure

Experiments were conducted in sessions lasting no more than one hour. All respondents completed one training session with ratio productions. Since some observers participated in multiple experiments, the total practice that individuals had prior to any one experiment varied substantially. Depending on the experiment, practiced respondents typically completed 60-64 estimates per session.

Estimating the $\circ_p$ operation

Let $\langle A, B \rangle$ denote a presentation of A followed by a temporally displaced presentation of B. Here, a temporal delay of 450 ms between A and B was used. An estimate of $x \circ_p 0 = v$ is, thus, obtained using the trial type

$$\langle (x, x), (v, v) \rangle \tag{21}$$

where the value of v is under the respondent's control. In practice, at the beginning of an ratio production, the value of the proportion p was displayed on the monitor. Then, the respondent heard the tone (x, x) followed, 450 ms later by the tone (v, v). Then the respondent used key presses to either adjust the intensity of v, to repeat the previous trial, or to indicate satisfaction with the ratio production. Adjustments were freely done in steps of 0.5, 1, 2 or 4 dB.

After an intensity adjustment, the altered tone sequence was played. The process was repeated until the respondent indicated satisfaction with the estimate.

Instructions to respondents consisted of a description of the task coupled with graphical examples in which, e.g., the intensity of (x, x) was represented as a reference bar, and the task of producing a tone (v, v) with $p = 200\%$ and $p = 60\%$, respectively, were represented by another bar whose height was 200% and 60%, respectively, times the reference bar. In verbal instructions to respondents, the task was explained as that of making the second stimulus a given percent of the first.

3.1.6 Statistical method

As the previous three articles, we again examine a number of parameter-free null hypotheses of the form $L_{\text{side}} = R_{\text{side}}$. This reflects the nature of the empirical axioms being tested. If the hypothesis $L_{\text{side}} = R_{\text{side}}$ is correct, it is equivalent to asserting that both L_{side} and R_{side} are drawn from the same distribution. Because we do not yet have a theory that predicts the distributions of our estimates, we chose the nonparametric Mann-Whitney U test for our statistical evaluation, with a significance level of .05.

Given that no distributional assumption is made, it would be preferable to report medians to means. However, the discrete nature of the signal values renders it difficult to find accurate estimates of medians, thereby making the mean a better estimator provided that the distribution of signal values is approximately Gaussian, which they appear to be. So we report means. To indicate variability in adjustments, we report the standard deviations.

A particular concern is whether the sample sizes for L_{side} and for R_{side} are sufficiently large so that a true failure of the null hypothesis can be distinguished within the power of the statistical method employed. To address this issue all statistical results were verified using Monte Carlo simulations based on the bootstrap technique (Efron & Tibshirani, 1993) (see Steingrimsson & Luce, 2005a, for details). We asked whether L_{side} and R_{side} could, at the .05 level, be argued to come from the same underlying distribution. This was our criterion for accepting the null hypothesis as supporting the behavioral property.

3.2 Experiment 1: 1-multiplicative property (1-MP)

As discussed in Section 2, Narens (1996) formulated, as a part of what he believed to underlie the theoretical thinking of S. S. Stevens (e.g., Stevens, 1975), the multiplicative form (10), namely $t = pq$, which has the associated behavioral propriety of compound production, (7).

The testing relies on data from Experiment 2, but the logical progression of what is being tested makes this the natural first experiment to present.

3.2.1 Method

In Experiment 2 we collected data for the compound production (7). Several estimates of $v = (x \circ_p 0) \circ_q 0$ where made and we then obtained the estimate $\hat{t}$

of t such that $\overline{v} = x \circ_{\hat{t}} 0$, where $\overline{v}$ was the average of the v's. This means that if 1-MP holds, then $\hat{t} = pq$ should hold statistically. Thus, we can test 1-MP by reanalyzing these data.

The estimate $\hat{t}$ was obtained using the Up-Down staircase, by design the data in an Up-Down process are cluster symmetrically around the average of the run, hence we can use a one-sample t-statistic to evaluate whether $\hat{t} = pq$. For further details, see Section 3.3.1

3.2.2 Result

Data were derived from 12 respondents. For $p < 1$, $q < 1$, the property was rejected in 14/18 and for $p > 1$, $q > 1$,was rejected in 19/21 tests, i.e. rejected in 33/39 tests overall.

3.2.3 Discussion

The failure of 1-MP in 33 of our 39 tests strongly agrees with the conclusions of previously published data (Ellermeier and Faulhammer, 2000 for $p > 1, q > 1$ and Zimmer, 2004 for $p < 1, q < 1$). The rejection of 1-MP is equivalent to a rejection of the multiplicative form (10) and thus W as the power function in (11).

Of some interest is whether these failures form a discernible pattern. To address this question, we explored the relative relationship between the pq and $\hat{t}$ by tallying the number of cases for which the statistical trend suggests $pq \gtreqless \hat{t}$. In addition, it seemed reasonable to compare these results to previously published data so we added those to the tally.[8] These results are shown in Table 1.

Table 1
Experiment 1: Summary of existing data on the relative relationship of pq to $\hat{t}$

Row	Condition	$<$	$>$	$=$	Comment
1	1: $p < 1, q < 1$	9	5	4	Experiment 1
2	1: $p < 1, q < 1$	0	17	2	Zimmer (2004)
3	2: $p > 1, q > 1$	2	17	2	Experiment 1
4	2: $p > 1, q > 1$	15	1	1	Ellermeier & Faulhammer (2000)

[8] Ellermeier and Faulhammer (2000) collected data for $v = (x \circ_2 0) \circ_3$ and $v' = x \circ_6 0$, using two-ear productions. When they found that $v > (<) v'$, we assume that a larger (smaller) proportion t would be required in order for $v = v'$. Thus when $v > (<) v'$, we assume it means $pq < (>) t$. Using similar methodology, Zimmer (2004) collected data for $v = (x \circ_{1/2} 0) \circ_{1/3}$ and $v' = x \circ_{1/6} 0$ and, by the previous logic, when $v < (>) v'$, we assume it means $pq > (<) t$.

Assuming that k-MP holds, then when $pq < (>)\hat{t}$ it means that $k > (<)1$, which if (15) holds is equivalent to $W(1) > (<)1$.

Particularily notable in Table 1 is the asymmetries between rows 1 and 2 and between rows 3 and 4, which suggests, on average, quite different direction for $W(1)$ within the two consistent p, q conditions.

Although statistically less than plausable, we cannot exclude the possibility that this is due to the test coming from different respondents. We have no information about whether there is an overlap between respondents in the two published studies. In our case, there are three respondents that participated in both Condition 1 and 2. In Appendix E we explore these three respondent in detail and conclude that they are not clearly inconsistent across the two conditions.

Regardless, the contrast between Conditions 1 and 2 and between Conditions 3 and 4 are so contradictory and lack any straightforward theoretical explanation that they invite careful thought. It seems prudent to look first for methodological explanations: We note that the results in rows 1 and 3 shared a common procedure whereas rows 2 and 4 (largely) shared a different one. This suggests two tests of interest: reproduce the Ellermeier and Faulhammer (2000) and Zimmer (2004) studies (i) by using their stimuli and our method and (ii) by using their method and our stimuli. However, because this inconsistency is not a focus of our present investigation, we do not attempt to address it further here, but merely note that it deserves further exploration.

3.3 Experiment 2: kmultiplicative property (k-MP)

We have that W has the general power function form (13) iff p, q, and t satisfy the k-MP (14):

$$t = kpq \text{ where } k = \begin{cases} W(1)^{1/\omega}, & p > 1, q > 1 \\ W(1)^{1/\omega^*}, & p < 1, q < 1 \end{cases} .$$

Because k is a constant independent of p and q the property can be tested by estimating $k = t/pq$ for several values of p and q and evaluate whether the obtained k's are equal. For this purpose, recall, using the convention of Section 3.1.2, that the compound ratio production, (7), is written

$$(x \circ_p 0) \circ_q 0 \sim x \circ_t 0.$$

So we can test the property by obtaining estimates $\hat{t}$ of t for several combinations of p and q and from those calculate k followed by a statistical evaluation of whether the k's are equal.

3.3.1 Method

The task is to obtain an estimate for the compound ratio production,(7), reproduced above. The estimation was done in two steps, S1 and S2.

S1 For given x, p, and q collect an estimate of $w = (x \circ_p 0) \circ_q 0$. This requires making the two successive estimates: $v = x \circ_p 0$ and, using v, $w = v \circ_q 0$. In practice, estimates $v_1, v_2, ..., v_j, ..., v_m$ were collected intermingled with estimates $w_1, ...w_m$ in such a way that $w_j = v_j \circ_q 0$. That is, each instance of v_j was used in the consequent estimation of w (this allows variance to propagate from one estimation step to the next. See Appendix E and Steingrimsson & Luce, 2005a, Appendix A.4 for details). Each estimate was obtained using the trial type given by (21).

S2 Using $\overline{w}$, the average of the w_m's from Step 1, estimate a proportion t such that $x \circ_t 0 \sim \overline{w}$. We used the simple Up-Down method (Dixon and Moon, 1948; Wetherill, 1963; Levitt, 1971) to arrive at this estimate. Briefly, an initial proportion of t_0 was presented and participants produced a corresponding intensity $w' = x \circ_{t_0} 0$, using a trial form given by expression (21). If $w' < \overline{w}$, then t_0 was increased by constant amount δ otherwise t_0 was decreased by the same amount; the new value is t_1. This process was repeated 35 times and the average of the last 30 values t_5 to t_{35} was taken as an estimate for t. Different staircases were interwoven within a block. A formal description is provide in Appendix F.

Proportions were presented as percentages, e.g., $p = 1.6$ was presented as 160%. In Step 2, $\delta = 10\%$. Data for each of the two steps were collected within a block of trials and these blocks were run in two corresponding sessions.

Data were collected using two sets of parameters

Condition	x dB	p%	q_1%	q_2%	q_3%
1: $p < 1$, $q < 1$	75	80	60	40	20
2: $p > 1$, $q > 1$	66	130	160	200	300

With one common p and three q's, each of the two conditions gave rise to the three estimates k_1, k_2, k_3

$$t_n = k_n p q_n, \quad n = 1, 2, 3 \tag{22}$$

Using the data, we estimate the k_n's, which from (22) is $k_n = t_n / pq_n$. Clearly, if k is a constant, then the k_n's form a line with a slope 0.

At first glance, this is easily explored by using linear regression of k_n on pq_n and statistically testing whether the slope parameter equals 0. However, the way in which each t_n is estimated poses a problem of statistical power. Specifically, each t_n is estimated based on the long-run average of the Up-Down processes. Thus, even though the staircase consists of (typically) 35 trials, only the final average can be regarded as a stable estimate of t_n, thus with one staircase per estimate, we have, for the purpose of statistical testing, only one observation, albeit a robust one, of each. Although estimating more than 3 t's would make for a more accurate slope estimate, that would not fundamentally improve the statistical situation. To increase the number of estimated t_n's, regardless of the range of n, would effectively mean collecting repeated data from Step 1 and 2 for each subject, logistically a rather formidable task. [9]

As an alternative, we take the following approach. Let r denote the respondent; let $k_{n,r}$ be a data set for r under condition pq_n; and let $\overline{k_r}$ be the mean of the 3 k_n estimates for respondent r and let $\overline{pq}$ be the mean of the pq_n. Then we plot all of the data for all respondents in terms of $k_{n,r} - \overline{k_r}$ against $pq_n - \overline{pq}$. By design, when fit to a line, these data will all have an intercept at $y = 0$. Thus, we can use the linear regression $k_{n,r} = apq_n$, where some statistical power is gained by removing the intercept term. From the end of Section 2.2, we know that $\overline{k_r} < (>)1 \iff W(1) < (>)1$. Thus, in addition, we examine the value of $\overline{k_r}$.

3.3.2 Results

Data were collected from 8 respondents and a total of 13 data sets were obtained. Of those, two data sets were excluded: one for Condition 1 due to the level of volatility suggesting that the respondent had severe difficulty with the task, and for one in condition 2 due to the safety limit's repeatedly being hit, making the "true" data inaccessible. The resulting 11 data sets are plotted in Figure 1.

Note, in both graphs, the y-axis ($k_{n,r} - \overline{k_r}$) has the same range, whereas the x-axis ($pq_n - \overline{pq}$) is varied to fit the different proportion ranges.

In Table 2, the estimated slopes for each participant are listed along with the result of statistically asking whether $a = 0$. In the last column the average $\overline{k_r}$ for each participant is indicated.

[9] Another formidable problem involved here is the fact that, although the pattern of responses in ratio productions between sessions is generally stable, the values produced are not at all (see Appendix G; Steingrimsson & Luce, 2005a, Appendix A.2, for details). Thus, the problem is not simply one of collecting multiple sessions for each step, but ideally—and as we attempt in practice—to collect instances of all related ratio productions within a session, in closely spaced sessions, to inspect inters-session results for numerical consistency etc.

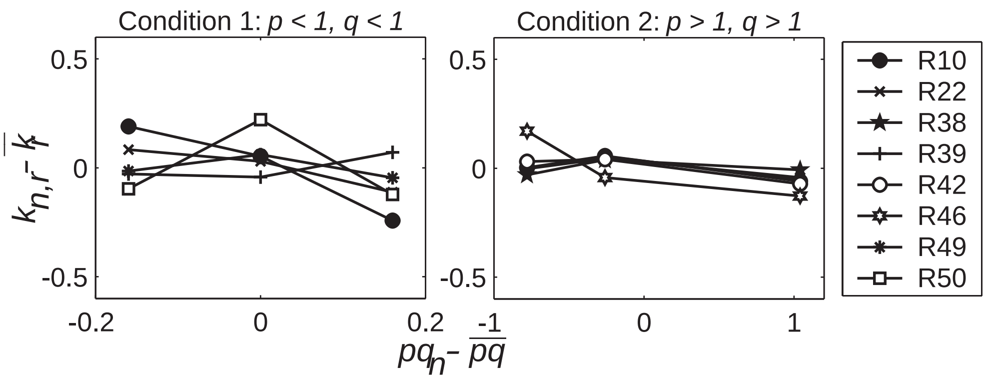

Fig. 1. Experiment 2: Results for testing the k-multiplicative property

Table 2
Experiment 2: Testing generalized multiplicative property. Result of the regression of $k_{n,r} = apq_n$

Resp.	Slope a	p_{stat}	Statistical trend	$\overline{k_r}$
Condition 1: $p < 1, q < 1$				
R10	-1.375	.022	$a \neq 0$	1.03
R22	-0.608	.033	$a \neq 0$	1.23
R39	0.309	.201	$a = 0$	1.03
R49	-0.098	.718	$a = 0$	0.80
R50	-0.081	.932	$a = 0$	0.96
Condition 2: $p > 1, q > 1$				
R10*	-0.044	.938	$a = 0$	0.91
R22	-0.031	.345	$a = 0$	0.82
R38	0.003	.925	$a = 0$	0.92
R39	-0.039	.311	$a = 0$	0.72
R42	-0.063	.064	$a = 0$	0.81
R46	-0.145	.116	$a = 0$	0.72

*Data average over three separate runs.

In sum, k-MP is found to hold in Condition 1 for 3/5 respondents and Condition 2 for 6/6. In Condition 2, the value of $\overline{k_r}$ is less than 1 for all, but somewhat mixed in Conditions 1.

3.3.3 Discussion

In general, variability tends to increase as stimulus intensity decreases. Thus one expects variability to be higher in Condition 1 ($p < 1, q < 1$) than in 2 ($p > 1, q > 1$), which clearly is the case.

In Condition 2, the results are similar for all respondents and the property is not rejected for any of the 6 respondents. Furthermore, for all but (possibly) one respondents, R46, the numerical values of the slopes are exceedingly close to 0, lessening any lingering worry about the result being an artifact of low statistical power.

Worthy of note is that the numerical direction of the slopes is negative for 5/6 respondents and for the remaining one, effectively 0, a priori giving rise to a suspicion of a slight bias in the data. As it happens, just such bias is expected due to the time-order error being a function of intensity (Steingrimsson & Luce, in preparation). In any case, it seems that over a fair range, k-MP holds well in this condition.

In Condition 1 the evidence is more mixed, with 2/5 rejecting k-MP. The 3/5 who do not reject k-MP behave much as in Condition 2. This raises the intriguing question: do the two respondents, R10 and R22, who reject k-MP, pass D-RI (20)? Both pass k-MP in Condition 2 so we certainly would expect them to pass D-RI for that condition (because a general power function is a special case of the general Prelec function) and were they to pass it for Condition 1, it would mean we had a complete description for the behavior of all respondents in both Condition 1 and 2. Indeed we did collected these data (Experiment 3), and found the property held for R10 and R22 in both Condition 1 and 2.

Note carefully, that aside from R22, who in any event does not appear to satisfy $k-$MP in Condition 1, only R10 and R39 in Condition 1 have $\overline{k_r} > 1$, but their estimated values of 1.03 are so close to 1 that one has little confidence that k_r itself is really > 1. The data, therefore, are not inconsistent with the assumption that $W(1) < 1$ in both conditions. Of course the estimated ω and ω' vary considerably over respondents and do not appear to be equal—see Appendix E for details.

The three main conclusions are:

- The property of k-MP holds well when $p > 1, q > 1$, and holds for 3 of the 5 respondents when $p < 1, q < 1$. For these respondents, this is equivalent to the weighting function W being a generalized power function (13)—for some in the condition $p < 1, q < 1$, we will see in the next experiment that the general Prelec function yields a better description (Experiment 3).
- At least for those for whom W is a generalized power function, we have

$W(1) < 1$.

- The form of the weighting function, as a general rule, differs for proportions above and below 1

3.4 *Experiment 3: Testing reduction invariance (RI) and double reduction invariance (D-RI)*

We collected data on RI prior to (i) realizing that the assumption of $W(1) = 1$, on which the RI property rests, was likely wrong, (ii) the discovery of D-RI, and (iii) having the results of Experiment 2. In hindsight we would anticipate that the RI property to fail overall, if for no other reason than $W(1) \neq 1$. Suffice it to summarize that Zimmer (2005), who was first to empirically test RI, used proportions $p < 1$, $q < 1$ and found the property rejected for 4/7 respondents and using a stricter method of analysis, for 6/7. We tested RI with 7 respondents for both $p < 1$, $q < 1$ and $p > 1$, $q > 1$ and rejected it overall in 15/24 tests.

The failure of k-MP for R10 and R22 in Experiment 2 leaves D-RI as a test to be done. Thus, we proceed to testing D-RI for R10 and R22.

For reference and using the notation convention of Section 3.1.2 D-RI (20) is given by

$$(x \circ_{p^N} 0) \circ_{q^N} 0 \sim (x \circ_{t^N} 0) \circ_{t^N} 0.$$

3.4.1 *Method*

Testing was done in three steps, S1-3, using two-ear productions.

S1 With a given x, p, and t, estimate $v = x \circ_p 0$ and $w = (x \circ_t 0) \circ_t 0$ in a manner analogous to that of Step 1 of Experiment 2.

S2 With the average of the v's from Step 1, estimate a proportion q such that $v \circ_q 0 \sim \overline{w}$, the average of the w's—method analogous to Step 2 of Experiment 2.

S3 With the estimate of q from Step 2, obtain estimates with proportions raised to the power N. The value of N was chosen such that it would provide proportions close to a multiple of five for each of p^N, $\widehat{q}^N$, and t^N and the result was rounded to the nearest 5% [10]. Then, we obtained the estimates

[10] The reason for this rounding was mainly due to respondents' not being particularly comfortable with percentage values that were finer than the rounding to the nearest 5%. This is not odd as the maximal range of rounding (2.5%) is less than JND.

of $L_{\text{side}} = (x \circ_{p^N} 0) \circ_{\widehat{q}^N} 0$ and $R_{\text{side}} = (x \circ_{t^N} 0) \circ_{t^N} 0$. L_{side} and R_{side} in the same manner as w in Step 1.

The property is said to hold if L_{side} and R_{side} are found statistically indifferent.

Data were collected using two sets of parameters

Condition	x dB	p%	t%
1: $p < 1$, $q < 1$	75	80	50
2: $p > 1$, $q > 1$	62	150	200

In this experiment, p and t are given and q is estimated. Data for each of the three step were collected within a block of trials and these blocks were run in three corresponding sessions.

3.4.2 Results

Data from R10 and R22 are reported in Table 3. In Table 3 are listed, by respondent and condition, the estimate of q, the value chosen for the power N, the means and standard deviations of the intensities whose equality is statistically tested along with the result of those tests.

Table 3
Experiment 3: Testing double reduction invariance. In Steps 1, 2, & 3, $n = 40, 55, 30$, respectively.

Resp.	Step 2	Step 3				
	$\widehat{q}$%	N	Mean (s.d.)		p_{stat}	Statistical trend
			$\widehat{L}_{\text{side}}$	$\widehat{R}_{\text{side}}$		
		Condition 1: $p < 1$, $t < 1$				
R10	29.4	0.64	63.29 (2.64)	63.50 (1.93)	.569	$L_{\text{side}} = R_{\text{side}}$
R22	25.0	0.50	67.00 (2.63)	67.33 (2.81)	.333	$L_{\text{side}} = R_{\text{side}}$
		Condition 2: $p > 1$, $t > 1$				
R10	252	0.75	80.17 (2.59)	81.69 (2.96)	.053	$L_{\text{side}} = R_{\text{side}}$
R22	233	0.82	82.33 (1.61)	83.03 (1.72)	.079	$L_{\text{side}} = R_{\text{side}}$

In sum, we find D-RI holding in 4/4 cases, 2/2 respondents.

3.4.3 Discussion

The generalized Prelec function (18) with $\mu = 1$ (for $p < 1$) and $\mu' = 1$ (for $p > 1$) reduces to the generalized power function (13). Although 3 respondents did not reject k-MP for $p < 1, q < 1$, R10 and R22 did. None of the 6 rejected it for $p > 1, q > 1$. Consistent with this, we tested D-RI, the generalization of $k-$MP, for R10 and R22 with $p < 1, q < 1$ and found it to hold, which is equivalent to the weighting function W being a generalized Prelec function (18) with $\mu \neq 1$. We also tested the property with $p > 1, q > 1$, where we expected it to hold consistent with the two respondent passing k-MP in that condition because it is the special case of the generalized Prelec function (18) with $\mu = 1$. Encouragingly, that expectation was met for both respondents.

Important issues of variance: In the current experiment, a particular statistical issue arises in a particularly onerous way. Principally, the problem lies in unaccounted for accumulated variance in compound judgments: The variance accumulated in Step 1, is "lost" when, in Step 2, a single number is used to estimate q, an estimate which is itself also represented by a single number. Consequently, in Step 3, we have an estimate of q which contains, in the end tree or four sources of bias (depending on how you count): a bias due to variance to which we add a rounding to the nearest 5% as we raise q to the power N, a bias also introduced into p^N and t^N. Hence, upon testing statistically whether $L_{\text{side}} = R_{\text{side}}$, substantial portion of the variance is "lost" to the statistical analysis, making the task of rejecting the equality hypothesis much easier one than it would be otherwise. Given this, the results here should be considered rather robust. We address this issue in more detail in Appendix G.

4 Summary and conclusions

- **Background:** Assuming $W(1) = 1$ and based on the failure of 1-MP (10), a power function form for W had been rejected. In this article, we have explored what happens if, instead, we abandon the assumption that $W(1) = 1$.
- **Form of the weighting function W and behavioral equivalents assuming $W(1) = 1$:** Narens' (1996) $1-$multiplicative property, (10), is shown to be equivalent to the power function form, (11). For a more general class of function, of which the power function, (11), is a special case, reduction invariance, (17), is shown to be equivalent to W being a Prelec function, (16).
- **Generalization assuming $W(1) \neq 1$:** The results are: the generalized power function form, (13), and an the associated behavioral equivalent, the $k-$multiplicative property, (14), and the generalized Prelec function, (18), and the corresponding behavioral property, double reduction invariance,

(20).

- **Experimental evaluation of generalized weighting function forms**: The k-MP property was evaluated in Experiment 2. The property was accepted for 6/6 respondents for $p > 1, q > 1$, and 3/5 for $p < 1, q < 1$. In Experiment 3 the two respondents rejecting k-MP for $p < 1, q < 1$ were both shown to pass D-RI. In Appendix E, we examined k-MP for the mixed condition $p < 1, 0 < 1$ for 3 respondents, and the data did not support $\omega \neq \omega'$ in (13), meaning that the respondents treated numbers above and below 1 differently.
- **Main conclusions:** For $p > 1, q > 1$, the generalized power function, (13), describes all respondents and 3 of 5 respondents for $p < 1, q < 1$. In those other cases, the generalized Prelec function, (18), is a good description.

4.1 Summary of this article in context of Steingrimsson and Luce, 2005a,b, 2006

Figure 2 displays graphically the relations among the current article and the preceding three which, together, have been aimed at testing Luce's (2002, 2004) theory of global psychophysics in the auditory domain.

The first article, Steingrimsson and Luce (2005a) established first that one could not assume the two ears were equal in behavior (5) and second that the p-additive representation (3) and the proportion representation (4), which we here simply refer to as the numerical distortion function $W(p)$, were supported separately.

Steingrimsson and Luce (2005b) tested properties that linked (3) and (4) in such a way that it was assured that the psychophysical function Ψ was the same one in both representations. This result left two functions whose exact form was unknown, namely, the psychophysical function Ψ and the weighting function W.

The third paper of the series, Steingrimsson and Luce (2006), established that for at least half of the respondents, Ψ was well described by a sum of power functions; the other half is still being thought about.

The current article explored forms for W. The theory developed substantially with the realization that the failures of 1-MP (10) and of RI (17) might not be due to a failure of the power function form (11) or the Prelec form (16) forms for W, per se, but rather might be due to a failure of the assumption $W(1) = 1$. This realization allowed us to generalize both the power form to (13) and the Prelec one to (18) and develop corresponding testable behavioral properties. The result is both clear and complete: most respondents' numerical distortions is well described by a general power function (13) and the few for whom that

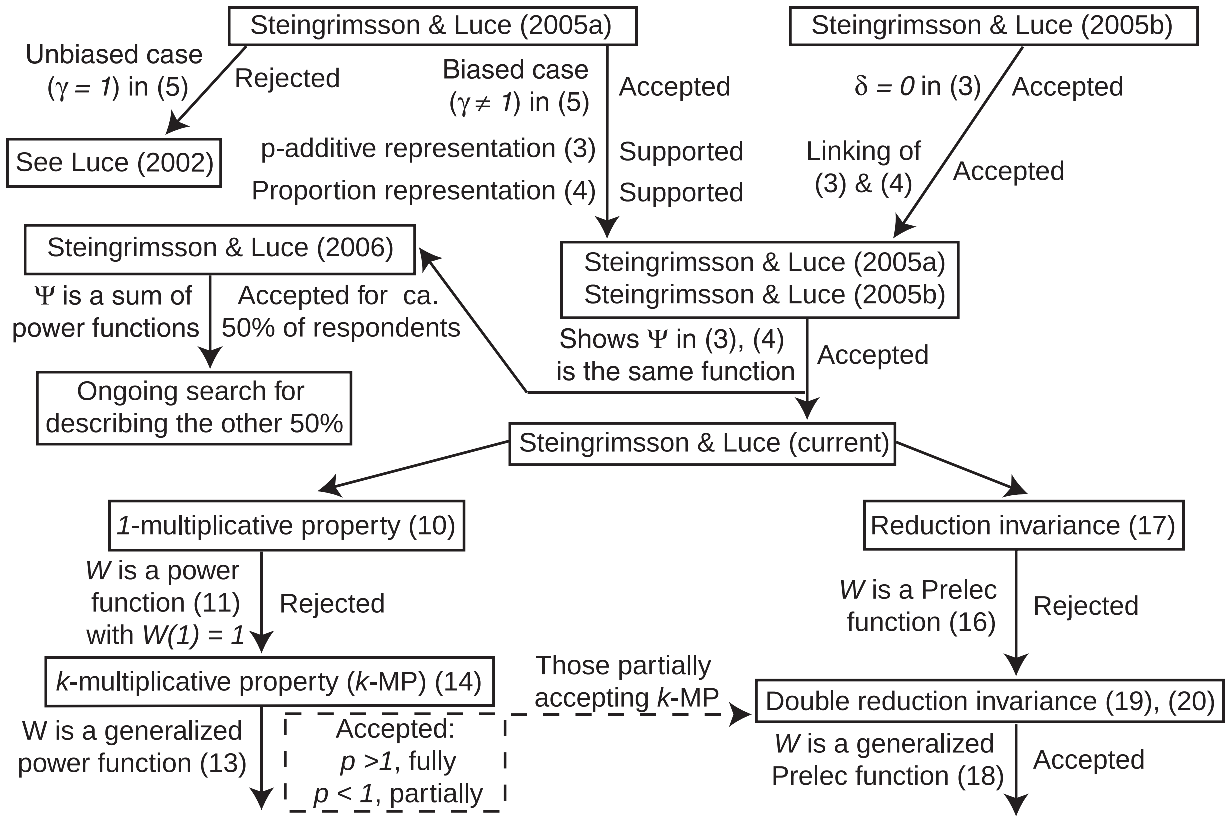

Fig. 2. The diagram shows the work done here in the context of the previous three papers in this, now, four paper series on the *Empirical Evaluation of a Model of Global Psychophysical Judgments*.

description is not adequate are well described by a generalized Prelec function, (18). Furthermore, these forms differ for judgments of numbers below and above 1.

Acknowledgments

This research was supported in part by National Science Foundation grant SBR-9808057 to the University of California, Irvine. Additional financial support was provided by the U.C. Irvine School of Social Sciences and by their Department of Cognitive Sciences. We thank the Center for Neural Science and the Department of Psychology at New York University, especially Dr. Malcolm Semple and Dr. Marisa Carrasco, for providing us with space and equipment for collecting much of the data reported here.

References

Aczél, J. (1966). *Lectures on Functional Equations and Their Applications*. Academic Press, New York/London.

Aczél, J., Luce, R. D. (submitted). *Remark: A Behavioral Condition for Prelec's Weighting Function without Restricting Its Value at* 1.

Attneave, F. (1962). Perception and related areas. In S. Koch (Ed.), *Psychology: A study of a science. Vol. 4*. New York: McGraw-Hill, 1962.

Baird, J. C. (1975). Psychophysical study of numbers IV. Generalized preferred state theory. *Psychological Research*, 38, 175-187.

Banks, W. P., Hill, D. K. (1974). The apparent magnitude of number scaled by random production. *Journal of experimental psychology monographs*. 102.

Banks, W.P. & Coleman, M.J. (1981). Two subjective scales of number. *Perception & Psychophysics*, 29, 95-105.

DeCarlo, L. T. (2003). An application of a dynamic model of judgment to magnitude production. *Perception & Psychophysics, 65*, 152–162.

DeCarlo, L. T. (in press). Sequential effects in successive ratio estimation. *Perception & Psychophysics*.

DeCarlo, L. T., Cross, D. V. (1990). Sequential effects in magnitude scaling: Models and theory. *Journal of Experimental Psychology: General, 119*, 375–396.

Dixon, W. J., Moon, A. M. (1948). A method for obtaining and analyzing sensitivity data. *Journal of the American Statistical Association, 43*, 109–126.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Chapman and Hall, New York.

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

Green, D. M., Luce, R. D., Duncan, J. E. (1977). Variability and sequential effects in magnitude production and estimation of auditory intensity. *Perception & Psychophysics, 22*, 450–456.

Ham, L. B., Biggs, F., & Cathey, E. H., Jr. (1962). Fractional and multiple judgments of loudness. *Journal of the Acoustical Society of America*, 34, 1118-1121.

Hellman, R.P., & Zwislocki, J. (1961). Some factors affecting the estimate of loudness. *Journal of the Acoustical Society of America,* 33, 687-694.

Hellström, Å. (1985). The time-order error and its relatives: Mirrors of cognitive processes in comparing. *Psychological Bulletin, 97*, 35–61.

Hellström, Å. (2003). Comparison is not just subtraction: Effects of time- and space-order on subjective stimulus difference *Perception & Psychophysics, 65*, 1161-1177.

Jesteadt, W., Luce, R. D., Green, D. M. (1977). Sequential effects in judgments of loudness. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 92–104.

Karpiuk, P., Jr., Lacouture, Y., & Marley, A. A. J. (1997). A limited capacity, wave equality, random walk model of absolute identification. In A. A. J. Marley (Ed.) *Choice, Decision, and Measurement: Essays in Honor of R. Duncan Luce. Mahwah, NJ: Lawrence Erlbaum Associates*. Pp. 279-299.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of Measurement, Vol. I.* New York, San Diego: Academic Press. (To be reissued by Dover Publications.)

Levitt, H., 1971. Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America, 49*, 467–477.

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

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

Luce, R. D. (2004). Increasing increment generalizations of rank-dependent theories. *Theory and Decision, 55*, 87–146.

Narens, L. (1996). A theory of ratio magnitude estimation. *Journal of Mathematical Psychology, 40*, 109–129.

Rule, S.J. & Curtis, D.W. (1973). Conjoint scaling of subjective number and weight. *Journal of General Psychology,* 97, 305-309.

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

Schneider, B., Parker, S., Ostrosky, D., Stein, D., Kanow, G. (1974). A scale for the psychological magnitude of number. *Perception & Psychophysics, 16*, 43-46.

Steingrimsson, R., & Luce, R. D. (2005a). Empirical evaluation of 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). Empirical evaluation of 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 intensity filtering. *Journal of Mathematical Psychology*, *50*, 15-29.

Steingrimsson, R., & Luce, R. D. (in preparation). The time-order error in the light of a model of global psychophysicaljudgments.

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

Wetherill, G. B. (1963). Sequential estimation of quantal response curves. *Journal of the Royal Statistical Society, B25*, 1–58.

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

Appendices

A Proof of proposition 1

We use the notational abbreviation of Section 3.1.2 and write $\psi(x) = \Psi(x, x)$. Because the latter is order preserving, applied to (7) and using (6) yields,

$$\begin{aligned} &(x \circ_p 0) \circ_q 0 \sim x \circ_t 0 \\ \Leftrightarrow\ &\psi((x \circ_p 0), \circ_q 0) = \psi(x \circ_t 0) \\ \Leftrightarrow\ &\psi(x)W(p)W(q) = \psi(x)W(t), \end{aligned}$$

which by canceling $\psi(x)$ is equivalent to (8).

B An argument for (13) via magnitude estimation

In his work using data averaged over participants, Stevens (1975) argued that both magnitude estimates[11] and productions are power functions although he demonstrated empirically they do not prove to be simple inverses of one another. Indeed, he spoke of there being an unexplained "regression" effect which has never really been adequately illuminated.

Steingrimsson and Luce (2006) showed that on the assumption that $\psi_l(x) = \Psi(x, 0)$ (the argument is equally valid for $\psi_r(x) = \Psi(0, x)$) is a power function, i.e.,

$$\psi_i(x) = \alpha_i x^{\beta_i} \quad (x \geq 0, i = l, r), \tag{B.1}$$

then the following inverse relationships hold between ratio productions and ratio estimates:

$$t_i(p) = W(p)^{1/\beta_i} \quad (p \text{ given, } i = l, r), \tag{B.2}$$

$$p_i(t) = W^{-1}(t^{\beta_i}) \quad (t \text{ given, } i = l, r). \tag{B.3}$$

Consider the possibility that, as Stevens (1975) claimed,

$$p_i(t) = \rho_i t^{\eta_i} \quad (t > 0, \rho_i > 0, \eta_i > 0). \tag{B.4}$$

Substituting (B.4) into (B.3) and rearranging yields (13) with $\omega = \beta_i/\eta_i$ and $W(1) = (1/\rho_i)^\omega$ and so by (15) we have $k = 1/\rho_i$. The latter two expressions mean that ρ must be independent of index.

[11] Ratio estimates with $y = 0$ and without a specified standard.

If we were to suppose that the model also holds for all $p > 0, q > 0$, thereby including $p > 1 > q$, then $\omega = \omega'$, independent of whether $p < 1$ or > 1. The literature (Hellman & Zwislocki, 1961; Ham, Biggs, & Cathey, 1962) and data in Appendix E suggest that this is not a good assumption, so we will experimentally examine (14) as asserted in Section 2.2.1.

Because of the way data are usually plotted in dB form, it is convenient to convert the predictions into dB form: $p_{i,\mathrm{dB}} := 10 \log p_i$, etc. For the power function form (13), and substituting that and its inverse into (B.2) and (B.3), respectively, and then if we write $w := W(1)$ and take logarithms, we have for $p < 1$

$$t_i(p)_{\mathrm{dB}} = \frac{1}{\beta_i} \left(w_{\mathrm{dB}} + \omega p_{\mathrm{dB}} \right), \tag{B.5}$$

$$p_i(t)_{\mathrm{dB}} = \frac{1}{\omega} \left(\beta_i t_{\mathrm{dB}} - w_{\mathrm{dB}} \right). \tag{B.6}$$

A similar expression holds for $p > 1$ with the constant ω'.

If production and estimation data are averaged over functions with a break at different standards for different people or with sequential effects of the sort described in Appendix D, it is no surprise that the results are not strict inverses of one another. This may provide an account of the regression effect. This is a possible answer to a question raised in Steingrimsson & Luce (2006, p. 19).

C Proof of proposition 2

Using the notation convention above, if we apply (6) to (7), then (14) yields the functional equation

$$W(p)W(q) = W(t) = W(kpq).$$

If we set $q = 1$, we obtain $W(kp) = W(1)W(p)$, whence

$$W(p)W(q) = W(1)W(pq).$$

This is the well known Pexider equation (Aczél, 1966) with solution (13).

Conversely, this form applied to (7) implies

$$W(1)^2 p^\omega q^\omega = W(1) t^\omega,$$

and so (14) is satisfied with

$$k = W(1)^{1/\omega}.$$

To show the Corollary, suppose first that $p < 1, q < 1$. With $kpq < 1$, the above argument proves the first part of (15). If however, $kpq > 1$, then we have

$$W(1)^2(pq)^\omega = W(p)W(q) = W(kpq) = W(1)k^{\omega'}(pq)^{\omega'}$$
$$\Leftrightarrow \frac{W(1)}{k^{\omega'}} = (pq)^{\omega'-\omega}.$$

Because the left term is constant, $\omega' = \omega$ and so by (15) $k' = k$. The other cases, $p > 1, q > 1$, and $(1-p)(1-q) < 0$ are similar.

D Sequential effects

In either magnitude estimation or magnitude production as usually conducted with no standard, the respondent presumably does one of two things. Either he or she sets a personal standard. This could be some sort of internalized standard at different locations for different people. These, when averaged over respondents, tend to produce a function that is less bowed than any of them alone. Or, when no standard is provided, the respondent could use the previous signal-response pair as the standard, in which case there will necessarily be sequential effects, perhaps of the sort reported by Green et al. (1977, Fig. 4, p. 455) for both magnitude estimates and productions when no standard was provided. We re-create their Fig. 4 as our Figure D.1.

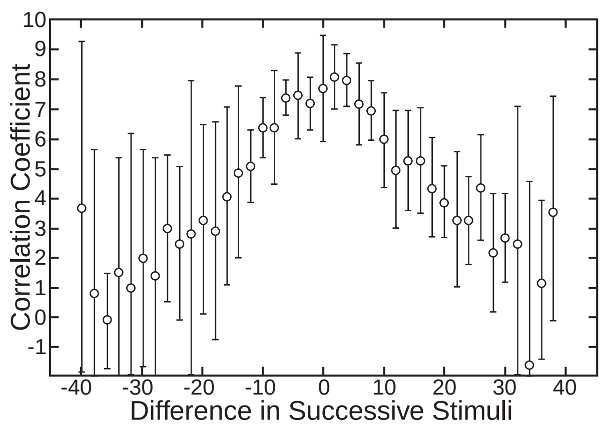

Fig. D.1. Correlation between successive responses (numbers) in magnitude estimation averaged over constant differences in the stimuli (dB) vs. these stimulus differences.

It is evident that the responses are very highly correlated when the successive signals are relative close and the correlation decays monotonically with signal separation until they are uncorrelated. This means that the data exhibited sequential effects. Various authors have formulated sequential models where the response on trial n in dB, $R_{n,\mathrm{dB}}$, depends linearly on the present signal in dB, $S_{n,\mathrm{dB}}$, the previous one, $S_{n-1,\mathrm{dB}}$, the previous response $R_{n-1,\mathrm{dB}}$, and in some cases $S_{n-2,\mathrm{dB}}$ (DeCarlo, 2003; DeCarlo and Cross, 1990; Jesteadt, Luce, and Green, 1977; Karpiuk, Lacouture, and Marley, 1997, and see references there).

D.1 Sequential effects due to the Prelec function

We explore how sequential effects may arise within our framework. The basic idea is to suppose that the respondent uses the immediately preceding signal/response pair as a departure point for assigning a response to the current signal. Thus, he or she makes the following identifications of magnitude estimation and production, (B.3) and (B.2),

$$t_n = \frac{S_n}{S_{n-1}},\ p_n = \frac{R_n}{R_{n-1}}, \tag{D.1}$$

Given that, we work out what happens with the Prelec function, (18), and by a simple specialization find the predicted sequential effects for power functions. To this end, we assume that a power function form holds for the psychophysical function, and so by the magnitude estimation form (B.3) and using the identification (D.1) we see that

$$\begin{aligned}
\left(\frac{S_n}{S_{n-1}}\right)^{\beta} &= W\left(\frac{R_n}{R_{n-1}}\right) \\
\Leftrightarrow \beta \ln\left(\frac{S_n}{S_{n-1}}\right) &= \ln\left(\frac{R_n}{R_{n-1}}\right) \\
&= \ln W(1) + \begin{cases} -\omega\left(-\ln\left(\frac{R_n}{R_{n-1}}\right)\right)^{\mu}, & S_n \le S_{n-1} \\ \\ \omega'\left(\ln\left(\frac{R_n}{R_{n-1}}\right)\right)^{\mu'}, & S_n > S_{n-1} \end{cases}.
\end{aligned}$$

If we define $\rho = 10/\ln 10 \approx 4.3429$ and $w_{\mathrm{dB}} = 10\log W(1)$, etc., then rearranging we have for sequential effects the prediction

$$R_{n,\mathrm{dB}} = R_{n-1,\mathrm{dB}} + \begin{cases} -\rho^{(1-1/\mu)}\left[\frac{w_{\mathrm{dB}}-\beta\left(S_{n,\mathrm{dB}}-S_{n-1,\mathrm{dB}}\right)}{\omega}\right]^{1/\mu}, & S_n \le S_{n-1} \\ \\ \rho^{(1-1/\mu')}\left[\frac{\beta\left(S_{n,\mathrm{dB}}-S_{n-1,\mathrm{dB}}\right)-w_{\mathrm{dB}}}{\omega'}\right]^{1/\mu'}, & S_n > S_{n-1} \end{cases}. \tag{D.2}$$

The following prediction for the power function follows simply by setting $\mu = \mu' = 1$, yielding

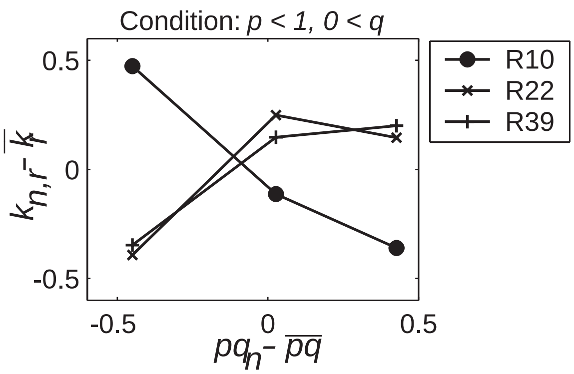

Fig. E.1. Experiment 2: Results for testing the k-multiplicative property in the mixed condition $p < 1, 0 < q$.

$$R_{n,\mathrm{dB}} = R_{n-1,\mathrm{dB}} + \left[\beta\left(S_{n,\mathrm{dB}} - S_{n-1,\mathrm{dB}}\right) - w_{\mathrm{dB}}\right] \begin{cases} \frac{1}{\omega}, & S_n \leq S_{n-1} \\ \frac{1}{\omega'} & S_n > S_{n-1} \end{cases} . \quad \text{(D.3)}$$

Models of the form (D.3) have been postulated and evaluated empirically without complete success (see references above). The more general form of (D.2) may be worth exploring empirically.

E Data suggesting $\omega \neq \omega'$

A long standing suspicion is that respondents deal with numbers above and below 1 differently, which, in the context of the generalized power form (13), is equivalent to $\omega \neq \omega'$ (see Ham, Biggs, & Cathey, 1962; Hellman & Zwislocki, 1961). Attempts to fit weighting functions and data we present here suggest that $\omega \neq \omega'$.

In Experiment 2 we collected data on k-MP (14) for $p < 1, q < 1$ and $p > 1, q > 1$, separately. For 3 respondents, we also collected equivalent data using $p < 1, 0 < q$, specifically:

Condition	x dB	p%	q_1%	q_2%	q_3%
$p < 1,\ 0 < q$	66	40	80	200	300

The results are presented in Figure E.1.

In Table E.1, we list $\overline{k_r}$, both for the data reported here as well as those for the same respondents from the non-mixed conditions (see Table 2).

Table E.1
Listing of obtained $\overline{k_r}$, for three conditions of p and q.

Condition	R10	R22	R39
1: $p < 1, q < 1$	1.03	1.23	1.03
2: $p > 1, q > 1$	0.91	0.82	0.72
3: $p < 1, 0 < q$	1.78	1.48	1.99

Visual inspection of Figure E.1 reveals a marked difference from comparatively straight lines formed by the non-mixed data of Figure 1. Likewise, looking at Table E.1 it is eye-catching how radically $\overline{k_r}$ in the mixed condition diverges from the other $\overline{k_r}$ estimates. Apart from R22, who in any case fails k-MP for $p < 1, q < 1$, and recognizing that $\overline{k_r} = 1.03$ is poor evidence for $\overline{k_r} > 1$, as well as all other estimates of $\overline{k_r}$ we have (including Table 2) are consistent with $\overline{k_r} < 1$. Knowing that $k < (>)1 \iff W(1) < (>)1$, then most of the data suggest $W(1) < 1$. Assuming that $W(1) < 1$, then from (15) it is quite clear that $\omega \neq \omega'$, a result consistent with non-constant functions of Figure E.1 and the very different $\overline{k_r}$ estimates obtained in the mixed condition compared to the other two.

F The Up-Down method for estimating proportions

The following details the use of the Up-Down method (Dixon & Moon, 1948; Wetherill, 1963; Levitt, 1971) to estimate a proportion p such that for some intensities v and w, $v \circ_p 0 = w$. With

- $\mathbf{P}_n$ a random variable representing requested proportion on trial n;
- p_0 the initial proportion;
- p a number representing the target proportion;
- $\mathbf{V}_n$ a random variable representing intensity estimate on trial n
- w a fixed intensity;
- δ a constant;
- $\mathbf{Y}_n$ a random variable taking the value $\begin{cases} 0 & \text{if } \mathbf{V}_n > w, \\ 1 & \text{if } \mathbf{V}_n \leq w, \end{cases}$

The Up-Down method assigns the initial and subsequent values of the requested proportion in the following way,

$$\mathbf{P}_{n+1} = \begin{cases} p_0 & \text{if } n = 0 \\ \mathbf{P}_n + \delta & \text{if } n > 0 \text{ and } \mathbf{Y}_n = 0 \\ \mathbf{P}_n - \delta & \text{if } n > 0 \text{ and } \mathbf{Y}_n = 1 \end{cases} \tag{F.1}$$

For sufficiently large n, an estimate $\widehat{p_{1/2}}$ of the median value for p, or $p_{1/2}$, can be obtained by averaging all the P_n's, but in practice excluding the $j < n$ first trials to avoid estimation bias caused by an ill-chosen initial value. That is, the statistic

$$\frac{1}{n-j}\sum_{i=j}^{n}\mathbf{P}_{j+i} \approx p_{\frac{1}{2}}.$$

G The problem of lost variance information

Variability in data is always problematic but in Experiment 3, it is particularly vexing. The problem lies in an accumulation of variance in compound judgements. For instance, in Step 1, the right side of D-RI, $(x \circ_t 0) \circ_t 0$, is reduced to one number by two successive ratio productions. To "preserve" the accumulating variance, we first obtain n estimates of $z_j = x \circ_t 0$, $j = 1...n$ and then use each individual estimate v_j as an input into the second one, to obtain $w_j = z_j \circ_t 0$ (see Appendix A.4 of Steingrimsson & Luce, 2005a, for a discussion this technique); on the right side, we obtain the estimates $v_j = x \circ_p 0$. When in Step 2 we obtain an estimate of q, we have to use the average of the v_j's and w_j's for the Up-Down process. An average always contains some error, which is now transferred into the estimate of q, which is itself taken as an average of the Up-Down run. Thus, we enter Step 3 using an estimate q which contains three sources of bias: a bias due to variance to which we add rounding to the nearest 5% following the raising of q to the power N—a bias introduced to p^N and t^N as well. As a result, the values that are actually subjected to a statistical test of equality ($L_{\text{side}} = R_{\text{side}}$) contain several levels of accumulated variance which are, in effect, "lost" for purposes of the statistical analysis and thereby making the task of rejecting equality a much easier one than it would be otherwise.

One possible solution to the problem is to estimate the magnitude of the "lost" variance and then add it into the final result, e.g., by way of a Monte Carlo simulation and explore its effect on the statistical test. Indeed if we let σ_z be the variance of $\overline{z}$ and σ_w the variance of $\overline{w}$, then, in principle, we have an indicator of both the total variance lost in compound ratio production as well as the increase in variance as the compounding occurs, i.e., the difference between σ_z and σ_w (recall, in Step 1, this variance is carried forward and is thus not "lost". However, several factors complicate a straight forward use of this information. We mention three:

(1) We know (e.g., Steingrimsson & Luce, 2005a, Appendix A) that variance tends to decrease with increases in intensity. This means that, in fact, σ_w is affected by both the inherited variance of σ_z as well as whether the ratio production is for a proportion larger or smaller than 1.
(2) We know that (Steingrimsson & Luce, 2005a, Appendix A) inter-session variability in ratio productions typically has the feature that, while the inter-relationship between estimated values tends to remain the same, the actual estimates can vary substantially. This means that it is very different to talk about inter-session variability and within-session variability. This means that any use of data pooled over multiple sessions has to be done quite carefully. Ideally, when dealing with ratio productions, all steps of estimation should be done within a single session.
(3) Even if we had a way to deal with the previous two items, we do not have a good estimate of how much the variability does impact the estimate of q and how that variability in q changes the final outcome—this requires independent investigation.

The upshot of this is that the final statistical test ($L_{\text{side}} = R_{\text{side}}$) in cases of this nature, e.g., D-RI, is unduly strict. However, while the problem is clearly both real and serious, it is not (to us at least) obvious what the appropriate remedy should be. Because we did not reject any conditions of D-RI, we are not compelled to seek a solution here. However, "crude" checks might be to identify cases in which equality of means is rejected but the differences between the means is less than σ_w and to regard those rejections as weak.