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

RAGNAR STEINGRIMSSON, R. DUNCAN LUCE, and LOUIS NARENS
University of California, Irvine

Recent studies based on testable behavioral axioms have concluded that psychological scales of subjective intensive attributes involving the ears and eyes form ratio scales. These studies have shown that a certain commutativity of proportion property must hold under either successive increases or successive decreases, with all other independent dimensions fixed. However, until recently limited attention has been paid to whether such subjective intensity scales differ when a dimension independent of intensity, such as frequency or wavelength (e.g., pitch in audition, hue in vision) is varied. Using a simple and favorably tested theoretical model for global psychophysics, Luce, Steingrimsson, and 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 conditions.

Luce, Steingrimsson, and Narens (2010) formulated an axiomatic theory for an 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 the behavioral property of commutative proportion judgments, which is necessary and sufficient for there to be 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 of loudness independent of pitch. The present research aims to evaluate this property for perception of luminance–hue pairs. Favorable results invite subsequent work involving the commonality of scales across two distinct intensity modalities (e.g., loudness and brightness).

Luce et al. (2010) also provided some historical background and a complete report of the theory of commutativity of proportions, including proofs. Here we only summarize what is needed to make the present studies intelligible; for greater detail the reader should consult Luce et al. (2010).

Background

OUR APPROACH

Our approach, which mimics classic static physics, is based on a behavioral model composed of several behavioral (and testable) axiomatic invariances that

American Journal of Psychology
Fall 2012, Vol. 125, No. 3 pp. 321–333 •

imply a numerical representation—a measurement scale and its properties—of the assumptions (Luce, 2004, 2008, 2012). One of these properties is a form of magnitude production, which models the well-known method pioneered by S. S. Stevens (1975). Simply put, suppose that x is a (physical) luminance, and (x,x) denotes the joint presentation to the two eyes of a respondent, and suppose $p > 0$ is a number. The respondent is asked, using some variation on the method of adjustment, to produce the luminance x_p such that (x_p,x_p) appears to be p times as bright as (x,x).

PREVIOUS WORK

Our approach has its ideological roots in the work of Fechner, specifically what he called outer psychophysics (1860/1966), which entailed the view that psychophysics should, as had been true for physics, proceed from the physical stimulus and via a mathematical formalism map responses into a numerical measurement scale. However, the necessary tools have only recently matured to a point where the topic explored here could be placed on a sound theoretical basis. Therefore, it is not surprising that, despite more than a century of color research, a literature search reveals little material that is relevant to our approach. In particular, the interest in classic psychophysics in the spirit of Stevens, who used approaches based on magnitude production and magnitude estimation, appears to have waned in the late 20th century. Applying Luce's model to brightness, Steingrimsson (2009, 2011) established that brightness ordering is a ratio-scaled attribute, a conclusion that is consistent with numerous function fitting studies, summarized in S. S. Stevens (1975).

The studies undertaken by Stevens on luminance, most specifically Stevens and Stevens (1962), evaluated only achromatic stimuli. In fact, to our knowledge, only Ekman, Eisler, and Künnapas (1960) ever reported analogous data for chromatic stimuli. They addressed whether the same brightness function is satisfied when monochromatic wavelength is varied using both magnitude estimation and production methods. The results, pooled over respondents, were consistent with a common scale of intensity across all six frequencies, and by fitting power functions to average production data of 30 respondents, Ekman et al. concluded that the brightness function is approximately the same at hues ranging from 459 to 672 mμ. Their estimated exponents ranged from 0.256 to 0.300. The intensity variable that they used was physical intensity minus an estimated constant that appears to be fairly close to the absolute threshold. Their result is consistent with the studies of achromatic stimuli, summarized in S. S. 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 article report any citations of Ekman et al. (1960).

In contrast, our approach does not require function fitting but rather testing a theoretical commutativity property, which derives from our well-sustained representation of the method of magnitude production.

Given this dearth of applicable brightness studies, it is somewhat curious that the literature on the commonality of scales in cross-modal situations is a bit richer than for hues. Those studies appear to provide favorable evidence for the commonality hypothesis for several cross-modal situations. Among these studies are Marks, Szczesiul, and Ohlott (1986), who broadly supported the commonality notion for both auditory loudness versus vibratory touch and auditory loudness versus visual brightness. Working with auditory and visual stimuli, Ward (1990) reported similar results; 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.

Indeed, we conclude, with certain caveats, that a single ratio scale obtains for perceived luminance across different hues.

RATIO SCALABILITY

For a single attribute, Narens (1996) presented a model and showed that ratio scalability of intensity is equivalent to a commutative property that asserts that the order of successive productions does not matter provided that both numbers are greater than 1 or both are less than 1. More specifically, for brightness magnitude production, suppose that $x > 0$ is an arbitrary luminance measured above 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 is

$$x_{p,q} = x_{q,p}, \tag{1}$$

and it obtains if and only if both $p,q \geq 1$ or $p', q' < 1$ and brightness is a ratio scale. Because whether a proportion is greater or less than 1 plays an important role in this context, we have separated these case by adding a prime such that $p' < 1$ and $q' < 1$ but otherwise write $p \geq 1$ and $q \geq 1$.

Narens (1996) formulated another property implied by his model called multiplicativity. Both properties have been extensively evaluated. Equation 1 has consistently been found to hold, but multiplicativity has equally consistently been rejected in various domains: loudness (Ellermeier & Faulhammer, 2000; Steingrimsson & Luce, 2005a; Zimmer, 2005), size of circles (Augustin & Maier, 2008), brightness (Steingrimsson, 2009), and perceived contrast (Steingrimsson, 2012a). Steingrimsson and Luce (2007) formulated another version of multiplicativity, k-multiplicativity, which they found to hold, but specifically, they concluded that the key property underlying ratio scale measurement was just commutativity (1).

Building on Luce (2004), Narens (2006), and Steingrimsson and Luce (2007), Luce et al. (2010) developed a new kind of paradigm whose main conceptual novelty was to extend the commutative property to two stimulus attributes, leading to a common ratio scale representation of subjective intensity over the two attributes. 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 the same hypothesis for physical attributes of luminance and hue.

Summary of Luce et al.'s (2010) Theory

NOTATIONAL CONVENTION

The stimuli are the same luminance x presented simultaneously to the two eyes. So this pair is written (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 emphasize that the theory itself is cast in terms of intensity increments above the corresponding threshold, and so if τ_l and τ_r are threshold intensities in the left and the right eye, respectively, then the effective stimulus is (x_l,x_r), where $x_l = x - \tau_l$ and $x_r = x - \tau_r$. However, because all signals were well above threshold and the respondents were selected for normal vision, the error in reporting intensities x in candelas per square meter is negligible. So in all cases, because the same intensity is presented to each eye, we simplify the stimulus reporting to writing x to indicate the intensity of this stimulus.

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, 2012) 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 two parameters ρ_i, $i = l,r$ called reference signals, that together satisfy the constraint (we do not write the case for $p' < 1$ explicitly because it is completely analogous).

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

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 require that. Regrettably, we currently lack a theory for the reference signals.

The W representation has received substantial empirical support in loudness (Steingrimsson & Luce, 2005a, 2005b), achromatic brightness[1] (Steingrimsson, 2009, 2011), and perceived contrast (Steingrimsson, 2012a, 2012b).

Three Propositions of Luce et al. (2010)

The following three propositions, along with Figure 1, are intended to make clear the several cases being described.

SINGLE-FREQUENCY CASES

Proposition 1: Assuming that the general model (2) 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 if and only if $\rho_+ = \rho_-$.

Case 1 of the proposition is illustrated in Figure 1, Panel A; the other two cases are analogous.

CROSS-DIMENSION COMMUTATIVITY

Next, these results are extended to cases involving two hue dimensions, f and g. Superscripting is used to identify which dimension is being matched to which, and, as before, subscripting identifies which proportions are used. For instance, magnitude production by p staying in one dimension, f to f, is written $x_p^{f,f}$.

With successive productions changing dimension, the one from $f \to g$ is denoted $x_p^{f,g}$. 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 $x_{p,q}^{f,g,f}$. Of course, when the roles of p and q are reversed, the notation is changed accordingly.

We assume that the reference point chosen depends on whether $p \geq 1$ or $p' < 1$ and on whether it concerns the f or g dimension.

Proposition 2: Assuming that the general model (2) 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 $\rho_+^g = \rho_-^g$.

Case 1 of the proposition is illustrated in Panel B of Figure 1; the other two cases are analogous.

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 with a notation that makes clear that ρ^g need not be the same when going from f to g as when going from g to f; as we shall see, the data imply that this distinction is needed. We distinguish these as, respectively, $\rho^{f,g}$ and $\rho^{g,f}$. This comparison is depicted in Panel C of Figure 1.

Proposition 3: Assuming that the general model (2) holds, then

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

if and only if

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

Luce et al. (2010) evaluated all three propositions for loudness and pitch. In sum, for Propositions 1 and 2 they found support for Cases 1 and 2 but rejected Case 3, implying that $\rho_+ \neq \rho_-$.[2] Furthermore, they rejected Proposition 3, implying that $\rho^{f,g} \neq \rho^{g,f}$.

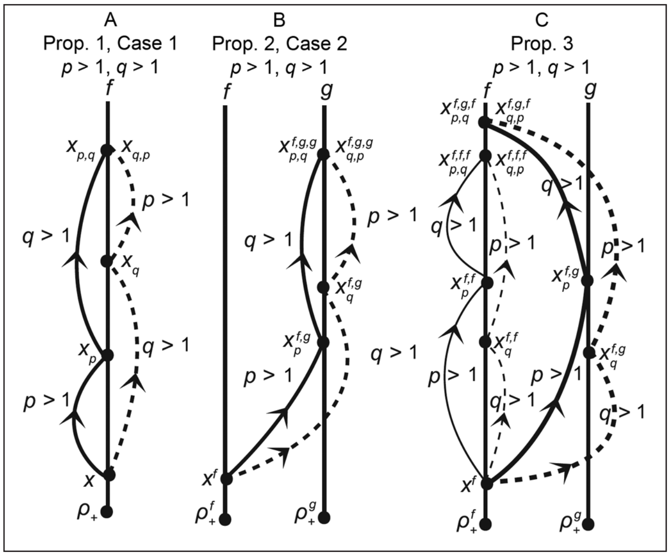

FIGURE 1. Cases 1 of Propositions 1 and 2 and the condition of Proposition 3. In each graph, the solid lines depict the case of magnitude production, with p followed by q, and the dotted ones depict the case of magnitude production, with q followed by p. In C, the narrow lines (solid and dotted) are on a single dimension and the thicker lines (solid and dotted) go across the two dimensions f and g. Commutativity is said to hold if statistically in A, $x_{p,q} = x_{q,p}$, and in B if $x_{p,q}^{f,g,g} = x_{q,p}^{f,g,g}$. In C, commutativity in the single-dimension case holds if statistically $x_{p,q} = x_{q,p}$, for the cross-frequency case if $x_{p,q}^{f,g,f} = x_{q,p}^{f,g,f}$, and $\rho^{f,g} = \rho^{g,f}$ holds only if $x_{p,q} = x_{q,p} = x_{p,q}^{f,g,f} = x_{q,p}^{f,g,f}$. (The figure summarizes Figures 1, 2, and 3 of Luce et al., 2010, with small modifications)

EXPERIMENT

METHOD

Respondents

A total of 10 respondents, of whom 8 were students from the University of California, Irvine, participated in the experiment reported (age range was 18–23, with one person aged 35 and one aged 43, 7 women, 3 men). All respondents reported normal or corrected-to-normal vision. All respondents, except the participating coauthor (R22), received compensation of $12 per session, no more than 1 hr long. Some participants had participated in some of our other experiments, so overall familiarity with similar tasks varied greatly. Each person provided written consent and was treated according to the Ethical Principles of Psychologists and Code of Conduct (American Psychological Association, 2002). Consent forms and procedures were approved by the University of California, Irvine's institutional review board.

Apparatus and Calibration

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 compensate for luminance fluctuations caused by ambient temperature and passage of time and a built-in gamma correction. The diagonal size is 56 cm, with a maximum resolution of 1,536 × 2,048 and luminance of 742 cd/m². Luminance measures were taken using Photo Research's PR-670 SpectraScan Spectroradiometer, which verified the monitor calibration and determined luminance values of stimuli.

When we write *f* for wavelength, for the monitor this is expressed as a triple (R, G, B), where each of R, G, and B is an 8-bit value (i.e., an integer from 0 to 255, or the monitor's video card lookup table [LUT] of integer values 0–255). We measured the intensity of each of the three guns and fit the result, as is customary in the vision literature, to a power function with an exponent γ to convert it to candles per square meter. The resulting function is traditionally called a gamma function.[3] The resulting exponents were $\gamma_R = 2.28$, $\gamma_G = 2.31$, and $\gamma_B = 2.14$. So although 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). The conversion from LUT to candles per square meter involves a power transform, so we report the transformed mean LUT values as well as the normalized standard deviations (to maintain relative magnitude vis-à-vis the mean in candles per square meter; we thank Dr. J. Yellott for this suggestion). For the purpose of standardization, the International Commission on Illumination (CIE) xyz measures are provided for all fixed stimuli (standards).

Stimuli

The basic stimuli consisted of two chromatic squares x^f (the standard) and z^g (the user-varied stimulus), where x and z are intensities and f and g are wavelengths, both subtending 10° of visual angle, presented on a uniform achromatic background of 4 cd/m², and arranged in vertical position with an 8° separation. A sample stimulus is depicted in Figure 2.

Statistical Method and Presentation of Results

The statistical method was identical to that used by Luce et al. (2010). We evaluated the same parameter-free null hypotheses of the generic form $L_{\text{side}} = R_{\text{side}}$. Because we did not know how individuals relate, all data analysis was done on the data for each individual separately (Luce, 1995, p. 20). The statistical approach consisted of three interlocking components.

C1. THE STATISTICAL TEST

We used the nonparametric Mann–Whitney U (M-W) test at a significance level of .05. Intensity adjustments were made using the discrete step sizes of LUTs, but because the obtained estimates seemed reasonably Gaussian, we report their means as an indicator of central tendency and standard deviations for variability.

C2. SAMPLE ADEQUACY

A Monte Carlo procedure was used to ascertain whether the sample size was sufficiently large to detect a true failure of the null hypothesis. The typical number of individual estimates was 30.

C3. EFFECT SIZE

No standard method for calculating effect size using the M-W exists, so instead we demanded that no two medians could differ by more than the applicable Weber's fraction. Steingrimsson (2009) concluded that the appropriate fraction for a compounding match-

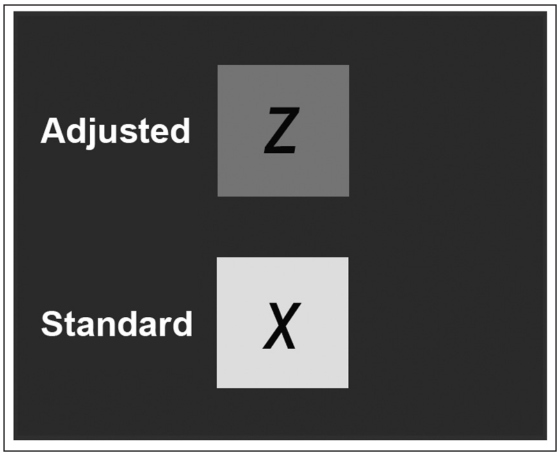

FIGURE 2. Stimulus consists of two hues (here illustrated monochromatically) displayed as squares. The lower one is the standard and the upper one is a variable one; the respondent is instructed to select it to be a proportion *p* as bright as the lower one. The labels in the figure are not present in the experiments. Not shown in the figure but indicated in the actual experiment is the block number as well as the proportion instructions (upper left corner of the screen)

ing task was between .05 and .08. Variability in ratio productions is generally substantially higher than for matching, so we used the upper bound of .08 as a conservative limit.

The result was taken as supportive of the hypothesis if and only if all three components agreed; otherwise, it was taken as not supported.

Procedure

The initial session involved obtaining written consent, explaining the task, and running practice trials. Each respondent trained for an additional session. Rest periods were encouraged, but the respondent controlled their frequency and duration. The experiment was conducted in a dark room, and each respondent received a minimum of 10 min of dark adaption. Information about the current block and trial number was displayed in the upper left corner of the screen.

As shown in Panel B of Figure 1, representing Case 1 of Proposition 2, we presented a standard x^f and, as shown by the solid arrow, we first sought an estimate of $x_p^{f,g}$; with that estimate as the next standard, following the second solid arrow obtained an estimate of $x_{p,q}^{f,g,g}$; and in the q, p order, the dotted arrows show the corresponding estimates $x_q^{f,g}$ then $x_{q,p}^{f,g,g}$. The statistical hypothesis is $x_{p,q}^{f,g,g} = x_{q,p}^{f,g,g}$.

In the test of Proposition 3, Panel C of Figure 1, the respondents dealt with judgments of first $x_q^{f,g}$ and then $x_{q,p}^{f,g,f}$. Procedurally, the process is analogous to the evaluation of Case 2 of Proposition 2. (Note that the test of $x_{p,q}^{f,g,f} = x_{q,p}^{f,g,f}$ happens also to be a valid test for Proposition 2, as are several other variation on the commutative theme.)

A magnitude production was carried out by presenting a fixed stimulus of luminance x^f and a variable stimulus of luminance z^g placed above it, whose initial value was picked at random in a ±20% interval around a rough estimate of the likely final estimate (individual variations were estimated using the training data). Figure 2 illustrates one instance of the stimulus setup. Not present in the illustration, a proportion instruction (e.g., "The proportion is 150%") was displayed 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 respondent increased or decreased the variable luminance using key presses in one of four step sizes, called large, medium, small, and extra small, which corresponded to 8, 4, 2, and 1 values on the active RGB channels.

Following a key press, the variable stimulus was redisplayed with the adjustment included. The redisplaying was signaled by setting the entire screen to the value of the background for one frame, resulting in a just perceptible flicker.

A respondent was free to make as many and varied changes to the variable stimulus as desired. When satisfied with the magnitude production, he or she indicated it by another key press, and the final value of z^g was taken as the estimate for the magnitude production. Then the program advanced to the next production.

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 study was to evaluate the cross-dimension commutativity in three ways:

1. Test the three p,q relationships of Proposition 2.
   Case 1: $p \geq 1$, $q \geq 1$ (Figure 1, Panel B).
   Case 2: $p' < 1$, $q' < 1$
   Case 3: $p' \geq 1 > q'$.
2. Test of Proposition 3, to examine whether the reference point is the same on any two frequencies, that is, $\rho^{f,g} = \rho^{g,f}$.
3. Test the conditions for several wavelengths and also a large range of the visible wavelengths. This we aimed to accomplish in various ways: Magnitude productions were made using a variety of standards with differing central wavelengths and over the range of wavelengths producible by our monitor.

Table 1 lists the 19 stimulus conditions that together addressed Proposition 1 and this three-pronged strategy.

Table 2 lists the four stimulus conditions under which Proposition 3 was evaluated. In the course of evaluating the proposition, four estimates were made: $x_{p,q}$, $x_{q,p}$, $x_{p,q}^{f,g,f}$, and $x_{q,p}^{f,g,f}$. These gave rise to two tests of the proposition's hypothesis: $x_{p,q} = x_{q,p}^{f,g,f}$ and $x_{q,p} = x_{p,q}^{f,g,f}$. However, these estimates also gave rise to one test of Proposition 1, $x_{p,q} = x_{q,p}$, and one test of Proposition 2, $x_{q,p}^{f,g,f} = x_{p,q}^{f,g,f}$.

RESULTS

The detailed results are depicted in Figures 3–7, and these results are summarized in Table 3.

TABLE 1. Experimental Conditions Under Which the Hypotheses of Propositions 1 and 2 Were Evaluated

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

Note. *For Proposition 1, Cases 1 (1, Panel A) and 2 are evaluated under 5 conditions, and for Proposition 2, Cases 1 (Figure 1, Panel B), 2, and 3 are evaluated under 9, 7, and 2 conditions, respectively, for a total of 18 conditions. For each condition, the standard is indicated in International Commission on Illumination (CIE) xyz coordinates, and the color gun (red [R], green [G], and blue [B]) that provided the largest contribution to the hue is indicated as well. Conditions 16-19 are marked by asterisks to indicate that they derive from the test of Proposition 3 (see Table 2).*

These figures share a similar structure and information.

Axes

The y-axis lists luminance in candles per square meter as a function of the individual tests on the x-axis. On the x-axis, the axis label for each individual test provides three lines of information. The first line indicates the results for each of the three components of the statistical test criterion (in the order they are listed under *Statistical Method and Presentation of Results*). The "+" symbol indicates the test was passed and the "–" that it failed. All three tests must be passed (+++) for the evidence to be taken as support of the hypothesis. The second line indicates the condition of the test, whose num-

TABLE 2. Experimental Conditions Under Which the Hypothesis of Proposition 3 Was Evaluated

Condition	Proportion p (%)	q (%)	Hue standard (CIE xyz) f	Hue variable g
Proposition 3: Equality of reference points $\rho^{f,g} = \rho^{g,f}$				
16a,b	150	200	(6.6,10,2,12,8) G	G-B
17a,b	150	200	(9.2,11.1,2.8) R-G	G-B
18a,b	75	50	(173.2,377.1,61.1) G	G-B
19a,b	75	50	(379.8,468.0,68.3) R-G	G-B

Note. *The table format is the same as in Table 1. The four estimates needed are depicted in Panel C of Figure 1, which give rise to the two tests, a and b, $x_{p,q} = x^{f,g,f}_{q,p}$ and $x_{q,p} = x^{f,g,f}_{p,q}$, for a total of 8 conditions. The hue of the standard is indicated in International Commission on Illumination (CIE) xyz coordinates, and the color gun (red [R], green [G], and blue [B]) that provided the largest contribution to the hue is indicated as well.*

TABLE 3. Testing of the Commutativity Hypotheses of Propositions 1, 2, and 3

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

Note. *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.*

ber refers to the details listed in Tables 1 and 2. The third and final line is an identifier for the respondent performing the test.

Bars

With reference to the first two bars of Figure 4, the left bar indicates the results of production by p and then by q, and the right bar indicates the results of production by q and then p. The values of p and q are indicated at the top of the bars along with the number n of observations made for each sequence. Both bars start at the luminance (y-axis) value of the first standard, x^f; the next horizontal line is the luminance value of the productions for, respectively, $x^{f,g}_p$ and $x^{f,g}_q$. The tops of the bars indicate the averaged luminance of the averaged productions for, respectively, $x^{f,g,g}_{p,q}$ and $x^{f,g,g}_{q,p}$. The statistical test is of whether $x^{f,g,g}_{p,q} = x^{f,g,g}_{q,p}$ (but in a nonparametric way), so the test is an indication of whether the height of the bars is statistically indistinguishable.

Colors

With reference to the first bar of Figure 4, indicated by the red color at the bottom of the bar is actual approximation of the color of the standard x^f (i.e., f). The color of $x^{f,g}_p$ (i.e., the g dimension) was green. That the production was from a red color f to a green color g is shown by a gradient change from the red to the green. The second production was from g to g, and so the corresponding part of the bar is a solid green. The exact colors are indicated in Tables 1 and 2.

Variations

In the case of $p' < 1$ and $q' < 1$, the bars read from the top down. In the case of $p' < 1$ and $q \geq 1$ (Figure 6), the first and the second columns are each broken up in two parts where one is read from the top and down and the other from the bottom to the top.

The pattern of results in Table 3 is clear: The ratio scaling of brightness on a dominant wavelength,

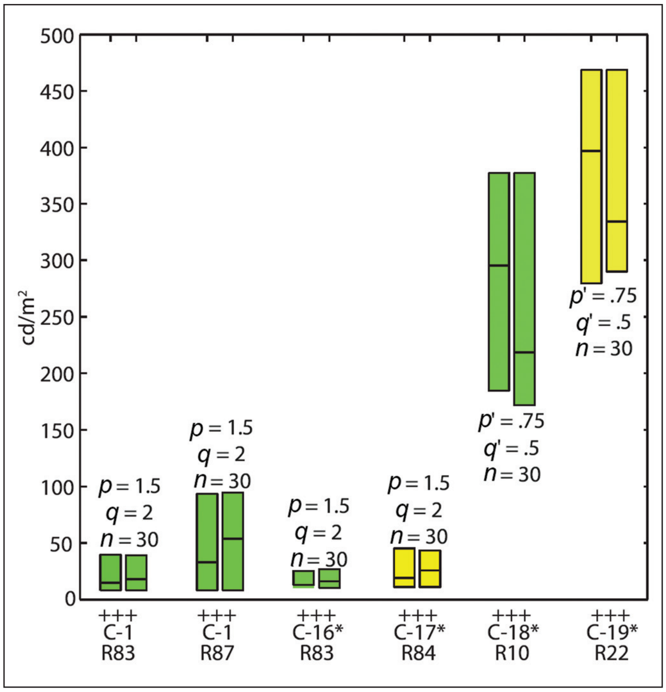

FIGURE 3. The 6 tests of Proposition 1 (see text for description). The tests were of Cases 1 (Figure 1, Panel A) and 2 of the proposition, and all were found to hold. No test of Case 3 is presented for the proposition. Asterisks indicate conditions stemming from the evaluation of Proposition 3

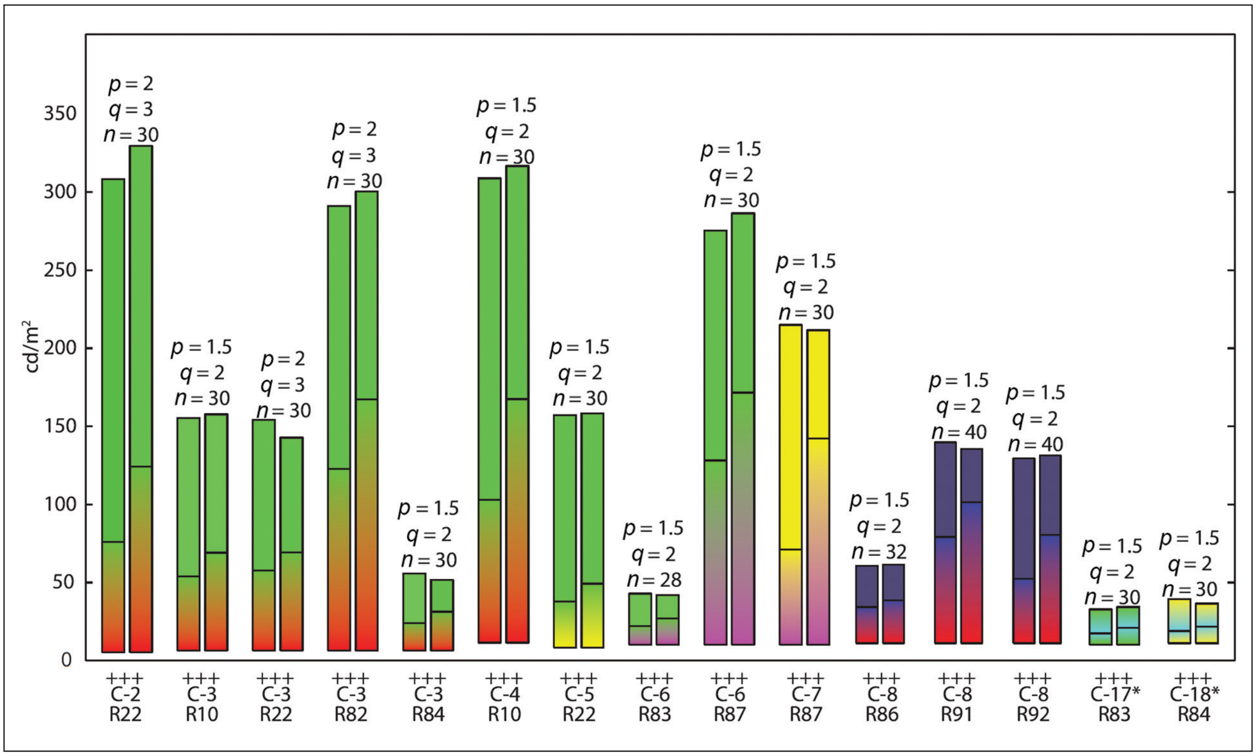

FIGURE 4. Results for 15 tests of Case 1 (Figure 1, Panel B) of Proposition 2 (see text for description); the tests of Cases 2 and 3 are reported in subsequent figures. In 15/15 cases, Case 1 of the proposition was found to be supported. Asterisks indicate conditions stemming from the evaluation of Proposition 3

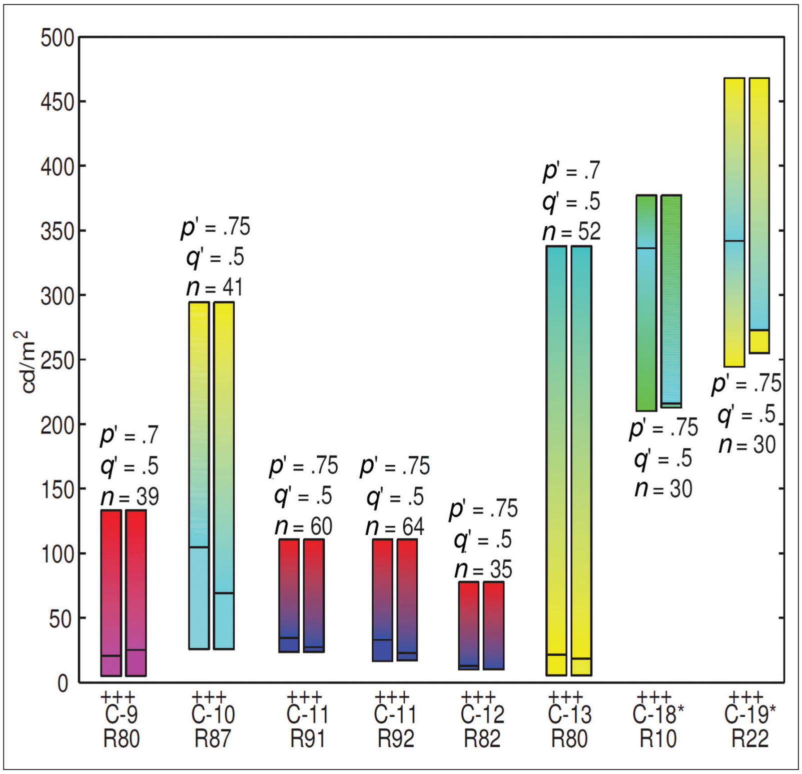

FIGURE 5. Results for 8 tests of Case 2 of Proposition 2 (see text for description). In 8/8 cases, Case 2 of the proposition was found to be supported. Asterisks indicate conditions stemming from the evaluation of Proposition 3

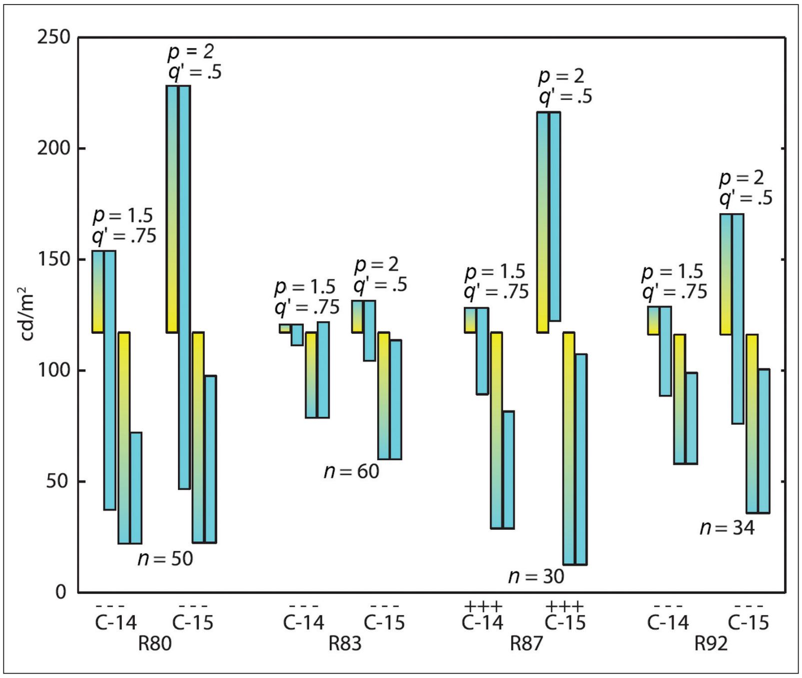

FIGURE 6. Results for 8 tests of Case 3 of Proposition 2 (see text for description). In 2/8 cases, Case 3 of the proposition was found to be supported. In Case 3 both reference points are in play, and commutativity is predicted only if $\rho_+ = \rho_-$

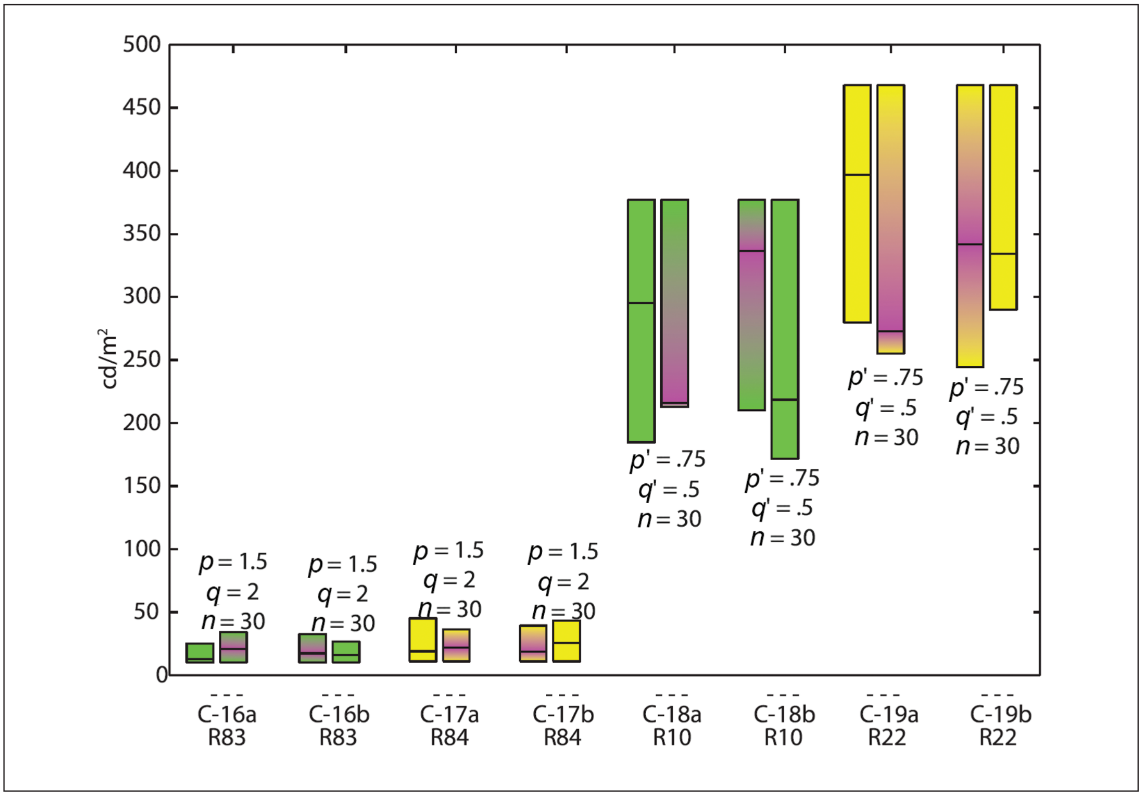

FIGURE 7. Results for 8 tests of Proposition 3 (see text for description). In 0/8 cases, the proposition was found to be supported. The proposition is depicted graphically in Figure 1, Panel C. Here both reference points are in play, and $\rho^{f,g} = \rho^{g,f}$ holds only if $x_{p,q} = x_{q,p} = x^{f,g,f}_{p,q} = x^{f,g,f}_{q,p}$, whose tests are depicted as in $x_{p,q} = x^{f,g,f}_{q,p}$ and $x^{f,g,f}_{p,q} = x_{q,p}$.

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 Panel B, Figure 1) 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 Panel 3, Figure 1), 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 the 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. The experimental program was based on those theoretical results.
2. We replicated previous evidence for a ratio scale of brightness for achromatic stimuli using chromatic stimuli.
3. The empirical evidence supports commutativity for $p,q \geq 1$ and for $p',q' < 1$, and the currently reported data indicate that individuals rely on a single scale for brightness regardless of stimulus wavelength. Our conclusion is much the same, but more detailed, as the claim of Ekman et al. (1960) based on comparing fitted curves.
4. The evidence implies that reference signals for $p' < 1$ and for $p \geq 1$ do differ.
5. The evidence implies that the reference signals do differ 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 that it is as long as reference signals 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., use of reflected rather than emitted light). One issue that has escaped us all along is a principled theory of reference signals. Such signals are clearly important in our data, but although we treat them as parameters (to be estimated), we can say a great deal about them (see, e.g., Items 4 and 5). A second question is, With the common ratio scale result established for loudness (Luce et al., 2010) and now for 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 S. S. Stevens (1975) identified as prothetic. In this connection, Luce (2012) has established an important sense in which these other modalities are richer in their potential for individual differences than for the ears and eyes. In each case, there are three inherently different types of psychophysical functions. For example, in the case of utility for money, these types correspond to risk-seeking, risk-neutral, and risk-averse utility functions. This greatly complicates the predictions for cross-modal matching.

For 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 subjective scale of intensity). If that turns 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 yet been entirely excluded empirically.

NOTES

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 article. Furthermore, we thank our reviewer Dr. H. Zhang and two anonymous reviewers for their comments, which substantially improved our presentation.

Address correspondence about this article to Ragnar Steingrimsson, Institute for Mathematical Behavioral Science, University of California, Irvine, Irvine, CA 92697-5100 (e-mail: ragnar@uci.edu).

1. There is nothing that suggests that the results would differ in the case of chromatic stimuli. Indeed, here we report a successful partial replication of previous experiments using chromatic stimuli; see Figure 3 and Item 2 of *Discussion and Conclusions*.

2. For Proposition 1 the evidence includes data summarized in Figure 1 of Steingrimsson and Luce (2006), some

data reported in Appendix E of Steingrimsson and Luce (2007), and systematic study of it by Steingrimsson (2012a).

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

REFERENCES

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

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 & Psychophysics, 62,* 1505–1511.

Fechner, T. G. (1966). *Elemente der psychophysics* [Elements of psychophysics] (Vol. 1). D. Howers & E. G. Boring (Eds.), H. E. Adler (Trans.). New York, NY: Holt, Rinehart & Winston. (Original work published 1860)

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

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. (2012). Predictions about bisymmetry and cross-modal matches from global theories of subjective intensities. *Psychological Review, 119*, 373–87.

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.

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

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. (2011). Evaluating a model of global psychophysical judgments for brightness: II. Behavioral properties linking summations and productions. *Attention, Perception & Psychophysics, 73,* 872–885.

Steingrimsson, R. (2012a). *Evaluating a model of global psychophysical judgments for perceived contrast: I. Behavioral properties of summations and productions.* Manuscript in preparation for publication.

Steingrimsson, R. (2012b). *Evaluating a model of global psychophysical judgments for perceived contrast: II. Behavioral properties linking summations and productions.* Manuscript submitted for publication.

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

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

Ward, L. M. (1990). Cross-modal additive conjoint structures and psychophysical scale convergence. *Journal of Experimental Psychology: General, 119,* 161–175.

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