Brightness of Different Hues is a Single Psychophysical Ratio Scale of Intensity

Ragnar Steingrimsson, R. Duncan Luce, 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

Recent studies based on testable behavioral axioms have concluded that psychological scales of subjective intensive attributes form ratio scales. These studies have shown that a certain commutativity of proportion property must hold under either successive increases or successive decreases provided, e.g., all other independent dimensions are fixed. These data, as well as much data in the literature, suggests that such psychophysical functions are closely approximated by power functions of physical intensity. However, until recently, limited attention has been paid to whether or not such subjective intensity scales differ when an independent dimension such as frequency or wavelength (pitch in audition; hue in vision; etc.) is varied. Using a simple and favorably tested theoretical model for global psychophysics, Luce, Steingrimsson, & Narens (2010) arrived at a necessary and sufficient cross-frequency, commutativity condition for there to exist a common intensity ratio scale. Here we show that brightness—already established to be a ratio-scalable dimension—and hue satisfy the same condition. The comparison of cross-dimensional commutativity of stimuli with that of commutativity of stimuli within a dimension discriminates between a general representation of the ratio scale property and a special case of it. Future questions to be investigated: Does the theory extend to other intensive continua (prothetic attributes)? If so, which ones? And does it extend to cross-modal matching?

Luce, Steingrimsson, and Narens (2010) formulated an axiomatic theory for the attribute of intensity to have a ratio scale type that is unchanged when another relevant variable, such as signal frequency, is varied. At the center of that development was a behavioral property of commutative proportion judgments which, if true, is necessary and sufficient to establish a ratio scale of subjective intensity independent of, say, frequency. They evaluated this theory of intensity/frequency pairs for loudness/pitch pairs and found strong support for a common ratio scale for these dimensions. The present research aims to evaluate this property for perception of luminance/hue pairs. Should the results be favorable, it invites subsequent work involving the commonality of scales across two distinct intensity modalities, e.g., loudness and brightness. And, if so, then to the intriguing question of whether sensation of intensity over all prothetic (intensive) continua may be described as a unitary mechanism of subjective intensity.

Luce et al. (2010) provided some historical background in addition to a report of the theory including proofs. Here we summarize that material and add some material specific to the brightness domain.

The article is structured as follows:

- Some relevant background is presented.
- The commutativity property is presented along with the necessary theoretical background.
- An experiment evaluating the commutative property for the perception of luminance and hue is reported.
- The final discussion, summarizes the evidence for the commonality of scales of subjective intensity, and it outlines the natural next steps to be taken.

Background

Psychophysical measurement from Fechner to present

Fechner, the proverbial father of scientific psychology in general and qualitative modeling of it in particular, wrote that

> [P]sychophysics refers to the *physical* in the sense of physics and chemistry, to the *psychical* in the sense of experiential psychology. ... Physical processes that accompany or underlie psychical functions, and consequently stand in a direct functional relationship to them, we shall call psychophysics ... Psychophysics, already related to psychology and physics by name must on one hand be based on psychology, and the other hand promises to give psychology a mathematical foundation ... [T]he general principles of psychophysics will involve the handling of qualitative relationships, just as in physics. (p. 9–10, 1860/1966).

The modeling approach he and we favor—discovering descriptive qualitative invariances that together form a model from which behavioral predictions follow—is very much in line with the tradition of classic static physics. However, Fechner carefully made a distinction between this mapping and the nature of the mechanism that exists in between. He wrote:

This research was supported in part by National Science Foundation grant BCS-0720288 and in part by the U. S. Air Force Office of Research grant FA9550-08-1-0468—any opinion, finding, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or the Air Force. We thank Drs. G. Iverson and K. A. Jameson for detailed and helpful comments on a draft version of this paper.

> By its nature, psychophysics may be divided into an outer and inner part, depending on whether consideration is focused on the relationship of the physical to the body's external aspects, or on [its] internal functions...The truly basic empirical evidence of psychophysics can be sought only in the realm of outer psychophysics...as it is the only part that is available to immediate experience...From physics outer psychophysics borrows aids and methodology; inner psychophysics leans more to physiology and anatomy...[outer psychophysics seeks] lawful relationships between sensation and the stimulus...Based on them...physical measurement yields psychic measurement on which we can base [important] arguments. (p. 9–11, 1860/1966).

Interestingly, Fechner seems to suggest that in order to solve effectively the problem of inner psychophysics it is first necessary to address the one of outer psychophysics. According to that view, any current brain imaging experiment, for instance, should be accompanied by a well understood outer psychophysical model to be correlated with brain activity. Our approach is very much in the spirit of Fechner's outer psychophysics.

Our Approach

Our approach, which mimics classical static physics, constructs a behavioral model composed of several descriptive (= testable) axiomatic invariances from which a numerical representation—a measurement scale and its properties—is derived. So we begin by saying just what constitutes such an axiom and how its accuracy is empirically evaluated.

- These invariances are in qualitative mathematical form and, as such, are not domain specific. This means they can apply to any domain for which their primitives can be interpreted.
- A concrete example: Consider the case when x and u as physical light signals, then (x, u) denotes a *joint presentation* of these signals where x is presented to the left eye and u to the right eye.
- The invariance should be testable and provide some constraint on behavior. An example is the property of *commutativity* of joint presentations, i.e., $(x, u) \sim (u, x)$, where $\sim$ denotes indifference. Note the analogy to arithmetic commutativity.
- A psychological interpretation could be: Does the perception of (x, u) actually commute? That is, Are the stimuli (x, u) and (u, x) perceptually equivalent? The empirical answer to this question is that such an interpretation of commutativity often fails in psychology.
- An empirical way to evaluate, for brightness signals, whether

$$(x, u) \sim (u, x)$$

is satisfied, it to present two square homogenous signals each subtending, say, 8^o of visual angle in a horizontal alignment with, respectively, the intensity of x and u. By using a stereo-scope, respondents see a fused image of x and u resulting in a unitary cyclopic image which is the joint presentation (x, u). Also on the screen are two squares of intensity z (which also fuse to a cyclopic image), an intensity the respondent can adjust and does until the match of $(x, u) \sim (z, z)$ is accomplished. Analogously, the match $(u, x) \sim (z', z')$ is accomplished on a different trial. The final step is to evaluate statistically the hypothesis that $z = z'$.[1]

- This process is repeated a number of times and for several instantiation of (x, u) pairs.

The Primary Psychophysical Method. The primary psychophysical method employed is Magnitude Production (MP). This well-known method was pioneered by S. S. Stevens (1975 provides a comprehensive account of his discovery and subsequent evolution of this technique). Simply put, a respondent is presented with, e.g., (x, u) and a number $p > 0$ and is instructed to produce, using some variation on the method of adjustment, an intensity, e.g., light, z_p such that (z_p, z_p) appears to be p-"times" as intense, or bright, as (x, u). The aforemention matching technique is simply MP with $p = 1$.

Scales of Measurement. Fechner's foundational goal of placing the study of psychology on a firm mathematical basis has long been pursued but yet is still quite incomplete. In particular, as Fechner suggested (1860/1966, p. 10), the primary problem of psychology is that the object which we seek to measure is not directly observable—only the stimulus and the response are available as data. The ratio scalability of intensive continua of psychological variables has given rise to extensive theoretical work over the past 150 years. A compendium of such measurement results was formulated in the 3-volumes of *Foundations of Measurement* (1971, 1989, 1990). However, this general foundational work did not directly extend to Stevens' form of measurement until Narens (1996) undertook the task of axiomatizing what he thought Stevens' underlying assumptions for his methods must have been. In the context of representational theory of measurement, Narens did arrive at a mathematically sound version of Stevens' methods in the form of testable axioms which, if correct, allowed for subjective measurements on ratio scales. Narens' (1996) theory led to two behavioral properties that together are equivalent to ratio scalability: a commutative one and a multiplicative one of magnitude productions. Both were extensively evaluated in a variety of domains, including loudness, brightness, perceived contrast, and perceived size (Ellermeier & Faulhammer, 2000; Zimmer, Luce, & Ellermeier, 2001; Ellermeier, Narens, & Dielmann, 2003; Steingrimsson & Luce 2005a, 2007; Augustin & Maier, 2008; Steingrimsson, 2009, Steingrimsson, in preparation b), and all have concluded that the commutative property holds, but that multiplicativity fails. Steingrimsson and Luce (2007) pointed out that the negative results for the multiplicative property arose from just one of Narens' (1996) assumptions that echoed the one Stevens had made namely, that subjective weighting W (called f by Narens, 1996) of numbers exhibited the special case of $W(1) = 1$. Auditory studies showed that this assumption failed for loudness,[2] and Steingrimsson & Luce (2007) showed that altering the multiplicative property slightly to what they called the k-multiplicative property, which was subsequently found to hold for loudness, then one

[1] The reader may wonder why not carry this out as a discrimination task by simply asking whether two squares appear equal. In that specific case, it may well work, but for other somewhat more complicated properties, discrimination methods fail to adaptable to what is needed. Further, place-order can create a bias, which is avoided using matching, but any sequential presentation will result in a Time-Order Error. So in general, the discrimination methods have not proved to be ideal for this type of study. Rather, some variation on the method of adjustment has generally been the choice of researchers.

[2] In loudness, the stimuli must be presented sequentially, and, empirically, this leads to the second of two identical stimuli to have different perceived loudness than the first.

could still prove that magnitude production has a ratio scale representation on continua of physical intensities. A post-hoc analysis of data existing before this result remained consistent with a ratio-scale for the tested domains.

Ratio scalability—Narens' commutativity property

Narens' (1996) commutative property asserts that two successive productions, one for "p-times", $p \geqslant 1$, and one for "q-times", $q \geqslant 1$, end up at the same intensity independent of the order of their productions–in the present context see later in the current article and formally Luce et al. (2010) for a generalization to $p \geqslant 0$ and $p \geqslant 0$. More specifically, for brightness magnitude production, it says: Suppose that x is an arbitrary luminance that is equal or greater than its threshold intensity and let p and q be two positive numbers. The respondent is first asked to produce a luminance x_p that is seen as "p times as bright as x", and then to produce a luminance $x_{p,q}$ that is "q times as bright as x_p". In a similar manner, the respondent produces the luminance $x_{q,p}$, using the same estimation sequence with the order of p and q reversed. The commutativity property holds if and only if

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

Naturally, this axiom needs testing before any assumption related to ratio-scalability is used. This property has been found to hold in several domains: loudness (Steingrimsson & Luce, 2005a), size of circles (Augustin, T., & Maier, K., 2008), brightness (Steingrimsson, 2009), and perceived contrast (Steingrimsson, in preparation b).

Commonality of scales over pair of attributes

Luce et al. (2010), building upon Luce (2004), Narens (2006), and Steingrimsson and Luce (2007) developed a new kind of paradigm applicable to a wide variety of psychophysical situations, including ones involving mixed modalities (see the section "The magnitude production representation"). The main conceptual novelty was to extend the commutative property to two attributes, such as frequency, leading to a common ratio scale representation intensity over that attribute. And they demonstrated the effectiveness of this new paradigm for loudness when frequency (pitch) was varied, which yielded evidence for a common scale of loudness across frequencies. Here we evaluate this hypothesis for physical attributes of luminance and hue.

Related literature

Our approach, although historically stemming from Fechner's original formulation of outer psychophysics, has only quite recently matured to a point where the studies undertaken here cold be based on a sound theoretical basis. It is perhaps, therefore, not surprising that, despite over a century of color research, a literature search reveals little by way of material that is relevant to our approach. In particular, the interest in classical psychophysics in the spirit of Stevens, who used approaches based on production and estimation, appears to have waned in the latter part of the 20th century. As the Fechner quote above clearly stated, a central task of psychophysics is to discover the mapping from the sensation evoked by a physical stimulus into a numerical structure. Using our amended method of Stevens' applied, specifically, to the studies of brightness, Steingrimsson (2009, 2010) established it to be a ratio scaled variable, a result that happens to be consistent with numerous function fitting-studies, summarized in Stevens (1975).

Of all the studies undertaken by Stevens on luminance, and most specifically Stevens and Stevens (1962), evaluated only achromatic stimuli. And in fact, surprisingly, to our knowledge, only Ekman, Eisler, and Künnapas (1960) have ever reported analogous data for chromatic stimuli. Ekman et al. (1960) established scales of brightness as a function of luminance for light at six frequencies using both magnitude estimation and production methods and the results, pooled over respondents, were consistent with a common scale of intensity across all six frequencies and a psychophysical function hat was a power function with an exponent of $\sim 1/3$. Their result is consistent with the studies of a achromatic stimuli, summarized in Stevens (1975) . The same conclusion may be derived from a recent computational model of color perception (Romney & Chia, 2009). Outside of these studies, the approach appears not to have captured the interest of researchers until now—in fact, none of the relevant citation databases up to the publication of this paper report any citation to Ekman et al. (1960) study.

It is curious that the literature on the commonality of scales in cross-modal situations is a bit richer than for hues and in sum appears to provide favorable evidence for the commonality hypothesis. Among these studies are Marks, Szczesiul, & Ohlott (1986) broadly supporting the commonality notion for both auditory loudness versus vibratory touch and auditory loudness versus visual brightness. Ward (1990), working with auditory and visual stimuli reported similar results and, in addition, he favorably evaluated the property of double cancelation in the cross-modal situation, establishing an additive conjoint structure for the stimulus pairs. Finally Nordin (1994) reported similar results for the intensity of odor, loudness, and brightness. All three studies also agree on and, by way of data, demonstrate substantial absolute intra-individual differences but similarity in patterns of responses, i.e., in producing responses consistent with a common scale of intensity across the tested domains.

Clearly, on the basis of the aggregate of the extant literature, we may be justified in expecting our inquiry to yield confirmatory results for the commonality of scales across hues. Indeed, we conclude, with certain caveats that a single ratio-scale obtains for perceived luminance across different hues.

Summary of Luce et al's (2010) Theory

The magnitude production representation

When a respondent is asked to produce the signal x_p that stands in ratio p to a given signal x, Luce (2004, 2008) showed that two functions exist: ψ a strictly increasing, psychophysical ratio scale over the set of signal intensities, and W a cognitive distortion function over numbers, as well as parameters ρ_i called *reference signals* that satisfy the constraint

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

For our purposes here, we treat the reference signal ρ_i to be a physical intensity parameter generated by the respondent. The reason for assuming $\rho_+ \neq \rho_-$ is that the data themselves support it. We currently have no theory of the reference signals, an unfortunate fact.

As mentioned earlier, the article concerns successive productions. For example with x_p satisfying (1), then $x_{p,q}$ is constructed relative to x_p using (1) with x_p playing the role of x in that equation and x_q playing the role of x_p in that equation. The article addresses whether or not for brightness commutativity obtains in the sense

$$x_{p,q} = x_{q,p}, \tag{2}$$

which under our model is the signature of a ratio scale. Proposition 1, quoted below from Luce et al. (2010) states conditions when (2) follows from the representation (1).

A great deal of literature cited by Stevens (1975) provides data supporting that ψ in brightness is well-fit by a power function, i.e.,

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

However, these studies averaged the data over subjects. Using our theory, Luce (in preparation) has outlined a way to estimate β for individual respondents, which we have not yet employed.

The weighting function W appears to be well described by the following functional form due to Prelec (1998):

$$W(p) = W(1) \begin{cases} \exp\left[-\omega\left(-\ln p\right)^{\mu}\right] & (0 < p \leq 1) \\ \exp\left[\omega'\left(\ln p\right)^{\mu'}\right] & (1 < p) \end{cases} . \tag{4}$$

Note that this specializes to a power function when $\mu = \mu' = 1$. Luce (2001) showed that a behavioral property called "double reduction invariance" (D-RI) is equivalent to (4) and is simpler to test than the behavioral condition that Prelec (1998) had presented earlier. Steingrimsson and Luce (2007) tested established that for most respondents and conditions $\mu = \mu' = 1$ and for the remaining respondents D-RI was evaluated and sustained.

Four propositions of Luce et al. (2010)

The following four proposition and accompanying figures designed to make clear the cases being described (these are inspired by and similar to those used by Luce et al,2010). Note that because p and q are involved, four cases are distinct: 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 $q \geq 1$. The numbers less than 1 are identified by a prime. Sometimes [see (1) and Case 3 of Proposition 1] we use $+$ and $-$ to refer to the sign of $p - 1$, with $+$ including $p = 1$.

Single frequency cases. Our first two Propositions concern the single intensity dimension shown as Case 1 in Figure 1.

Proposition 1 *Assuming that the general model (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 $\rho_+ = \rho_-$.*

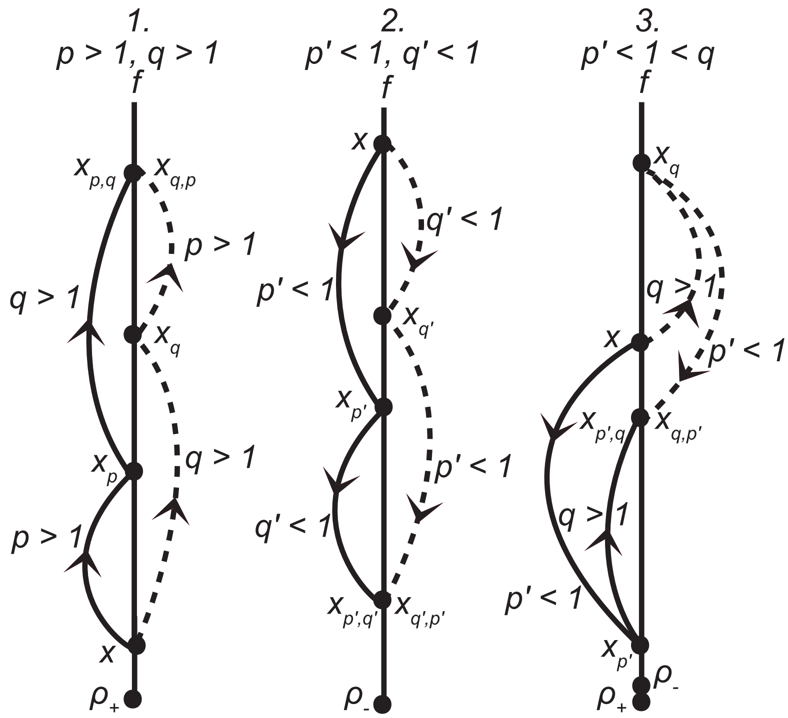

Figure 1. Depicted are the three cases of Proposition 1. In each graph, the solid lines depict the case of magnitude production with p followed by q and the dotted ones depict the case of magnitude production with q followed by p. Commutativity is said to hold if statistically in Case 1 if $x_{p,q} = x_{q,p}$, in Case 2 if $x_{p',q'} = x_{q',p'}$, and in Case 3 if $x_{p',q} = x_{q,p'}$. The reference point for $p, q > 1$ is ρ_+ and for $p', q' < 1$ it is ρ_-. In Case 3 both reference points are in play but commutativity is predicted only if $\rho_+ = \rho_-$.

Luce et al. (2010) describe the extensive support that exists for Cases 1 and 2. The extant evidence rejects Case 3, i.e., that $\rho_+ \neq \rho_-$. This evidence includes data summarized in Fig. 1 of Steingrimsson and Luce (2006), some data reported in Appendix E of Steingrimsson and Luce (2007), and systematic study of it by Steingrimsson (in preparation b).

Cross-dimension commutativity. The above results were modified to two dimensions, say f and g, by suitable super- and subscripting. The superscripts describe which dimension is being matched to which. The subscripts identify the production value of the match. So, in this notation staying in one dimension, f to f, is written $x_p^{f,f}$. In the case where the judgment is from dimension f to dimension g, then we write $x_p^{f,g}$. Then, as long as the roles of p and q are simply switched, the Proposition 1 continues to be satisfied.

With successive productions, that from $f \to g$ yields $x_p^{f,g}$. For the next production there are two possible cases: $f \to g \to g$ and $f \to g \to f$. These lead to the productions $x_{p,q}^{f,g,g}$ and $x_{p,q}^{f,g,f}$. When the roles of p and q are reversed, the notation is appropriately changed to reflect this.

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 dimensions.

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

Proposition 2 *Assuming that the general model (1) holds, 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 $\rho_+^g = \rho_-^g$.*

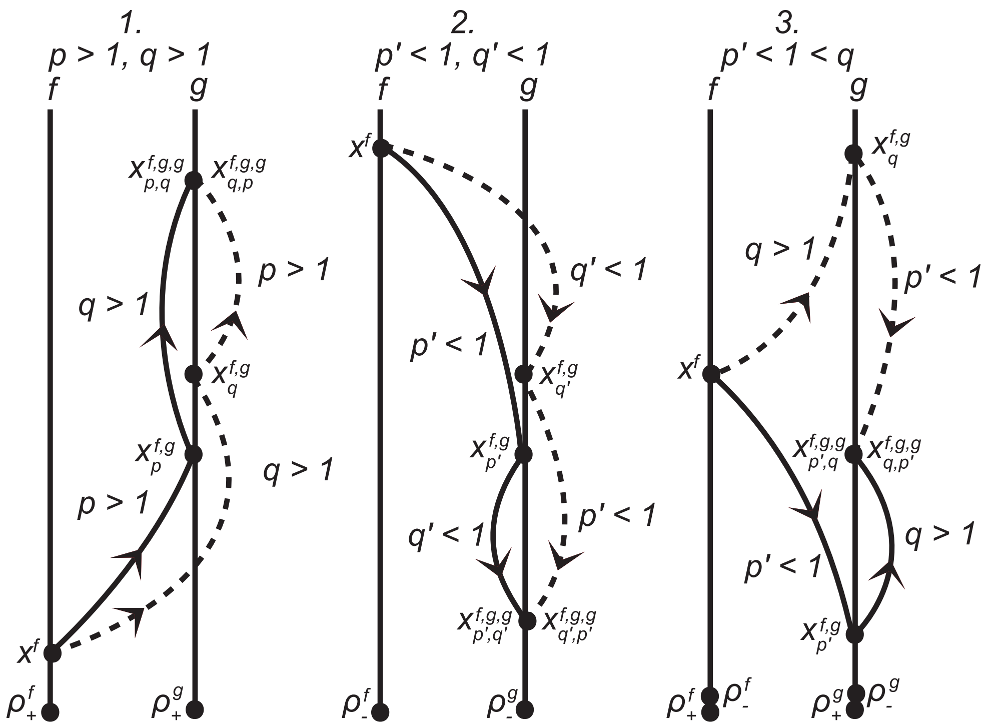

Figure 2. Depicted are three cases Proposition 2, which parallels that of Proposition 1 when extended to two dimensions, f, g. In each of the three cases, the solid lines depict the case of magnitude production with p followed by q, and the dotted lines depict magnitude production starting with q followed by p. Commutativity is found to hold if in Case 1, statistically $x_{p,q}^{f,g,g} = x_{q,p}^{f,g,g}$, in Case 2 if $x_{p',q'}^{f,g,g} = x_{q',,p'}^{f,g,g}$, and in Case 3 if $x_{p',q}^{f,g,g} = x_{q,p'}^{f,g,g}$. The reference point for $p, q > 1$ is ρ_+ and for $p', q' < 1$ it is ρ_-. In Case 3 both reference points are in play and commutativity is predicted only if $\rho_+ = \rho_-$.

There is an additional case to explore, the comparison of $f \to f \to f$ to $f \to g \to f$.

We have already 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?

When crossing dimensions in the order f, g, f, we need to replace ρ^f and ρ^g by a notation that makes clear that ρ^g need not be the same when going from f to g with going from g to f, and as we shall see the data show that this caution is needed. We distinguish these as, respectively, $\rho^{f,g}$ and $\rho^{g,f}$. This comparison depicted in Figure 3.

Proposition 3 *Assuming that the general model (1) holds, then*

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

if and only if

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

Because Narens' (1996) theory has no reference points, it predicts the above which, as we showed in loudness, is wrong and we will also show below is wrong for brightness.

Method

The goal of the experiment is to test the same scale hypothesis of Section "Cross-dimension commutativity" as well as whether the two reference signals on different dimension agree or not (Section "Comparison of $f \to f \to f$ to $f \to g \to f$ ") where the intensity dimension is brightness and the varied dimension is hue. In speaking of hue as a dimension, we can talk of the wavelength or its inversely related frequency. In the following, we will speak of changing of hue as a change in the wavelength of light. Hence we study the subjective scale of luminance (brightness) as wavelength (hues) is varied.

Respondents

A total of 10 respondents—mostly students from the University of California, Irvine—participated in the experiment reported. All respondents reported normal or corrected-to-normal vision. All respondents, except the participating coauthor (R22), received compensation of $12 per session, maximally one hour in duration. 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 Review Board. Respondents are given unique identifiers of the form *Rm* where *m* is a counter signifying only when an individual came to participate in any experiment we have conducted.

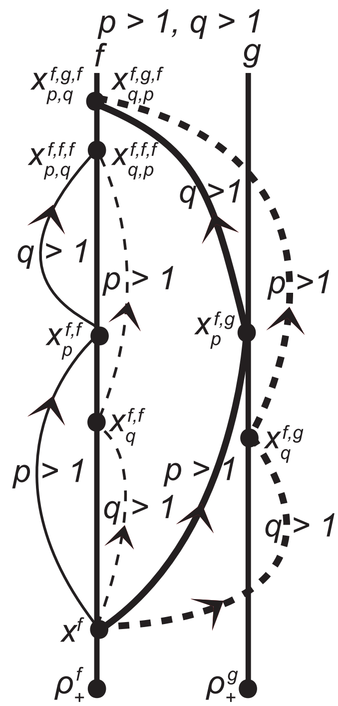

Figure 3. Depicted is the condition of Proposition 3. The narrow lines (solid and dotted) are on a single dimension and the thicker lines (solid and dotted) go across the two dimensions f, g. Independent of thickness, the solid lines depict the case of magnitude production with p followed by q whereas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

Apparatus

The stimuli were generated by a personal computer and displayed on a monitor (Eizo RadiForce RX320) with automatic luminance uniformity equalizer and blacklight sensor to compensation for luminance fluctuations caused by ambient temperature and passage of time as well as a built in *gamma correction* (the standard term in vision for correcting for the power relationship between LUT values and intensity). The diagonal size is 56 cm and maximum resolution of 1536 x 2048 and luminance of 742 cd/m^2. Luminance measures were taken using Photo Research's PR-670 SpectraScan Spectroradiometer which verified the monitor calibration and to determine luminance values of stimuli.

Notational convention and calibration

Stimuli were presented simultaneously to the two eyes. This stimulus is most accurately expressed as (x, x), which means the joint presentation of the luminance x to the left eye and luminance x to the right eye. Even though the same luminance is presented to both eyes, we need to note that the theory is cast in terms of intensity increments above threshold and if x_l and x_r are threshold intensities in the left and the right eye, respectively, then the effective stimulus consists of $x_l = x' - x_l$ and $x_r = x' - x_r$ where (x_l, x_r) are the actual intensities presented. However, because all signals were well above threshold and the respondents were selected for normal vision, the error in reporting intensities x in cd/m^2 is negligible and in all cases, the same intensity is presented to each eye, we simplify the stimulus reporting to writing x to indicate the intensity of this stimulus.

When we write, e.g., x^f, then f stands for wavelength, which for a monitor is expressed as a the triple (R, G, B) where each of R, G, and B is an 8-bit value, i.e., an integer from 0 to 255. Hence we measured the intensity of each of the three separate guns and fit the result, as is customary, to a power function in the vision literature traditionally expressed with an exponent γ and hence the resulting function as is traditionally referred to as a *gamma function*[3], to convert it to cd/m^2; the standard stimuli are reported using the CIE-xyz standard. The resulting exponents were $\gamma_{\mathrm{R}} = 2.28$, $\gamma_{\mathrm{G}} = 2.31$, and $\gamma_{\mathrm{B}} = 2.14$. So, while the exponents are quite close, it is clearly the case that when two or three guns form a stimulus, there will be a minor hue shift across the intensity spectrum. This is not a theoretical problem, and it is small enough that it was not noticeable to respondents (subjective reports). For the purpose of standardization, the CIE-xyz measures are provided for all fixed stimuli (standards).

Stimuli

The basic stimulus consisted of two chromatic squares having the intensity and wavelength x^f (the standard) and z^g (the user-varied stimulus) both subtending 10 degrees of visual angle, presented on a uniform achromatic background of 4 cd/m^2, and arranged in vertical position with an 8 decree separation. A sample stimulus is depicted graphically in Figure 4.

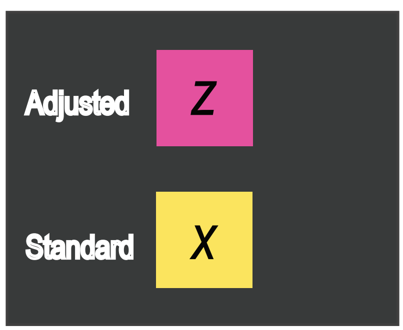

Figure 4. The stimulus consists of two hues displayed as squares. The lower one is the standard and the upper one is a variable one which the respondent is instructed to make a proportion p as bright as the lower one. The labels in the figure are not present in the experiments. However, not shown in the figure but indicated in the actual experiment, is the number of blocks of trials as well as the proportion instructions (upper left corner of the screen).

[3]This usage should not be confused with the term *gamma function* as used in mathematics, physics, and areas of engineering.

Statistical method and presentation of results

We evaluate parameter-free null hypotheses that have the generic form $L_{\text{side}} = R_{\text{side}}$. Lacking a model of how individuals relate, all data analysis is done on individual data (as argued by Luce, 1995, p. 20). The statistical approach is patterned on traditions in physics, where for the most part null hypotheses are theoretical invariances that the experimenter is attempting to evaluate: does the invariance hold or does it not? This takes the form of a statistical criterion consisting of three interlocked components.

• Test Statistic: Due to lack of a parametric information, we use the nonparametric Mann-Whitney U test at a significance level of .05, a choice presently quite common for similar situations (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, 2010). Intensity adjustments are made using discrete step-sizes, complicating direct report of the median, but because the resulting estimates appear reasonably Gaussian, means are good estimates of medians and so these are the central tendency indicators reported.

• Sample size (power): The adequacy of the sample size to detect a true failure of the null hypothesis is needed (Falmagne, 1978; Pratt 1964). To this end, we evaluate the adequacy of the sample of every test by a Monte Carlo simulation (suggested by, e.g., Mumby, 2002): A simulated sample consisted of drawing, without replacement, from the combined data set. The equality of medians was evaluated by means of the Mann-Whitney test. This process was repeated 1,000 times. If the distribution of the tests of the simulation agreed with actual test result, at the .05 level, the test's power was deemed adequate (see Steingrimsson & Luce, 2005a, and Steingrimsson 2009, for additional details)—in effect, this simulation favors rejection near the significance level.

• Effect Size: No generally accepted method exists for calculating the effect size for non-parametric tests. One approach is based on the observation that should two medians (means) differ by less than Weber's fraction, then the two are arguably not noticeably different to the observer (Dr. J. Yellott, personal communication, made this observation). Variability is well-known to be larger for magnitude productions than for matching and therefore certainly larger than for discrimination. A suitable Weber's fraction has not, to our knowledge, been worked out for magnitude/ratio productions and, in particular, not for colored stimuli. As a proxy, we convert the Weber constant to a proportion of the observed standard deviation in our non-magnitude production data and then use it as the proportion to set the cut-point in the magnitude production data. Teghtsoonian (1971) reports the Weber's fractions from reportedly conservative and independent studies to agree on approximately .08. Steingrimsson (2009) undertook the direct assessment of the fraction for an experimental situation similar to ours but for matching, which is less variable than magnitude production, and reported it to be .05. Thus, we regards .05 as a lower bound for Weber's fraction in the current situation. Applying this proportion in the above manner, we arrive at a fraction of ∼.08.[4] These bounds were used for evaluating effect size for all results.

[4]Note that even if we were to use the original, more restrictive .05, it would affect only two results, which did not alter the overall pattern of results.

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

Estimations are made using in discrete steps, namely in LUT (the monitor's video card **L**ook**U**p **T**able of integer values 0-255) values, but reported in cd/m^2. Because the estimates appear reasonably symmetrically distributed, the mean is known to be a good estimate for the median. The conversion from LUT to cd/m^2 involves a power transform, so we report the transformed mean LUT values as well as the normalized standard deviations (to maintain relative magnitude vis-a-vis the mean in cd/m^2—we thank Dr. J. Yellott for this suggestion).

In the current paper, we make several theoretical predictions, namely that some tests should hold, some should fail, and some that could go either way. The pattern of actual results is in line with these predictions as evaluated by the above outlined statistical criterion. Were it the case that our tests lacked power or sensitivity to detect true rejection, it would be nearly astronomically unlikely to get such a precise pattern of results. This fact alone gives us great confidence that our statistics are solid.

Procedure

Experiments were conducted in sessions of 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. The experiment was conducted in a dark room and each respondent received a minimum of 10 minutes of dark adaption. Information about the current block and trial number was displayed in the upper left corner of the screen.

With reference to Figure 2, Case 1, testing the same scale hypothesis consisted of presenting a standard x^f and, following the solid line, obtaining first an estimate of $x_p^{f,g}$, then using the estimate $x_p^{f,g}$ 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 $x_{p,q}^{f,g,g} = x_{q,p}^{f,g,g}$.

With reference to Figure 4, a magnitude production was carried out by presenting a reference stimulus of luminance x^f and a variable stimulus of luminance z^g placed above it, whose initial value is picked at random in a +/-20% cd/m^2 interval around a rough estimate of the likely final estimate (individual variations taken into account using training data). A proportion instruction, e.g., "The proportion is 150%" was printed in the upper left corner of the monitor. The respondent's task was to adjust the variable stimulus until it was perceived as being the given proportion of the standard stimulus.

The luminance of the variable stimulus could be increased or decreased, using key-presses, in one of four step-sizes, called "large", "medium", "small" and "extra small", corresponding to 8, 4, 2, and 1 values on the active RGB channel(s).

Following a key-press, the variable stimulus was redisplayed with the adjustment included. The re-displaying was signaled by setting the two squares to the (R, B, G) values of the background resulting in a just perceptible flicker.

Respondents were free to make as many and varied changes to the variable stimulus as desired. When they were satisfied with the magnitude production, they indicated it by another key-press, and the final value of z^g was taken as the estimate for the magnitude production. Following this, the program advanced to the next production.

Thus, to produce the estimate for $x_p^{f,g}$, the respondents adjusted the intensity of z^g until it appeared to be the prescribed proportion p of the standard stimulus x^f and the final value of z^g was taken to be an estimate of $x_p^{f,g}$. In the following step, the estimate for $x_p^{f,g}$ was placed in the role of a standard and the respondent adjusted a new z^g until it looked to be a proportion q of $x_p^{f,g}$. This final value was taken to be an estimate of $x_{p,q}^{f,g,g}$. An estimate for $x_{q,p}^{f,g,g}$ was obtained in an analogous fashion. Thus, arriving at estimates for $x_{p,q}^{f,g,g}$ and $x_{q,p}^{f,g,g}$ required four individual estimates and the typical number of individual estimates was 30. The statistical evaluation was of whether $x_{q,p}^{f,g,g} = x_{q,p}^{f,g,g}$, in which case the property is considered supported.

Design and condition

Although Proposition 1 (commutativity on a single dimension) has been evaluated (Steingrimsson, 2009), that evaluation involved achromatic stimuli only. Therefore, for completeness, we present a few tests of Proposition 1, Cases 1 and 2, for hued stimuli—there is no a priori reason to expect any difference in results from Steingrimsson (2009).

However, the main aim of this paper is to evaluate the same scale hypothesis in three ways

1. Test the three p, q relationships of Proposition 2.

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 3, 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 wavelengths, and also a large range of the visible wavelengths.

Listed in Table 1, are the 19 stimulus conditions that together address Proposition 1 and these 3 aims.

Aim 3 is addressed in various ways: magnitudes productions are made using a variety of standard with differing central wavelengths and over the range of wavelengths producible by our monitor.

Results

The detailed results are depicted in Figures 5, 6, 7, 8, and 9, and these results are summarized in Table 3.

The pattern of results in Table 3 is clear: The ratio scaling of brightness on a dominant wavelength, Proposition 1, shown by Steingrimsson (2009) to hold for achromatic stimuli, also holds for the tested chromatic stimuli. The same scale hypothesis of Proposition 2, (corresponding to Cases 1, 2, and 3 of Figure 2) is supported for Cases 1 and 2, whereas Case 3 is unambiguously rejected. The failure of Case 3 is evidence against the hypothesis that $\rho_+ = \rho_-$ for both wavelengths. For the test of Proposition 3 (corresponding to Figure

Condition	Proportions		Hue Standard (CIE-xyz)	Hue Variable
	$p\%$	$q\%$	f	g
Proposition 1, Cases 1 ($p > 1, q > 1$) & 2 ($p' < 1, q' < 1$)				
1	150	200	$(4.3, 8.3, 2.7)$ [Green]	Green
16*	150	200	$(6.6, 10.2, 12.8)$ [Green]	Green
17*	150	200	$(9.2, 11.1, 2.9)$ [Red-Green]	Red-Green
18*	75	50	$(173.2, 377.1, 61.1)$ [Green]	Green
19*	75	50	$(379.8, 468.8, 68.3)$ [Red-Green]	Red-Green
Proposition 2, Case 1: $p > 1, q > 1$				
2	200	300	$(9.6, 5.3, 1.7)$ [Red]	Green
3	150	200	$(11.7, 6.4, 1.9)$ [Red]	Green
4	200	300	$(21.3, 11.4, 2.2)$ [Red]	Green
5	200	300	$(7.0, 8.2, 2.7)$ [Red-Green]	Green
6	150	200	$(20.0, 10.2, 36.1)$[Red-Blue]	Green
7	150	200	$(20.0, 10.2, 36.1)$ [Red-Blue]	Red-Green
8	150	200	$(28.6, 1.1, 42.8)$ [R/G 20-B 40]	Green
16*	150	200	$(6.6, 10.2, 12.8)$ [Green-Blue]	Green
17*	150	200	$(9.2, 11.1, 2.9)$ [Red-Green]	Green-Blue
Proposition 2, Case 2: $p' < 1, q' < 1$				
9	70	50	$(257.4, 133.4, 12.6)$ [Red]	Red-Blue
10	75	50	$(241.3, 294.3, 43.0)$ [Red-Green]	Green-Blue
11	75	50	$(162.108.8, 10.0)$ [Red]	Blue
12	75	50	$(152.2, 79.0, 7.4)$ [Red]	Blue
13	70	50	$(223.6, 337.9, 449.8)$ [Red-Green-Blue]	Green-Blue
18*	75	50	$(173.2, 377.1, 61.1)$ [Green]	Green-Blue
19*	75	50	$(379.8, 468.0, 68.3)$ [Red-Green]	Green-Blue
Proposition 2, Case 3: $p > 1, q' < 1$				
14	150	75	$(97.5, 117.1, 17.5)$ [Red-Green]	Green-Blue
15	200	50	$(97.5, 117.1, 17.5)$ [Red-Green]	Green-Blue

Table 1: The Experimental Condition Under Which the Hypotheses of Proposition 1 and 2 were evaluated. For Proposition 1, Cases 1 and 2 of 1 are evaluated under 5 conditions, and For Proposition 2 , Cases 1, 2, and 3 of Figure 2 are evaluated under 10 conditions, for a total of 15 conditions. For each condition, the standard is indicated in CIE-xyz coordinates, but to facilitate the clarity of the actual hue the Red-Green-Blue color channel that dominates the standard and the user controlled variable are provided as well. The Conditions 16-19 are indicated by a *, to identify that these derive from the test of Proposition 3 (See Table 2).

Condition	Proportions		Hue Standard (CIE-xyz)	Hue Variable
	Proposition 3: Equality of reference points $\rho^{f,g} = \rho^{g,f}$			
16a,b	150	200	$(6.6, 10, 2, 12, 8)$ [Green]	Green-Blue
17a,b	150	200	$(9.2, 11.1, 2.8)$ [Red-Green]	Green-Blue
18a,b	75	50	$(173.2, 377.1, 61.1)$ [Green]	Green-Blue
19a,b	75	50	$(379.8, 468.0, 68.3)$ [Red-Green]	Green-Blue

Table 2: The Experimental Condition Under Which the Hypothesis of Proposition 3 was evaluated. Listed are the 4 main conditions, each is corresponding to Figure 3 wherefore each gives rise to two tests, a and b, for a total of 8 conditions with the corresponding standard, indicated in CIE-xyz coordinates. To facilitate the clarity which hues are used, the Red-Green-Blue color channel that dominates the standard are provided as well.

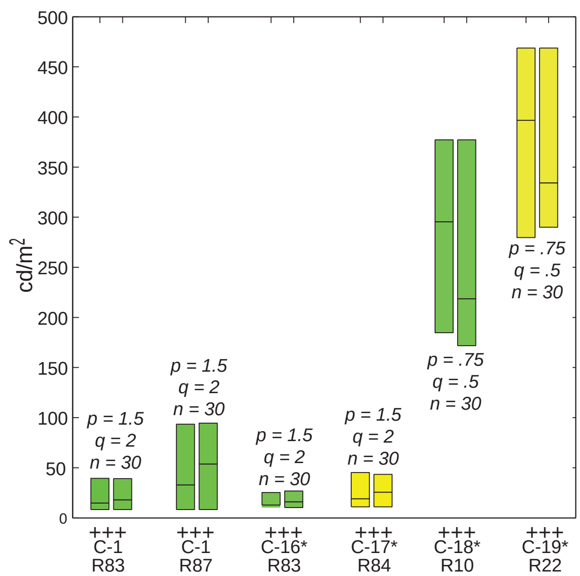

Figure 5. Depicted are the six tests of Proposition 1. In a group of two columns, the first column represents the results of magnitude production with p followed by q and the second column magnitude production with q followed by p. The results for each proportions is separated by a solid horizontal line in each column. Along with this information, listed are the values of the proportions, the number of observation for each test, and on the x-axis each + indicates, in order, the result for each of the three components of the statistical test. Three +'s are needed for the test to be taken as supported. That is in reference to Figure 1, Commutativity is said to hold if statistically in Case 1 if $x_{p,q} = x_{q,p}$, in Case 2 if $x_{p',q'} = x_{q',p'}$. No test of Case 3 is presented for Proposition 1.

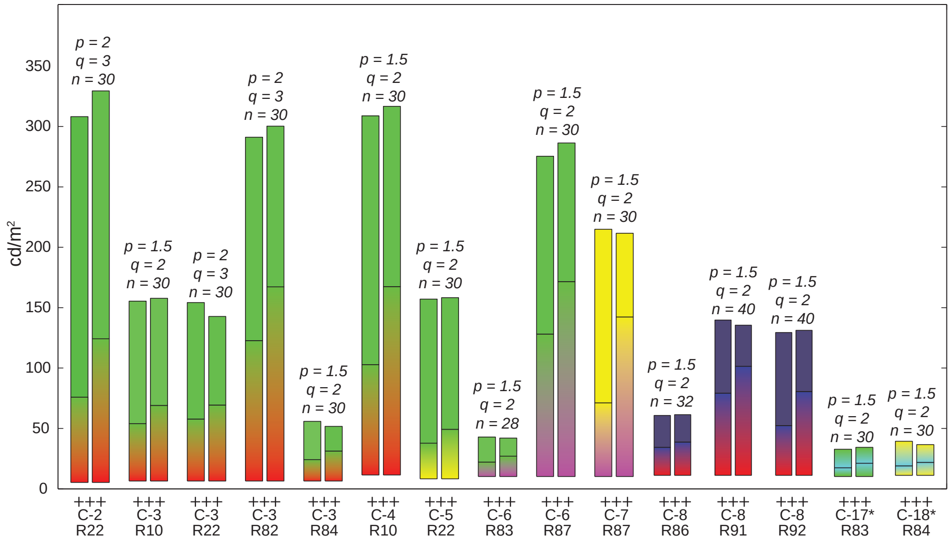

Figure 6. Depicted are the results for 15 tests of Case 1 of Proposition 2 ($p > 1$, $q > 1$). The columns and notation is the same as in Figure 5. However, since these are cross-frequency cases, the initial magnitude estimate is made from a standard of one hue and a variable one of a different hue. This is depicted by a gradient hue change from a hue that approximates the standard hue to the one approximating the variable hue. The statistical evaluation is of whether for Case 1, $x^{f,g,g}_{p,q} = x^{f,g,g}_{q,p}$.

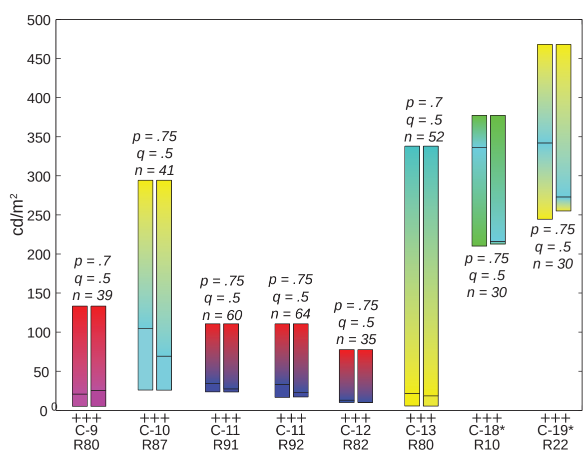

Figure 7. Depicted are the results for eight tests of Case 2 of Proposition 2 ($p' < 1$, $q' < 1$). The columns and notation is the same as in Figure 6. The statistical evaluation is of whether for Case 2, $x^{f,g,g}_{p',q'} = x^{f,g,g}_{q',p'}$.

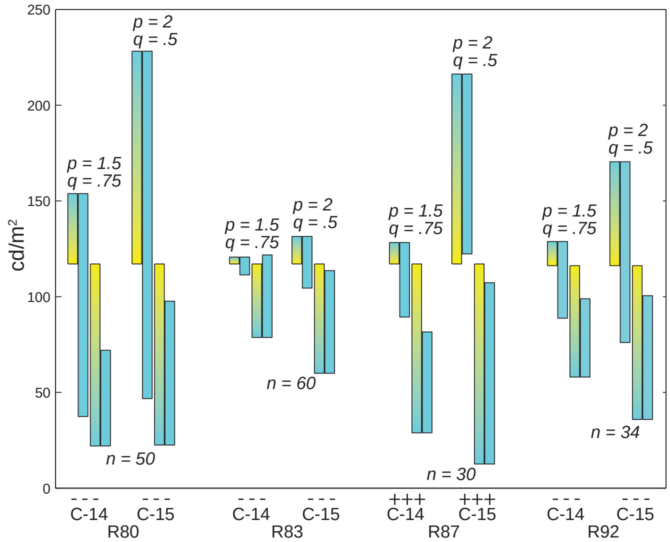

Figure 8. Depicted are the results for eight tests of Case 3 of Proposition 2 ($p' > 1$, $q' < 1$). The columns and notation is similar to the one in Figure 6. However, since in this mixed proportions case, the judgments do not yield monotonically increasing or decreasing values, each p and q judgment is represented separately. In the case of the first test, the first column depicts a production by $p > 1$, then the immediately adjacent column depicts a production by $q' < 1$ that has a staring point at the top of the first column. The second two columns represent the same but with the p and q judgments made in the reversed order. The statistical evaluation is of whether for Case 3, $x^{f,g,g}_{p,q'} = x^{f,g,g}_{q',p}$. In Case 3 both reference points are in play and commutativity is predicted only if $\rho_+ = \rho_-$.

Test of	#Tests	#Hold	#Fail	%Hold	Hypothesis
Prop. 1: $x_{p,q} = x_{q,p}$	6	6	0	100%	Supported
Prop. 2, Case 1: $p > 1, q > 1$	15	15	0	100%	Supported
Prop. 2, Case 2: $p' < 1, q' < 1$	8	8	0	100%	Supported
Prop. 2, Case 3: $p' < 1 < q$	8	2	6	25%	Rejected
Prop. 3: Equality of reference points $\rho^{f,g} = \rho^{g,f}$	8	0	8	0%	Rejected

Table 3: The table summarizes the testing of the commutativity hypotheses of Propositions 1, 2, and 3. 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.

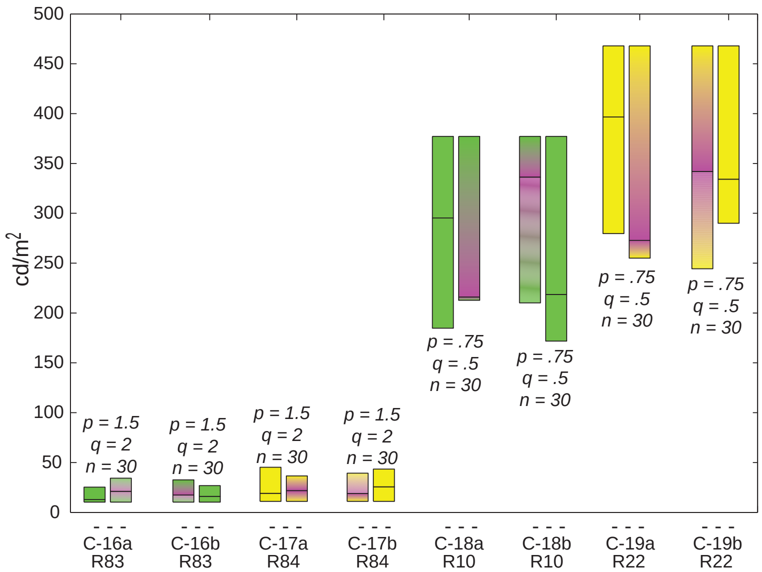

Figure 9. Depicted are the results for eight tests Proposition 3. The columns and notation is the same as in Figure 6. With reference to Figure 3, the comparison is between the judgments of the narrow lines (solid and dotted) are on a single frequency and the thicker lines (solid and dotted) which go across the two frequencies f, g. magnitude production starting with q followed by p. Commutativity in the single frequency case holds if statistically $x_{p,q} = x_{q,p}$, for the cross wavelength case if $x_{p,q}^{f,g,f} = x_{q,p}^{f,g,f}$, whose tests are depicted in Figs. 6 and 7. However, $\rho^{f,g} = \rho^{g,f}$ holds only if $x_{p,q} = x_{q,p} = x_{p,q}^{f,g,f} = x_{q,p}^{f,g,f}$, whose tests are depicted as in a ($x_{p,q} = x_{q,p}^{f,g,f}$) and b ($x_{p,q}^{f,g,f} = x_{q,p}$).

3), even though the data are somewhat mixed, they certainly do not support the general hypothesis that $\rho^{f,g} = \rho^{g,f}$.

Discussion and Conclusions

1. We began by summarizing how the commutativity property of intensive (prothetic) ratio scale, 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 or wavelength. Those results were the basis of the experimental program.

2. We replicated previous evidence for a ratio scale of brightness for achromatic stimuli using chromatic stimuli.

3. The empirical evidence supports the notion that for $p, q \geq 1$ and $p', q' < 1$, individuals rely on a single scale for brightness regardless of stimulus wavelength.

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

5. The evidence is consistent with different reference points for different wavelengths. Specifically $\rho^f \neq \rho^g$ as well as $\rho^{f,g} \neq \rho^{g,f}$.

6. We asked: Is brightness an intensity scale that is independent of hue? The data suggest the answer is yes as long as reference points are not assumed to remain fixed

independent of the wavelength and whether judged ratios are either all $p \geq 1$ or $p' < 1$.

These results open several paths for further exploration. Of course, it would be desirable to replicate the results and extend them to an even larger mix of stimuli and testing situations (e.g., to using reflected rather than emitted light). One issue that still escapes 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: With the result established for loudness (Luce et al., 2010) and brightness, do the same results hold for other intensive/prothetic continua, among which, in vision, are perceived contrast and saturation? The list can in principle extend to all domains that Stevens' (1975) identified as prothetic. For those domains where the cross-dimensional hypothesis is found to hold, a clear and intriguing next step is to extend the evaluation to cross-modal situations, e.g., to ask whether loudness and brightness may rely on a single scale of intensity. Should that turn out to be the case, the implication is that there may be a single notion of subjective intensity for a person—a somewhat sweeping idea. Large individual differences seem to rule out the possibility that individuals rely on the same-scale but that possibility has not been entirely excluded empirically.

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.

Ekman, G., Eisler, H., & Künnapas, T. (1960). Brightness of monochromatic light as measured by the method of magnitude production. *Acta Psychologica*, **17**, 392–397.

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.

Fechner, T. G. (1860/1966). Elemente der Psychopysics/Elements of Psychophysics. Vol. I. D. Howers & E. G. Boring (Eds.), H. E. Adler (Translation). Holt, Rinehart & Winston inc: New York.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of Measurement, Vol. I*, New York: Academic Press. (Reprinted 2007 by Dover Publications, Mineola, N. Y.)

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. (in preparation). Some predictions from Luce's global psychophysical theories.

Luce, R. D., Krantz, D. H., Suppes, P., & Tversky, A. (1990). *Foundations of Measurement, Vol. III*, New York: Academic Press. (Reprinted 2007 by Dover Publications, Mineola, N. Y.)

Luce, R. D., & Narens, L. (1985). Classification of concatenation structures by scale type. *Journal of Mathematical Psychology*, **29**, 1–72.

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.

Luce, R. D., Steingrimsson, R., & Narens, L. (2010). Are Psychophysical Scales of Intensities the Same or Different When Stimuli Vary on Other Dimensions? Theory with Experiments Varying Loudness and Pitch, *Psychological Review*, **117,** 1247–1258.

Marks, L. E., Szczesiul, R., & Ohlott, P. (1986). On the Cross-Modal Perception of Intensity. *Journal of Experimental Psychology: Human Perception and Performance*, **12**, 517–534.

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

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

Narens, L., & Luce, R. D. (1976). The algebra of measurement. *Journal of Pure and Applied Algebra*, **8**, l97–233.

Nordin, S. (1994). Context effects, reliability, and internal consistency of intermodal joint scaling, *Perception & Psychophysics*, **55**, 184–189.

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

Romney, A. K., Chia, C.-C. (2009). Functional computational model for optimal color coding. *Proceedings of the National Academy of Sciences*, **106**, 10376–10381.

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. *Attention, Perception, & Psychophysics.* Online First: DOI 10.3758/s13414-010-0067-5.

Steingrimsson, R. (in preparation a). Evaluating a Model of Global Psychophysical Judgments for Brightness: III. Forms for the Psychophysical and the Weighting Function.

Steingrimsson, R. (in preparation b). Evaluating a model of global psychophysical judgments for perceived contrast: I. Behavioral properties of summations and productions.

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

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

Steingrimsson, R., & Luce, R. D. (2006). Empirical evaluation of a model of global psychophysical judgments: III. A form for the psychophysical function and 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.

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.

Stevens, J. C. & Stevens, S. S. (1962). Brightness functions: Effects of adaptation. *Journal of the Optical Society of America*, **53**, 375–385.

Suppes, P., Krantz, D. H., Luce, R. D., & Tversky, A. (1989). *Foundations of Measurement, Vol. II*, New York: Academic Press. (Reprinted 2007 by Dover Publications, Mineola, N. Y.)

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

Ward. L. M. (1990). Cross-Modal Additive Conjoint Structures and Psychophysical Scale Convergence. *Journal of Experimental Psychology: General*, **119**, 161–175.

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.

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