Are Psychophysical Scales of Intensities the Same or Different When Stimuli Vary on Other Dimensions? Theory with Experiments Varying Loudness and Pitch

R. Duncan Luce, Ragnar Steingrimsson, and Louis Narens
Institute for Mathematical Behavioral Science
University of California, Irvine

Ragnar Steingrimsson
Institute for Mathematical Behavioral Science
University of California, Irvine
Irvine, CA, 92697-5100
e-mail: ragnar@uci.edu

Abstract

Most studies concerning psychological measurement scales of intensive attributes have concluded that these scales are of ratio type and that the psychophysical function is closely approximated by a power function. Experiments show, for such cases, that a commutativity property must hold under either successive increases or successive decreases provided, e.g., all other independent dimensions are fixed. A good deal of data support this. However, little or no attention has been paid to whether or not such subjective intensity scales differ when an independent dimension such as frequency (pitch in audition; color in vision; etc.) is varied. Using a (simple and favorably tested) theoretical model for global psychophysics, we arrive at a necessary and sufficient cross-dimension, commutativity condition for a common intensity scale to exist. For example, our data show that the loudness of a tone at frequency f and one at frequency g can each be viewed as arising from a common property of loudness over intensity/frequency pairs. Comparing one version of cross-dimensional commutativity with the corresponding one-dimensional commutativity property discriminates between a general representation of the ratio scale property and a special case of it. Future work: Does the theory extend to other intensive continua (prothetic attributes)? If so, which ones? And does it extend to cross-modal matching?

S. S. Stevens (1946, 1951) produced a new theory of measurement that allowed for the direct measurement of subjective intensities using the, now, well-known methods of magnitude estimation and production. These methods have been, and continue to be, widely used by psychologists and other behavioral scientists. Yet, Stevens' theory and his methods of measurement by magnitude estimation and production are considered by the current standards of measurement theorists, both within and outside of psychology, and by philosophers of science and others interested in the theory of measurement to be non-rigorous and/or ill-founded.

Narens (1996) offered an axiomatic analysis of Stevens' methods using the representational theory of measurement and he discovered alternatives to Stevens' methods which he showed were subjectively measured rigorously on ratio scales. The principal difference between two forms of measurement is that Stevens assumed that the respondents give a veridical interpretation of the instructions to produce either ratio estimations or ratio productions, whereas Narens required them to be represented as some ratio, but not necessarily the exact ratio requested in the instructions. Later, other theories (Augustin, 2006; Luce, 2004, see erratum 2008; Narens, 1997, 2006) expanded Narens (1996) approach. Further, the experimental results of several researchers cited throughout this article have established the predicted deficiencies of Stevens' approach and supported the predicted strengths of both Luce's and Narens' theoretical approaches.

Luce's and Narens' approaches are examples of a theory of subjective intensity called the "relation theory." The relation theory (Krantz, 1972; Shepard, 1981) holds that the fundamental object judged by an observer is a pair of stimuli rather than a single stimulus. We apply the relation theory to pairs of stimuli from different dimensions in a manner that is different from others in the literature (e.g., Krantz, 1972).

Narens' models rest on the idea that ratio scalability is essentially equivalent to a simple commutativity property for signals varying only on single dimension of intensity. For loudness magnitude production, the commutativity property is formulated as follows: Let x be an arbitrary sound intensity less its threshold intensity and let p and q be two positive numbers. The respondent is asked to produce a sound intensity x_p that is heard as "p times as loud as x", and later to produce a sound intensity $x_{p,q}$ that is "q times as loud as x_p". Next suppose the respondent is asked to reverse the order so that q is first and p second, i.e., to produce a sound x_q that is "q times as loud as x" and then a sound $x_{q,p}$ that is "p times as loud as x_q". The commutativity property holds if and only if $x_{p,q} = x_{q,p}$. As we said, this prediction has been supported empirically in several studies (see below).

Luce's (2004) global psychophysical model is somewhat richer in concepts than Narens' models are and applies to a richer set of experimental situations. Also, it has been extensively and favorably evaluated experimentally in loudness and brightness (Steingrimsson & Luce, 2005a,b, 2006, 2007; Luce & Steingrimsson, 2008; Steingrimsson, 2009, submitted). Further, Narens' model, when viewed as a model of the psychophysical function, becomes a special case of Luce's where the reference points are set to 0. Moreover, Luce's and Narens' models both predict commutativity, but the former appears to be more straightforward to present and has somewhat greater explanatory powers. So we have chosen it to formulate the theory presented in this article and its account of the experimental data we report.

This research was supported in part by National Science Foundation grant BCS-0720288 to the University of California, Irvine—any opinion, finding, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation. We thank A. A. J. Marley for his detailed comments on an earlier version.

Narens (2006) formulated various ways to extend the commutativity property and used those extensions to conclude when it is possible to represent various ratio scales all on a common same ratio scale. The latter is a general measurement-theoretic result that also applies outside of psychophysical situations. To apply the underlying idea to magnitude production when the ratio scales are on disjoint domains, we needed to develop a new kind of paradigm. We did so using a theoretically motivated paradigm that can be applied to a wide variety of psychophysical situations, including ones involving mixed modalities. The experimental part of this article demonstrates effectiveness of the new paradigm by investigating whether frequency dependent ratio scales for auditory loudness can be represented on a common ratio scale.

The paradigm can be used equally well to address more general questions. For example, assuming that brightness turns out to have properties similar to the loudness ones discussed in this article, then can loudness and brightness each be considered to be special cases of a common ratio scale of subjective magnitude for both modalities? If so, to what other domains does cross-dimensional commutativity generalize? The most general case would be to have a common scale for all intensity magnitudes of each type classed as prothetic by Stevens (1975). This could develop into quite an extensive experimental exploration.

The article begins with a concise statement of the theory, describes the experimental design, and reports the results and their implications.

A Psychophysical Theory of Ratio Scales

The behavioral, axiomatic theory for magnitude production has led to the following representation

$$W(p) = \frac{\psi(x_p) - \psi(\rho_i)}{\psi(x) - \psi(\rho_i)}, \quad \rho_i, \ i = \begin{cases} + & \text{if } p - 1 \geq 0 \\ -, & \text{if } p - 1 < 0 \end{cases} \quad , \tag{1}$$

where ψ is a strictly increasing, psychophysical ratio scale over signal intensities, and W is a strictly increasing, dimensionless, cognitive distortion of numbers, and ρ_i are called reference signals, which are respondent generated. For our purposes, we will treat the ρ_i as respondent generated free parameters to be estimated, which we detail at the end of this section. The reason for assuming $\rho_+ \neq \rho_-$ is that the data below require it. It would be most desirable to have a theory for reference signals, but currently we do not have one.

The representation (1) was derived from axioms including certain structural ones such as solvability and others that are behavioral and experimentally testable. One formulation of these conditions is given as Corollary 1 of Theorem 1 of Luce (2004). These testable properties have been evaluated favorably for both loudness and brightness (study of perceived contrast is underway). For such a simple representation, almost certainly other axiomatic formulations will be discovered in the future.

For some purposes, it is convenient to rewrite (1) in a linear form:

$$\psi(x_p) = W(p)\psi(x) + [1 - W(p)]\,\psi(\rho_i). \tag{2}$$

Further, as verified empirically by Steingrimsson and Luce (2005a,b), Luce & Steingrimsson (erratum, 2008), and Steingrimsson (2009, submitted), when a respondent selects

x_p to be the signal that is perceived to be "p times that of x" relative to its reference intensity ρ (magnitude production), the data are nicely described by this representation. Having an estimate of x_p, then it can play the role of x under the instruction to find $x_{p,q}$ that is q times x_p. Such data permit one to examine the commutativity property.

Steingrimsson and Luce (2006) provided data supporting the idea that ψ may be a power function, i.e.,

$$\psi(x) = \alpha x^{\beta}, \quad \alpha > 0, \beta > 0. \tag{3}$$

We very much need to devise a means for estimating β for each respondent. We do not need to estimate α because it cancels in (1).

Steingrimsson and Luce (2007) explored the form for W and supported the representation of Prelec (1998) but using a simpler property due to Luce (2001). We do not report this result in detail because it plays no role in this article.

Two propositions for single frequency case

Because we are using two numbers p and q, we need to be careful to distinguish four cases: both $p \geq 1$ and $q \geq 1$, both $p' < 1$ and $q' < 1$ and the two mixed cases $p \geq 1$ and $q' < 1$ and $p' < 1$ and $q \geq 1$. Note the we prime any number less than 1. Sometimes (see Case 3 of Proposition 1 we use $+$ and $-$ to refer to the sign of $p - 1$.

Our first two Propositions concern the single intensity dimension as shown as Case 1 in Figure 1

Proposition 1 *Suppose that the generic representation (1) holds. Then in the one dimensional case:*

1. *For $p \geq 1, q \geq 1$, commutativity $x_{p,q} = x_{q,p}$ holds.*
2. *For $p' < 1, q' < 1$, commutativity $x_{p',q'} = x_{q',p'}$ holds.*
3. *For $p' < 1 \leq q$ and for $p \geq 1 > q'$, commutativity holds iff $\rho_+ = \rho_-$.*

Proofs are given in Appendix A.

Cases 1 and 2 have been extensively studied and supported (Ellermeier & Faulhammer, 2000; Steingrimsson & Luce, 2005a; Zimmer, 2005; Augustin & Maier, 2008; Steingrimsson, 2009). So we take Cases 1 and 2 to be confirmed.

Case 3 has not been systematically studied empirically. Nothing that we know of precludes that $\rho_+ = \rho_-$, but equally well nothing a priori forces it. There has been very little study of the mixed case except for a limited amount of data reported in Steingrimsson and Luce (2007), Appendix E, that imply that $\rho_+ \neq \rho_-$.

The next proposition provides a means for estimating ρ_+ and ρ_-.

$$\tau_+(p, q) := W(p)W(q) \text{ if } p > 1, q > 1, \tag{4}$$
$$\tau_-(p', q') := W(p')W(q') \text{ if } p' < 1, q' < 1. \tag{5}$$

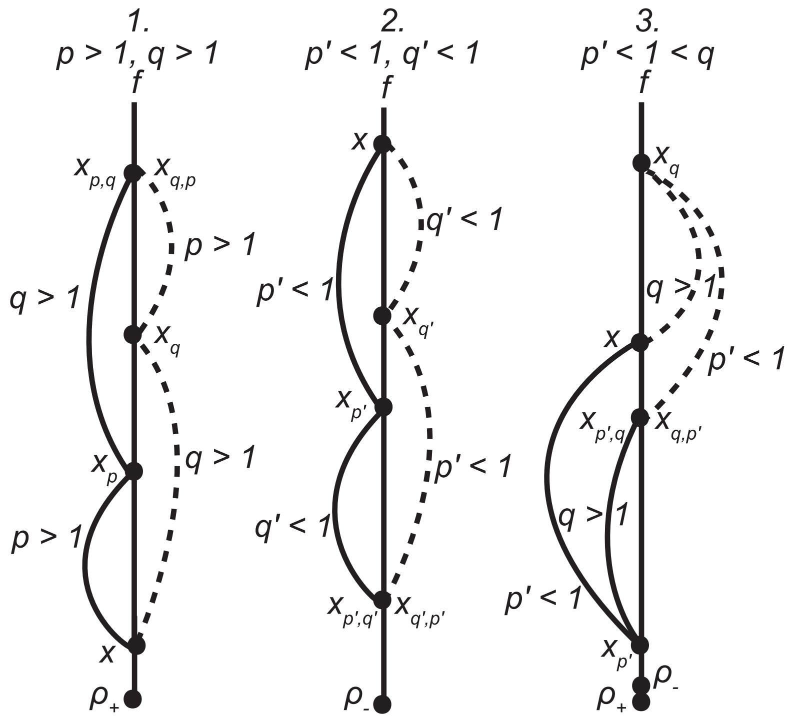

Figure 1. Depicted are the three cases of Proposition 1 In each part, the solid lines depict the case of magnitude production p followed by q and the dotted ones depict the case of magnitude production q followed by p. Commutativity is said to hold in Case 1 if $x_{p,q} \sim x_{q,p}$, in Case 2 if $x_{p',q'} \sim x_{q',p'}$, and in Case 3 if $x_{p',q} \sim x_{q,p'}$. Note that the reference point is ρ_+ if $p, q > 1$ and is ρ_- if $p', q' < 1$. And for Case 3, both reference point occur and commutativity is predicts only if $\rho_+ = \rho_-$.

Proposition 2 *Under the assumption of Proposition 1 and assuming (3), the parameters of Cases 1 and 2 must satisfy for all intensities x and empirically estimated $x_{p,q}$*

$$\rho_+^\beta = \frac{x^\beta \tau_+(p,q) - x_{p,q}^\beta}{\tau_+(p,q) - 1}, \tag{6}$$

$$\rho_-^\beta = \frac{x^\beta \tau_-(p',q') - x_{p',q'}^\beta}{\tau_-(p',q') - 1}. \tag{7}$$

With p, q (and p', q') held fixed, the data as we vary x should satisfy the linear expression (2), and so we may estimate ρ_i and τ_i separately for $i = +, -$. The "separateness" rests upon the empirical fact that respondents do behave differently.

Cross-frequency commutativity

General notation. The above arguments are readily modified to two frequency dimensions, say f and g, as in Figure 2, by suitable super- and sub-scribing. For example, in the single frequency case we wrote x_p whereas had we been explicit that we were mapping f

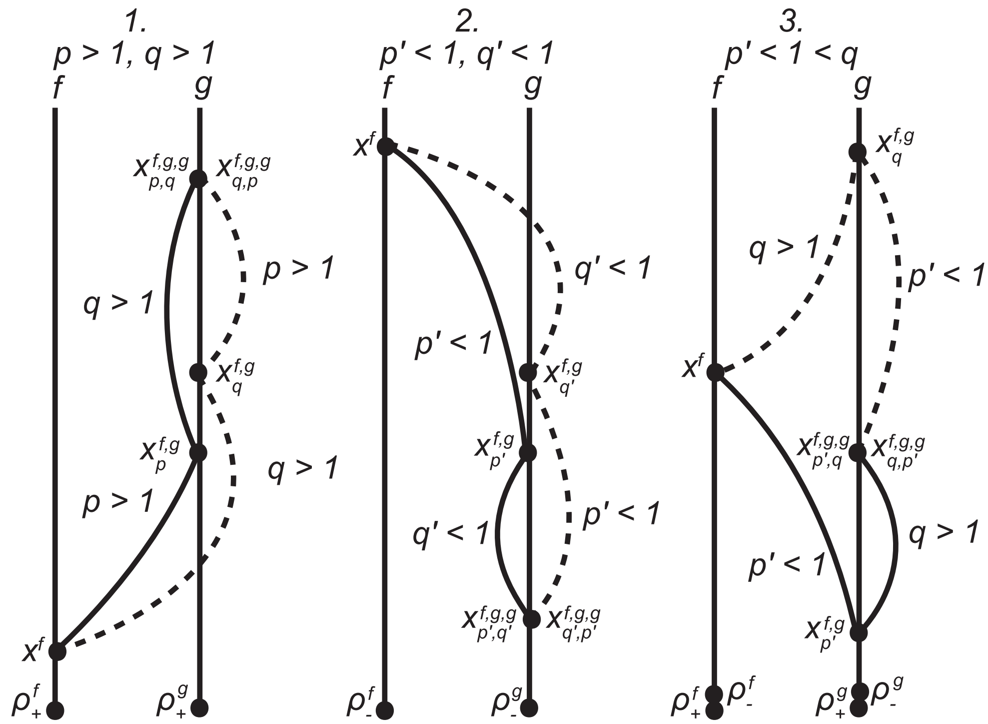

Figure 2. Depicted are the three cases of Proposition 1 when extended to two frequencies, f, g. In each of the three cases, the solid lines depict the case of magnitude production with p followed by q, "q, and the dotted lines depict magnitude production starting with q followed by p. Commutativity is found to hold if in 1. $x^{f,g,g}_{p,q} \sim x^{f,g,g}_{q,p}$, 2. $x^{f,g,g}_{p',q'} \sim x^{f,g,g}_{q',,p'}$, 3. $x^{f,g,g}_{p',q} \sim x^{f,g,g}_{q,p'}$ for each of the three cases, respectively. Note the reference point for $p, q > 1$ is ρ_+ and for $p', q' < 1$ it is ρ_-. In Case 3, both reference points appear and commutativity is predicted to hold only if $\rho_+ \sim \rho_-$.

to f, we should have written $x^{f,f}_p$. In the case where the judgment is from frequency f to frequency g, that must be explicit as $x^{f,g}_p$.[1] Then, as long as the roles of p and q are simply switched, the proposition continues to be satisfied.

The notation x^f simply indicates that x has the frequency f, not a power of x. The same is true of $x^{f,g}_p$. Because we are restricting ourselves to the two frequencies f and g, this convention appears the least cumbersome of those we have considered.[2]

More specifically, assume that loudness at each frequency is correctly described by Luce's (2004, 2008) theory. Let the resulting numerical ratio scales of loudness at f and g be denoted, respectively, by ψ^f and ψ^g (again, f and g refer to frequencies, not powers). Recall that each scale is assumed to have a reference point, ρ^f and ρ^g, from which the judgments are made.

We assume that the claim that ψ^f and ψ^g are "the same ratio scale" means that there

[1]Note that the x_p notation denotes the signal produced by the respondent. In earlier work, beginning with Narens (1996) we also used an operator notation such as $x \circ_p \rho$, which really is theoretical, not empirical.

[2]We use the simple f, g notation. However, were we to consider a property with $n > 2$ different dimensions, we would almost certainly switch to writing f_j, $j = 1, .., n$. For $n = 2$, we avoid the additional subscripts.

is a single ratio scale $\psi^{f,g}$ with arguments (x, h) for intensity x at frequency h that agrees with ψ^f when $h = f$ and with ψ^g when $h = g$.

Suppose a signal intensity x at frequency f is presented and the respondent magnitude produces an intensity at frequency g that is "p times x". We denote that production by $x_p^{f,g}$. If the next step is to produce on g the signal that is "q times $x_p^{f,g}$," the resulting signal is denoted $x_{p,q}^{f,g,g}$. Whereas, if the second magnitude production were to be at frequency f, the notation would change to $x_{p,q}^{f,g,f}$. When the roles of p and q are reversed, so they are in this notation.

We assume that the reference point "chosen" depends both upon whether or not $p \geq 1$ or $p' < 1$ and on whether it concerns the f or g frequency.

In the following, we write $f \to g$ to mean we are considering a cross-frequency production from frequency f to frequency g.

The cases $f \to g \to g$ and $f \to g \to f$. Three distinct cross-frequency cases, called 1, 2, and 3, are diagrammed in Figure 2. Note that in these cases, there is one cross dimension magnitude production followed by single dimension magnitude production. Figure 2 presents the realizations that we found to be most effective (it parallels Figure 1 for testing of single frequency); the first item we study begins with a judgment in frequency g of the given intensity at frequency f, namely $x_p^{f,g}$.

Proposition 3 *Under the assumption of Proposition 1, the cross mapping $f \to g \to g$ and $f \to g \to f$ both satisfy cross-dimensional commutativity for Cases 1 and 2 of Proposition 1 and hold for Case 3 if and only if $\rho_+^f = \rho_-^f$ and $\rho_+^g = \rho_-^g$.*

There is an additional case to explore.

Comparison of $f \to f \to f$ to $f \to g \to f$. We have shown commutativity on dimension f

$$x_{p,q} = x_{q,p},$$

and commutativity in the two dimensional case $f \to g \to f$,

$$x_{p,q}^{f,g,f} = x_{q,p}^{f,g,f}.$$

A natural question to ask is: Do these two cases agree?

Until now it has been sufficient to use simple superscripted reference points such as ρ^f and ρ^g. However, in crossing dimensions in the order f, g, f, we probably should not assume that ρ^g is the same when comparing going from f to g with going from g to f, and as we shall see the data agree with the possibility. We distinguish these as, respectively, $\rho^{f,g}$ and $\rho^{g,f}$.

Proposition 4 *Under the assumption of Proposition 1,*

$$x_{p,q}^{f,f,f} = x_{p,q}^{f,g,f},$$

if and only if

$$\rho^{f,g} = \rho^{g,f}.$$

Note that the Narens (1996) theory has no reference points and so it predicts that this agreement must hold. As we will see, the data do not support that prediction. The above argument is depicted in Figure 3.

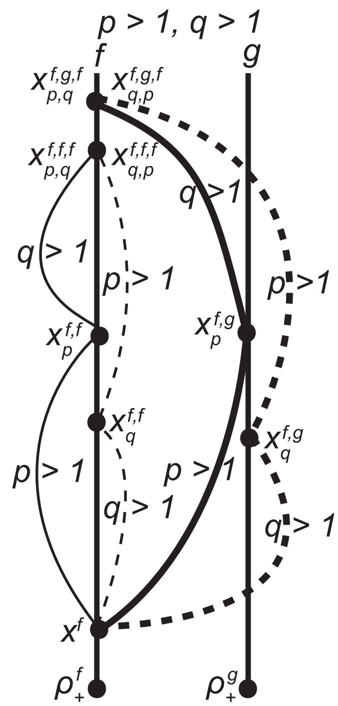

Figure 3. Depicted is the condition of Proposition 3. The narrow lines (solid and dotted) are on a single frequency and the thicker lines (solid and dotted) go across the two frequencies, f, g. Independent of thickness, the solid lines depict the case of magnitude production with p followed by q whereas the dotted line depict magnitude production starting with q followed by p. Commutativity in the single frequency case holds if $x_{p,q} \sim x_{q,p}$, for the cross frequency case if $x^{f,g,f}_{p,q} \sim x^{f,g,f}_{q,p}$, and $\rho^{f,g} = \rho^{g,f}$ holds only if $x_{p,q} \sim x_{q,p} \sim x^{f,g,f}_{p,q} \sim x^{f,g,f}_{q,p}$.

Method

The goal of the experiment is to test the same scale hypothesis of Section "Cross-frequency commutativity" as well as whether two reference points on different frequencies agree or not (Section "Comparison of $f \to f \to f$ to $f \to g \to f$ ").

Respondents

A total of 12 respondents—mostly graduate and undergraduates from the University of California, Irvine—participated in the experiment reported. All respondents reported normal hearing. Although desirable, for logistical reasons, not every respondent participated in every condition. All respondents, except the participating coauthor (R22), received compensation of $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 UC Irvine's Institutional Irvine's Institutional Review Board.

Stimuli

The basic signal was a sinusoidal tone presented for 100 ms, which included 10 ms on and off ramps. Because the theory is cast in terms of intensities less the threshold intensity, then for a left ear with a threshold of x_τ and a right ear one of u_τ, the effective stimulus (x, u) consists of $x = x' - x_\tau$ and $u = u' - u_\tau$ where (x', u') are the actual intensities presented. However, because all signals were well above threshold and the respondents were selected for normal hearing, the error in reporting intensities (x', u') in dB SPL (henceforth dB) is negligible. The basic stimulus consisted of two signals (tones), (x, x), i.e. the same tone in both ears, and (z, z), separated by 450 ms—see the Procedure Section for additional details.

Apparatus

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). Intensity and frequency was controlled through a programmable interface for the RP2.1 and stimuli were presented over Sennheiser HD265L headphones to the listener seated in an individual, single-walled IAC sound booth located in a quiet lab-room. A safety ceiling of 90 dB was imposed in all experiments.

Statistical method and presentation of results

We have no a priori model of how the data from individuals relate to one another, therefore all data analysis is done on individual data (e.g., Luce, 1995, p. 20).

We seek to evaluate parameter-free null hypotheses that have the generic form $L_{\text{side}} = R_{\text{side}}$. As a matter of logic, a null hypothesis can never be proven empirically, but it can be empirically supported. At present, we have no model of error for respondent produced estimates. This precludes any principled use of parametric testing. Further, as far as we are aware, but without intending any comment on the choices made by other authors, there is presently no clear or agreed upon approach to the issue in the psychological testing of representational models. The problem is common in physics and the solution in that field takes the form of providing a criterion by which data can be said to support (or not) a tested null hypothesis. To this end we developed a three pronged criterion.

• Statistical Method: Lack of parametric information, leads to the use of nonparametric test. Our choice is the Mann-Whitney U at a significance level of .05. This was the choice of others in similar studies (e.g., Ellermeier & Faulhammer, 2000; Zimmer, Luce, & Ellermeier, 2001; Ellermeier, Narens, & Dielmann, 2003; Zimmer, 2005; Steingrimsson & Luce 2005a, 2005b, 2006, 2007; Augustin & Maier, 2008; Steingrimsson, 2009, submitted). Because respondents adjust intensities in discrete steps, which estimates appear reasonably Gaussian, then medians are known to be well estimated by the mean, and thus central tendencies are reported as means with variability in standard deviations—the test itself, being a rank-order one, is on the medians. Put more formally, the tested version of (1) is

$$W(p) = \frac{\psi(\overline{\mathbf{x}}_p) - \psi(\rho_i)}{\psi(x) - \psi(\rho_i)}, \tag{8}$$

where $\overline{\mathbf{x}}_p$ is the median produced value.

• Sample size adequacy: To evaluate whether the sample size is sufficiently large to detect a true failure of the null hypothesis, all statistical results were verified using Monte Carlo simulations (see Steingrimsson & Luce, 2005a, and Steingrimsson, 2009, for details).

• Effect size: Currently, no accepted method for calculating effect size exists for the Mann-Whitney U. Instead, we use the simple observation that should two medians (means) differ by less than Weber's fraction, they are arguably not noticeably different to an observer.[3] Teghtsoonian (1971) reports that Weber's fractions for loudness from three, reportedly conservative and independent, studies all agreed on approximately .05.

Our criterion saying that a result supports our hypothesis is for all three indicators to agree; otherwise it is not supported.

Procedure

Experiments were conducted in sessions that lasted at most one hour each. The initial session was devoted to obtaining written consent, explaining the task, and running practice trials. All respondents trained for one additional session. Rest periods were encouraged but both their frequency and duration were under the respondent's control.

With reference to Figure 2, Panel 1, testing the same scale hypothesis consists of presenting a standard x^f and, following the solid line, obtaining first an estimate of $x_p^{f,g}$, then using $x_p^{f,g}$ as standard, obtaining an estimate of $x_{p,q}^{f,g,g}$; following the dotted line, the corresponding estimates are $x_q^{f,g}$ then $x_{q,p}^{f,g,g}$. The statistical hypothesis is $x_{p,q}^{f,g,g} \approx x_{q,p}^{f,g,g}$.

We use a two-ear presentation thus the standard is (x^f, x^f), which means the joint presentation of the tone x^f in the left ear and tone x^f in the right ear. Let $\langle (x^f, x^f), (z^g, z^g) \rangle$ mean that the joint presentation (x^f, x^f) is played, followed by 450 ms by the joint presentation (z^g, z^g)—a tone is a sinusoids of 100 ms duration including 10 ms on and off ramps (see Stimulus section). The intensity of the tone z^g is under the respondent's control, such that s/he can either increase or decrease its intensity. The intensity changes are in step sizes of .5, 1, 2, or 4 dB, each tied to the keys "a", "s", "d", and "f" for increasing intensity ";", "l", "k", and "j" for decreasing intensity. After each key-press, the altered tone sequence was replayed. Additionally, respondents could replay the tone sequence unaltered by pressing the "r" key.

To produce the estimate for $x_p^{f,g}$, the respondent's adjusted the intensity of z^g until the stimulus (z^g, z^g) sounded some prescribed proportion p of the standard stimulus (x^f, x^f). The respondent could adjust z^g as often as desired, but when satisfied with the production, s/he would press the "b" key. The final value of z^g was taken as an estimate for $x_p^{f,g}$. In a following step, this estimate was used as a standard $(x_p^{f,g}, x_p^{f,g})$ using the same process to obtain an estimate of $x_{p,q}^{f,g,g}$. The respondent produced estimates for $x_q^{f,g}$ and $x_{q,p}^{f,g,g}$ in an analogous fashion. Thus, to arrive at an estimate for $x_{p,q}^{f,g,g}$ and $x_{q,p}^{f,g,g}$ required a total of four individual estimates.

In most cases, two instantiation of the standard, (x^f, x^f) and (t^f, t^f), were used and randomly interleaved within a block of trials that thus consisted of eight estimates (four for each standard) to arriving at two test of the statistical hypothesis, namely, $x_{p,q}^{f,g,g} \approx x_{q,p}^{f,g,g}$ and $t_{p,q}^{f,g,g} \approx t_{q,p}^{f,g,g}$, respectively.

[3]We thank J. Yellott for this simple and elegant observation.

Two avoid inter-session variability, the eight estimates were collected within a block of trials (See Appendices *A.1* & *A.4* in Steingrimsson and Luce, 2005a, for details). Respondents typically completed between eight and 10 blocks per session. The value of the current proportion was displayed on monitor as well as the task instructions along with the current block number and trial number. The typical number of estimates for each of $x_{p,q}^{f,g,g}$and $x_{q,p}^{f,g,g}$ ($t_{p,q}^{f,g,g}$and $t_{q,p}^{f,g,g}$) was around 30 (excluding training), requiring 3–4 sessions to collect data for each condition.

Design and condition

The testing strategy aimed at evaluating the same scale hypothesis in three ways

1. Test the three p, q relationships of Proposition 3.

Case 1: $p \geq 1, q \geq 1$ (Figure 2, panel 1).

Case 2: $p' < 1, q' < 1$ (Figure 2, panel 2).

Case 3: $p \geq 1 > q'$ (Figure 2, panel 3).

2. Test of Proposition 4, whether $\rho^{f,g} = \rho^{g,f}$, i.e. is the reference point the same on any two frequencies. (Figure 3).

3. Test the conditions for several standards (frequencies), and frequencies, including a large frequency range.

Listed in Table 1, are the 13 stimulus conditions that together address these 3 aims.

Aim 3 is addressed in a various way: magnitudes productions are made both from higher to lower frequencies (conditions 1, 2, 3, 4, 6, 7, 8, 19, 12, 13), and lower to higher frequencies (conditions 3, 4, 9, 11) and over a wide range of frequencies (conditions 6, 7). For all except conditions 12 and 13, two standards are used for each condition, resulting in two tests for each condition.

Results

The results are summarized in Table 2. The detailed results are listed in Appendix B, in Tables B1 and B2. The condensed presentation here avoids burdening the reader with a list of 68 different tests, where as Tables B1 and B2 ensures full reporting of the appropriate details.

The pattern of results in Table 2 is clear: The same scale hypothesis of Proposition 3, is well supported for Cases 1 and 2, whereas Case 3 is unambiguously rejected (corresponding to conditions 1, 2, and 3 of Figure 2). The failure of Case 3 is evidence against the hypothesis that $\rho_+ = \rho_-$ for both frequencies. For the test of Proposition 4 (corresponding to Figure 3) the data do not support the general hypothesis that $\rho^{f,g} = \rho^{g,f}$.

From Tables B1 and B2, Appendix B, it can be seen that the two of the failures in test of Case 1 (Table 2) are for the same respondent in the same condition (failure for each of the two standards). Otherwise, there does not seem to be any pattern to the failures. Nine respondents attempted to exceed the 90 dB safety limit. For three, the standard was lowered and the problem resolved itself. Three participated only in test of Case 2 in which the proportions were less than 1 and thus the situation could not occur; the other respondents were unavailable to provide additional data. One respondent appeared not to have understood the instructions and showed dramatic inter-session variability ($\sim$10 dB) in three sessions. Those data are not reported.

Condition	Proportions		Frequencies (Hz)		Standards (dB)	
	p	q	f	g	x^f	t^f
Proposition 3, Case 1: $p > 1, q > 1$						
1‡	2	3	1000	2000	54	50
2	2	3	1000	2000	54	50
3	2	3	1000	2000 (x)/500(u)	54	50
4	150%	200%	1000	2000	54	50
5	150%	200%	1000	500	54	50
6	150%	200%	200	5000	54	50
Proposition 3, Case 2: $p' < 1, q' < 1$						
7	75%	50%	200	5000	82	72
8	75%	50%	1000	2000	82	76
9	75%	50%	1000	500	82	72
Proposition 3, Case 3: $p' < 1, q > 1$						
10	75%	150%	1000	2000	70 (60)†	74 (64)†
11	75%	150%	1000	500	70 (60)†	74 (64)†
Proposition 4: Equality of reference points $\rho^{f,g} = \rho^{g,f}$						
12	150%	200%	1000	2000	50	None
13	50%	75%	1000	2000	78	None

‡Stimuli presented in left ear only. †Standard differed for one subject.

Table 1: In the table are listed the 13 conditions under which the the commutativity condition was tested. First listed is an numerical identifer, followed by the values of the proportions, followed by the frequencies, and finally, the intensity of the two standards.

Test of	#Tests	#Hold	#Fail	%Hold	Hypothesis
Prop. 3, Case 1: $p > 1, q > 1$	38*	34	4	89%	Supported
Prop. 3, Case 2: $p' < 1, q' < 1$	12*	11	1	91%	Supported
Prop. 3, Case 3: $p' < 1 < q$	8	0	8	0%	Rejected
Prop. 4: Equality of reference points $\rho^{f,g} \approx \rho^{g,f}$	8	1	7	14%	Rejected

*Cases 1 and 2 include data from C_{12} and C_{13}, which in the course of evaluate $\rho^{f,g} \sim \rho^{g,f}$ provide additional tests of the two cases.

Table 2: The table summarizes the testing of the commutativity hypotheses of Propositions 3 and 4. First listed is the specific test conducted, followed by the number of tests of each, how many of those tests were found to hold, how many to fail, and the percentage of tests that held of the total number of tests. Finally, the conclusion about each hypothesis is listed.

Discussion and Conclusions

1. We began by showing how the ratio scale, commutativity property of prothetic attributes for signals that vary only in intensity can be extended to signals that vary both in intensity and in another variable such as frequency. Those results were the basis of the experimental program.

2. Empirical evidence supports the notion that for $p, q \geq 1$ and $p', q' < 1$, individuals rely on a single scale for loudness regardless of stimulus frequency.

3. The evidence is consistent with the idea that reference points for $p' < 1$ and for $p \geq 1$ do differ.

4. The evidence is consistent with reference points for different frequencies to differ. Specifically $\rho^f \neq \rho^g$ as well as $\rho^{f,g} \neq \rho^{g,f}$.

5. We asked: Is loudness an intensity scale that is independent of frequency? The data suggested the answer is yes as long as reference points are not assumed to remain fixed independent of procedure.

Several directions need to be explored. One that has so far escaped us is a principled theory of reference points. They are clearly important in our data and we have only been able to treat them as parameters to be estimated. A second question is: Do the same results hold for other prothetic continua, among them brightness of monochromatic colors? If so, expand the exploration to other prothetic continua that have been shown to be systematic using Stevens' magnitude methods. For those to which it does apply, the next step is to extend the model to inter-modal situations. Should that work out, the implication is that there is a single notion of subjective intensity for a person—a somewhat sweeping idea.

References

American Psychological Association (2002). Ethical principles of psychologists and code of conduct. *American Psychologist*, **57**, 1060–1073.

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

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.

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

Ellermeier, W., Narens, L., & Dielmann, B. (2003). Perceptual ratios, differences, and the underlying scale. In B. Berglund and E. Borg (Eds.) *Fechner Day 2003. Proceedings of the 19th annual meeting of the International Society for Psychophysics.* Stockholm, Sweden: International Society for Psychophysics. Pp. 71–76.

Krantz, D. H. (1972). A theory of magnitude estimation and crossmodality matching. *Journal of Mathematical Psychology*, **9**, 168–199.

Luce, R. D. (1995). Four tensions concerning mathematical modeling in psychology. *Annual Reviews of Psychology*, **46**, 1–26.

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

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

Luce, R. D. (2008). Correction to Luce (2004). *Psychological Review*, **115**, 601.

Luce, R. D., & Steingrimsson, R. (2008) Note on a changed empirical inference in several Steingrimsson and Luce articles due to C. T. Ng's correction of an error in Luce (2004). *Journal of Mathematical Psychology*, **52**, 263–264.

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

Narens, L. (1997). On subjective intensity and its measurement. In A. A. J. Marley (Ed.), *Choice, Decision, and Measurement: Essays in Honor of R. Duncan Luce.* Mahwah NJ: Lawrence Erlbaum and Associates. Pp. 189–206.

Narens, L. (2006). Symmetry, Direct Measurement, and Torgerson's Conjecture. *Journal of Mathematical Psychology*, **50**, 290–301.

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

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.

Shepard, R. N. (1981). Psychological relations and psychophysical scales: On the status of "direct" psychophysical measurement. *Journal of Mathematical Psychology*, **24**, 21–57.

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

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

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

Steingrimsson, R. (2009). Evaluating a model of global psychophysical judgments for brightness: I. Behavioral properties of summations and productions. *Attention, Perception, & Psychophysics*, **71**, 1916–1930.

Steingrimsson, R. (2010). *Evaluating a model of global psychophysical judgments for brightness: II. Behavioral Properties Linking Summations and Productions.* Manuscript submitted for publication.

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

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

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

Teghtsoonian, R. (1971) On the exponents in Stevens' law and the constant in Ekman's law. *Psychological Review*, **78**, 71–80

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

Zimmer, K., Luce, R. D., & Ellermeier, W. (2001). Testing a new theory of psychophysical scaling: Temporal loudness integration. Fechner Day 2001. In G. Sommerfeld, R. Kompass, & T. Lachmann (Eds.), *Proceedings of the 17th Annual Meeting of the International Society for Psychophysics.* Lengerirch, Germany: Pabst. Pp. 683–688.

Appendix A
Proofs

Proposition 1.

Case 1. This is the case $x \to x_p$ followed by $x_p \to x_{p,q}$:

$$W(p) = \frac{\psi(x_p) - \psi(\rho_+)}{\psi(x) - \psi(\rho_+)},$$
$$W(q) = \frac{\psi(x_{p,q}) - \psi(\rho_+)}{\psi(x_p) - \psi(\rho_+)}.$$

So,

$$W(p)W(q) = \left(\frac{\psi(x_p) - \psi(\rho_+)}{\psi(x) - \psi(\rho_+)}\right)\left(\frac{\psi(x_{p,q}) - \psi(\rho_+)}{\psi(x_p) - \psi(\rho_+)}\right) = \frac{\psi(x_{p,q}) - \psi(\rho_+)}{\psi(x) - \psi(\rho_+)}.$$

Similarly, for the order q, p,

$$W(q)W(p) = \left(\frac{\psi(x_q) - \psi(\rho_+)}{\psi(x) - \psi(\rho_+)}\right)\left(\frac{\psi(x_{q,p}) - \psi(\rho_+)}{\psi(x_p) - \psi(\rho_+)}\right) = \frac{\psi(x_{q,p}) - \psi(\rho_+)}{\psi(x) - \psi(\rho_+)}.$$

By the commutativity of multiplication, equate these and it is immediate that

$$x_{p,q} = x_{q,p},$$

i.e., commutativity holds.

Case 2. Here we have $p' < 1, q' < 1$, and a completely analogous argument using ρ_- again forces commutativity.

Case 3. Now assume we are working with $p > 1 > q'$. Then, according to the model when the presentation order is p, q', we have

$$W(p) = \frac{\psi(x_p) - \psi(\rho_+)}{\psi(x) - \psi(\rho_+)},$$
$$W(q') = \frac{\psi(x_{p,q'}) - \psi(\rho_-)}{\psi(x_p) - \psi(\rho_-)}.$$

Therefore

$$W(p)W(q') = \left(\frac{\psi(x_p) - \psi(\rho_+)}{\psi(x) - \psi(\rho_+)}\right)\left(\frac{\psi(x_{p,q'}) - \psi(\rho_-)}{\psi(x_p) - \psi(\rho_-)}\right). \tag{9}$$

Next, consider the presentation order q', p :

$$W(q') = \frac{\psi(x_{q'}) - \psi(\rho_-)}{\psi(x) - \psi(\rho_-)},$$
$$W(p) = \frac{\psi(x_{q',p}) - \psi(\rho_+)}{\psi(x_{q'}) - \psi(\rho_+)}.$$

Thus,

$$W(q')W(p) = \left(\frac{\psi(x_{q'}) - \psi(\rho_-)}{\psi(x) - \psi(\rho_-)}\right)\left(\frac{\psi(x_{q',p}) - \psi(\rho_+)}{\psi(x_{q'}) - \psi(\rho_+)}\right). \tag{10}$$

Equating (9) to (10), $x_{p,q'} = x_{q',p}$ holds iff $\rho_+ = \rho_-$, as claimed. ■

Proposition 2.

By the proof of Proposition 1, we have for the two p, q orders

$$\tau_+(p,q) = W(p)W(q) = \frac{\psi(x_{p,q}) - \psi(\rho_+)}{\psi(x) - \psi(\rho_+)} \tag{11}$$
$$\tau_+(q,p) = W(q)W(p) = \frac{\psi(x_{q,p}) - \psi(\rho_+)}{\psi(x) - \psi(\rho_+)}.$$

Because we know $x_{p,q} = x_{q,p}$ it suffices to look only at the first display. Holding p and q fixed, we may vary x over several values and determine $x_{p,g}(x)$ empirically. Thus, we have a set of equations with two parameters, $\tau_+(x)$ and ρ_+. Assume (3), i.e., ψ is a power function, and so (11) can be rewritten

$$\rho_+^\beta = \frac{x^\beta \tau_+(p,q) - x_{p,q}^\beta}{\tau_+(p,q) - 1}, \tag{12}$$

which permits us to estimate ρ_+. Exactly analogous calculations hold for $p' < 1, q' < 1$ with parameters $\tau_-(p', q')$ and ρ_-.

Given that monotonicity of W, then for $p' < 1 < p$ and $q' < 1 < q$ we know

$$\tau_-(p', q') < \tau_+(p,q).$$

■

Proposition 3.
Case 1. For the mapping $f \to g \to g$, i.e., where $x^f \to x_p^{f,g}$ and $x_p^{f,g} \to x_{p,q}^{f,g,g}$, we have

$$W(p) = \frac{\psi(x_p^{f,g}) - \psi(\rho_+^{f,g})}{\psi(x^f) - \psi(\rho_+^f)},$$
$$W(q) = \frac{\psi(x_{p,q}^{f,g,g}) - \psi(\rho_+^{g,g})}{\psi(x_p^{f,g}) - \psi(\rho_+^{f,g})}.$$

Thus,

$$\begin{aligned} W(p)W(q) &= \left(\frac{\psi(x_{p,p}^{f,g,g}) - \psi(\rho_+^{f,g})}{\psi(x^f) - \psi(\rho_+^{f)})}\right)\left(\frac{\psi(x_{p,q}^{f,g,g}) - \psi(\rho_+^{g,g})}{\psi(x_p^{f,g}) - \psi(\rho_+^{f,g})}\right) \\ &= \frac{\psi(x_{p,q}^{f,g,g}) - \psi(\rho_+^{g,g})}{\psi(x^f) - \psi(\rho_+^f)}. \end{aligned}$$

Similarly, for the order q, p,

$$\begin{aligned} W(q)W(p) &= \left(\frac{\psi(x_{q,p}^{f,g,g}) - \psi(\rho_+^{g)})}{\psi(x_q^{f,g}) - \psi(\rho_+^{f,g})}\right)\left(\frac{\psi(x_q^{f,g}) - \psi(\rho_+^{f,g})}{\psi(x^f) - \psi(\rho_+^f)}\right) \\ &= \frac{\psi(x_{q,p}^{f,g,g}) - \psi(\rho_+^{g,g})}{\psi(x^f) - \psi(\rho_+^f)}. \end{aligned}$$

By the commutativity of multiplication, equate these and it is immediate that

$$x_{p,q}^{f,g,g} = x_{q,p}^{f,g,g}$$

i.e., commutativity holds.

Case 2 is analogous. And following the pattern of Proposition 1, Case 3 requires $\rho_+ = \rho-$ for both frequencies. ■

Proposition 4.
For the case $f \to g \to f$, i.e., where $x^f \to x_p^{f,g}$ and $x_p^{f,g} \to x_{p,q}^{f,g,f}$, we have for $i = +, -$

$$W(p) = \frac{\psi(x_p^{f,g}) - \psi(\rho_i^{f,g})}{\psi(x^f) - \psi(\rho_i^f)},$$
$$W(q) = \frac{\psi(x_{p,q}^{f,g,f}) - \psi(\rho_i^f)}{\psi(x_p^{f,g}) - \psi(\rho_i^{g,f})}.$$

so

$$\begin{aligned} W(p)W(q) &= \left(\frac{\psi(x_p^{f,g}) - \psi(\rho_i^{f,g})}{\psi(x^f) - \psi(\rho_i^f)}\right)\left(\frac{\psi(x_{p,q}^{f,g,f}) - \psi(\rho_i^f)}{\psi(x_p^{f,g}) - \psi(\rho_i^{g,f})}\right) \\ &= \left(\frac{\psi(x_{p,q}^{f,g,f}) - \psi(\rho_i^f)}{\psi(x^f) - \psi(\rho_i^f)}\right)\left(\frac{\psi(x_p^{f,g}) - \psi(\rho_i^{f,g})}{\psi(x_p^{f,g}) - \psi(\rho_i^{g,f})}\right). \end{aligned}$$

For the case $f \to f \to f$ (Case 1 of Proposition 1), yielded

$$W(p)W(q) = \frac{\psi(x_{p,q}^{f,f,f}) - \psi(\rho_i^f)}{\psi(x^f) - \psi(\rho_i^f)}$$

which means

$$x_{p,q}^{f,g,f} \approx x_{p,q}^{f,f,f} \text{ iff } \rho_i^{f,g} \approx \rho_i^{g,f}.$$

■

Appendix B
Detailed Results

Tables B1 and B2 contain the detailed results of tests conducted.

Cond.	Resp.	$x_{p,q}^{f,g,g}$	$x_{q,p}^{f,g,g}$	$t_{p,q}^{f,g,g}$	$t_{q,p}^{f,g,g}$	N	p_x	p_t
Proposition 3, Case 1: $p > 1, q > 1$								
C_1	10	85.43 (3.65)	85.37 (3.59)	84.60 (3.18)	84.25 (4.03)	30	.847	.929
	22	78.10 (5.47)	78.53 (4.78)	77.22 (5.69)	76.10 (5.00)	30	.784	.374
	75	71.33 (3.49)	71.78 (3.41)	70.75 (4.39)	71.80 (3.86)	30	.711	.205
C_2	56	78.40 (5.89)	80.45 (5.05)	78.93 (6.92)	79.12 (6.00)	30	.151	.959
	66	82.07 (4.18)	81.13 (5.40)	79.40 (2.86)	79.53 (3.75)	30	.542	.726
	71	86.77 (3.33)	87.63 (1.96)	86.97 (1.76)	85.85 (3.18)	30	.372	.188
	75	75.88 (4.36)	76.16 (3.62)	74.86 (3.30)	76.09 (4.99)	37	.618	.137
C_3	66	77.85 (5.41)	78.19 (2.99)	82.00 (2.53)	82.46 (4.46)	30	.646	.544
	75	72.47 (3.69)	70.98 (4.68)	73.63 (4.24)	74.30 (4.57)	30	.396	.311
C_4	47	78.23 (4.89)	77.93 (5.06)	76.60 (4.25)	77.13 (5.51)	30	.635	.721
	66	80.85 (4.14)	81.30 (5.02)	79.07 (4.90)	79.23 (4.40)	28	.941	.875
	68	77.83 (5.90)	75.55 (6.61)	74.85 (6.52)	73.55 (4.59)	30	.157	.148
	71	86.47 (2.35)	86.17 (2.27)	86.37 (3.06)	84.98 (2.48)	30	.460	.021*
	75	78.53 (5.82)	78.12 (5.31)	77.93 (5.60)	78.52 (6.03)	28	.761	.824
	76	86.63 (2.00)	85.20 (2.77)	86.41 (1.82)	84.39 (5.17)	27	.075	.834
C_5	75	80.78 (2.94)	81.89 (3.96)	79.09 (3.96)	81.84 (3.89)	32	.107	.017*
	76	79.52 (3.54)	82.14 (3.00)	79.25 (4.17)	81.70 (3.56)	28	.001**	.012*
C_6	10	86.53 (5.81)	86.89 (4.38)	83.68 (5.41)	84.07 (4.70)	28	.497	.863
Proposition 3, Case 2: $p' < 1, q' < 1$								
C_7	47	71.30 (6.20)	71.40 (3.56)	57.80 (8.61)	57.30 (8.57)	30	.946	,894
C_8	59	75.92 (4.81)	74.23 (4.24)	72.93 (7.00)	70.50 (5.69)	30	.123	.082
	76	52.40 (7.48)	48.75 (10.20)	48.84 (12.45)	48.28 (7.42)	32	.081	.572
C_9	59	75.87 (3.45)	72.62 (5.41)	70.48 (6.54)	69.87 (7.63)	30	.023*	.976
	63	50.83 (10.29)	50.08 (13.10)	47.65 (10.49)	47.58 (12.71)	30	.496	.953
Proposition 3, Case 3: $p' < 1, q > 1$								
C_{10}	59	77.83 (2.66)	87.70 (3.27)	77.00 (2.48)	87.77 (2.95)	30	<.001**	<.001**
	76	64.35 (3.41)	75.25 (4.11)	62.30 (3.88)	76.05 (2.74)	30	<.001**	<.001**
C_{11}	59	75.20 (5.03)	83.50 (3.94)	75.30 (4.09)	84.70 (2.57)	30	<.001**	<.001**
	76	63.65 (3.73)	76.27 (4.00)	61.36 (4.60)	77.10 (3.59)	29	<.001**	<.001**

*Means rejection at the .05 limit. **Means rejection at the .01 limit.

Table B1: Tests of Proposition 3. The table lists by condition and by respondent the detailed results of the testing. For each condition is listed the participating respondent, his/her averaged estimates and their standard deviations as well as the number of observation produced be each person. In the final two columns list the results of the statistical testing where the notation p_x means $x_{p,q}^{f,g,g} = x_{q,p}^{f,g,g}$ and p_t means $t_{p,q}^{f,g,g} = t_{q,p}^{f,g,g}$

Cond.	Resp.	frequency	$x_{p,q}$	$x_{q,p}$	N	$p_{\rightarrow}$	$p_{\times}$
		Proposition 4, Case 1: $p > 1, q > 1$					
C_{12}	10	$f \rightarrow f \rightarrow f$	80.02 (2.43)	79.17 (3.37)	34	.279	.005
		$f \rightarrow g \rightarrow f$	83.06 (4.00)	83.47 (3.77)		.922	$< .001$
	22	$f \rightarrow f \rightarrow f$	71.70 (3.58)	72.30 (3.12)	30	.563	.001
		$f \rightarrow g \rightarrow f$	75.77 (4.06)	75.25 (4.22)		.733	.005
		Proposition 4, Case 2: $p' < 1, q' < 1$					
C_{13}	76	$f \rightarrow f \rightarrow f$	35.11 (4.50)	35.71 (3.55)	28	.873	$< .001$
		$f \rightarrow g \rightarrow f$	40.68 (4.99)	39.96 (3.19)		.423	$< .001$
	22	$f \rightarrow f \rightarrow f$	65.52 (3.31)	64.64 (2.04)	28	.702	.286
		$f \rightarrow g \rightarrow f$	67.39 (3.01)	66.44 (3.22)		.120	.007

Table B2: Tests of Proposition 4: Equality of reference points $\rho^{f,g} \approx \rho^{g,f}$. The table lists by condition and by respondent the detailed results of the testing. For each condition, the participating respondent, his/her averaged estimates and their standard deviations as well as the number of observation produced by each person. In the final two columns list the results of the statistical testing where the notation $p_{\rightarrow}$ means either $x^{f,f,f}_{p,q} = x^{f,f,f}_{q,p}$ or $x^{f,g,f}_{p,q} = x^{f,g,f}_{q,p}$ and that $p_{\times}$ means either $x^{f,f,f}_{p,q} = x^{f,g,f}_{q,p}$ or $x^{f,g,f}_{p,q} = x^{f,f,f}_{q,p}$