Evaluating a model of global psychophysical judgments for brightness: II. Behavioral properties linking summations and productions

Ragnar Steingrímsson

Published online: 24 December 2010

© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract Steingrímsson (*Attention, Perception, & Psychophysics*, 71, 1916–1930, 2009) outlined Luce’s (*Psychological Review*, 109, 520–532 2002, 111, 446–454 2004) proposed psychophysical theory and tested, for brightness, behavioral properties that, separately, gave rise to two psychophysical functions, $\Psi_{\oplus}$ and $\Psi_{\circ_p}$. The function $\Psi_{\oplus}$ maps pairs of physical intensities onto positive real numbers and represents subjective summation, and the function $\Psi_{\circ_p}$ represents a form of ratio production. This article, the second in a series expected to consist of three articles, tests the properties linking summation and production such that it forces $\Psi_{\circ_p} = \Psi_{\oplus} = \Psi$. The properties tested are a form of distributivity and, in three experiments, were subjected to an empirical evaluation. Considerable support is provided for the existence of a single function Ψ for both summation and ratio production. The scope of this series of articles is to establish the theory as a descriptive model of binocular brightness perception.

Keywords Binocular brightness · Brightness summation · Ratio production · Magnitude production · Magnitude estimation · Psychophysics · Matching · Bisymmetry · Joint-presentation decomposition · Distributivity

Luce (2002, 2004, 2008 [erratum]) has proposed a theory of global psychophysics.¹ This article is the second (Steingrímsson, 2009, being the first) in a series of articles whose aim is to apply this theory to binocular brightness perception and, in so doing, establish it as a descriptive model of that domain. In a parallel series of four articles, Steingrímsson and Luce (2005a, 2005b, 2006, 2007; Luce & Steingrímsson, 2008 [erratum]) evaluated the theory for loudness and provided broad support for the theory in that domain. Steingrímsson (2009) paralleled the work of Steingrímsson and Luce (2005a) for loudness, but for brightness. This work provided evidence for a model of binocular brightness (cyclopic image) in which the percept of physical stimuli arriving at the two eyes could be described by a *summation representation* (Eq. 3) and the respondents’ distortion of numbers in magnitude/ratio productions by a *proportion representation* (Eq. 4)—both are presented and discussed below.

A psychophysical function Ψ of the subjective intensity of brightness arises in both representations, but as far as the results of Steingrímsson (2009) went, the psychophysical functions of brightness intensity arising in Eqs. (3) and (4) were not guaranteed to be the same function. The topic of the present article is to evaluate necessary and sufficient behavioral properties (axioms) to link the summation and production representations by showing that the same function arises in both representations. In form and content, the article parallels the loudness article of Steingrímsson and Luce (2005b).

R. Steingrímsson (✉)
Institute for Mathematical Behavioral Sciences,
University of California,
Irvine, CA 92697-5100, USA
e-mail: ragnar@uci.edu

¹ Global psychophysics deals with stimuli that can be discriminated correctly with near certainty.

Steingrímsson (2009) discussed the background and underlying theory. To avoid unnecessary repetition, I will refer to the sections of that article, as applicable. However, for this article to be self-contained, a brief summary of the relevant background and theory is provided.

As an article in a series, the results reported here deal with a portion of a larger whole. To see how the results reported here fit into the larger picture, it is instructive to recall that the psychophysical function is a formal model that maps physical stimuli onto sensations. The linking of the two representations (Eqs. 3 and 4) is equivalent to stating that this description, the psychophysical function, does not depend on the operation the respondent is asked to perform. This is critical because a conceptual novelty of Luce's (2002, 2004) theory is the harnessing of the power of two operations to arrive at a complete model. This linking, typical in physics, has not been a feature of formal psychological modeling so the data and the question are novel (Fig. 3, as a summary, displays this graphically).

As we shall see, the results reported favor the single psychophysical function and, thereby, the model. This result, along with the auditory work (Steingrímsson & Luce, 2005a, 2005b), places the work into an even larger context, in that the same model is found to capture behavior in two domains that have traditionally mostly been modeled and studied separately. The unification of the two domains occurs on the level of description and, as such, suggests that many psychophysical phenomena can, as is the case in physics, be described using a common language and on the basis of common primitives. Indeed, as is mentioned in the General Discussion section, this fact has already shown itself useful for addressing several outstanding issues.

The paper is organized as follows:

- Relevant theory and its interpretation in brightness is summarized.
- Results from three experiments, which together link the two representations of the theory, are reported.
- A comprehensive summary of the present article in the context of the series' first article is provided, along with direction of subsequent research (see Fig. 3 for a quick overview).

Theoretical background and the theory's application to binocular brightness perception

The theoretical approach mimics the classical, static physical approach to measurement as described in the *Foundations of Measurement* (Vols. I and III; Krantz, Luce, Suppes, & Tversky, 1971; Luce, Krantz, Suppes, & Tversky, 1990) and characterizes most of physics to this day.

This means that we start by evaluating testable behavioral invariances—axioms—on the basis of which larger consequences can be derived. More specifically, the theory is one of axiomatic psychophysics in which the aim is to map a sensation elicited by a physical stimulus onto a numerical structure. Just as in physics, it is therefore not necessary to consider the mechanisms that give rise to the sensations being studied in order to achieve a complete behavioral description of them.

This approach was dominant in the early days of psychology, with one of the earliest results being, for example, the famous Fechner–Weber law. In recent years, psychology has increasingly focused on attempting to answer questions by, in effect, reverse-engineering the mechanism (the nervous system) that gives rise to sensations. Perhaps for that reason, the classical psychophysical approach is less familiar to many psychologists. For this reason, Steingrímsson (2009)—the first article in this series—addressed in some detail, in Appendices A–C, the axiomatic approach, as well as the specific theory's relation to other modeling efforts of binocular brightness.

The experimental setting involves a computer monitor, with respondents seated in a dark room altering, according to instructions, the luminance of achromatic stimuli (squares on a background) displayed on the monitor. In some cases, a stereoscope is used to help generate the stimuli (see Fig. 1 for a preview).

The luminance of the physical signal is taken to correspond to the sensation of brightness, and the change in luminance is experienced as a change in brightness—for example, the change that results from varying the luminance setting on a computer monitor. This understanding accords with that of numerous articles (see the reviews in Ding & Sperling, 2006; Grossberg & Kelly, 1999).

Fig. 1 Stimuli displayed on a monitor (A) and viewed through a stereoscope (B) produce the subjective percept seen by the respondents (C). The x , u , and z values are luminance. (Reprinted with permission from Steingrímsson, 2009)

Primitives

Luce's (2002, 2004, 2008) theory, as is typical of axiomatic theories, is non-domain-specific but, when applied to binocular brightness, as it is here, becomes a descriptive theory for that domain. The theory features three primitives. Applying the theory to brightness involves specifying and interpreting these primitives for binocular brightness. If empirically supported, it provides a reasonable description of the data, including having some predictive power. This aspect of the approach is discussed in some detail in Appendix B of Steingrímsson (2009), as is its interpretation for brightness (see the introduction). Thus, here I provide only a summary of this material.

Joint presentations, ordering, and matching Let x and u correspond to physical intensities—for example, the luminance of two achromatic squares on a monitor; then, the ordered pair (x, u) is a stimulus that means that a luminance of intensity x and one of intensity u are presented jointly (simultaneously), such that x is presented to the left eye and u to the right eye. This first primitive is called a joint presentation.

The ordered pairs have the property that $(x, u) \succsim (y, v)$, which means that the stimulus (x, u) is judged to be at least as bright as (y, v) . This *ordering*, $\succsim$, is the *second primitive*. The indifference relation $\sim$ is defined by: $(x, u) \sim (y, v)$ if and only if both $(x, u) \succsim (y, v)$ and $(y, v) \succsim (x, u)$. Note that $\sim$ and $\succsim$, are used, rather than $=$ and $\geq$, because the latter refer to ordering of real numbers and the former to psychological judgments, but the notational similarity suggests that $\succsim$ is behaving similarly to the ordering $\geq$ of the real numbers.

The ordering is assumed to agree with physical intensity, a feature shown to hold in all but the most extreme viewing conditions (produced under laboratory conditions); this phenomenon is called *Fechner's paradox*. The paradox is that if the intensity in one eye is 10%–15% or less of the intensity in the other, lowering that intensity further results in increased perception of brightness (Fechner, 1861; Levelt, 1965). Theoretically, testing is directed to what Luce (2004) called the *symmetric* version of the theory—that is, the case in which some light is assumed to reach both eyes in binocular viewing conditions, and outside of Fechner's paradox, this is rather reasonable for regular viewing, considering the 70%–80% overlap in the two eyes' visual fields. Steingrímsson (2009) devoted an entire section ("Fechner's Paradox") to this matter, so it will not be discussed further here.

Matching the perceived magnitude of a standard to another percept is a classic method in psychophysics (see Steingrímsson, 2009, Appendix A; see, e.g., Stevens, 1975, for a comprehensive discussion). Here, this method is used

and formalized as follows. Each joint presentation (x, u) can be matched perceptually by (z, z) —that is,

$$(z, z) \sim (x, u). \quad (1)$$

This is referred to as a *symmetric match*, but operationally, it will be referred to as *brightness matching* or just *matching*. It is sometimes useful to use a mathematical operator notation to specify matching:

$$z := x \oplus u, \quad (2)$$

where z is defined by Eq. (1) and the notation $A := B$ means that A is defined by B . Technically, $\oplus$ is a binary mathematical operator and is referred to as the *summation operator*.

Note that (x, u) refers to a joint presentation of a signal pair—an actual stimulus—whereas $x \oplus u$ is a mathematical expression that stands for, here, the resulting cyclopic image, equivalent to that of intensity z . The cyclopic image/percept is a unitary percept (our everyday binocular view) that results from the combining of two separate input signals—hence, referred to as subjective *summation*—of the two signals. Since the theory is descriptive, nothing is or needs to be hypothesized about how this summation is biologically accomplished.

Ratio production The third primitive is a generalization of magnitude production. Magnitude production was popularized by Stevens in the mid-20th century and subsequently became a standard psychophysical method (see Steingrímsson, 2009, Appendix A; see Stevens, 1975, for a comprehensive discussion). In magnitude production, the respondent produces a stimulus that is some prescribed proportion of a standard—for example, two times as bright as a standard.

In *ratio production*, suppose that $x > y \geq 0$ are intensities, and let $p > 0$ be a real number. Let (z, z) denote a signal that the respondent judges to make the brightness "interval"² from (y, y) to (z, z) stand in the ratio p to the brightness interval from (y, y) to (x, x) . As in matching Eq. (2), it is convenient to write the operation as a mathematical operator having the form: $(z, z) := (x, x) \circ_p (y, y)$, or as $z = x \circ_p y$ in shorthand (the subscript s means symmetric production). The *generalization* part of the ratio production can be seen by the fact that it agrees with *magnitude production* when $(y, y) = (0, 0)$.

Notational convention Let ε_l and ε_r denote thresholds for the left and the right eyes, respectively, and let x' and u' be intensity arriving at the left and the right eyes, respectively; then the notation is $x = x' - \varepsilon_l$ and $u = u' - \varepsilon_r$. Thus $x = 0$ denotes the threshold intensity (or less) of the left eye stimulus, and $u = 0$ denotes the same for the right eye. For

² The term *interval* is used figuratively to refer to the difference in brightness that respondents experience between two intensity pairs.

stimuli well above threshold, which are used, the difference $x - x'$ is negligible.

Representations of $\oplus$ and $\circ_p$

Luce (2002, 2004) formulated a set of necessary and sufficient (and testable) behavioral properties, formulated in terms of the presented primitives, that allowed him to construct a psychophysical function Ψ that is a strictly monotonic mapping of intensity pairs to the nonnegative real numbers that preserves the order $\succsim$, and for which there exists a constant $\delta = 0$ or 1, such that

$$\Psi(x, u) = \Psi(x, 0) + \Psi(0, u) + \delta \Psi(x, 0) \Psi(0, u), \quad (3)$$

and a strictly increasing, numerical distortion function, $W(p)$, from a nonnegative number p onto itself, such that

$$W(p) = \frac{\Psi[(x, u) \circ_p (y, v)] - \Psi(y, v)}{\Psi(x, u) - \Psi(y, v)}, [(x, u) \succ (y, v) \succeq (0, 0)]. \quad (4)$$

The representation (Eq. 3) captures the combining of inputs to the left and right eyes, respectively, and is referred to as the *summation representation* [also known as a *p (polynomial)-additive representation* in the literature]. The representation (Eq. 4) describes the ratio production operation and is referred to as the *production representation*. Mapping subjective inputs onto mathematical expressions allows us to bring to bear the toolbox of mathematics on unobservable subjective entities (see the [General Discussion](#) section for examples of other recent applications).

The representations consist of two unspecified functions, Ψ and W , and the constant δ . This allows for enormous freedom for capturing individual differences, as well as extending these representations to different domains. The key is that as long as certain parameter-free behavioral properties are satisfied, the functions and the representations are guaranteed to exist and have that status without the need for any fitting of data to functions (see Steingrimsón, 2009, Appendix B, for details).

Steingrimsón and Luce (2006, 2007) addressed the question of the forms of the unknown functions for loudness, and for the psychophysical function, Ψ , and the weighting function, W , respectively, they found support for $\Psi(x, 0)$ and $\Psi(x, u)$ being power functions and for W being a power or a Prelec function (Steingrimsón & Luce, 2007). Parallel work is ongoing for brightness.

What is known from Steingrimsón (2009)

Steingrimsón (2009) evaluated three behavioral axioms: joint-presentation symmetry, which was rejected, meaning that the two eyes cannot be taken to be behaviorally identical; the

Thomsen condition (a special case of double cancelation), which was supported and established additivity over the two eyes and, along with certain reasonable background conditions, established the summation representations (Eq. 3); and finally, production commutativity, which established the production representation (Eq. 4). Together, these results provided evidence for Eqs. (3) and (4) holding, but separately. That is, in reality there are two distinct psychophysical functions: $\Psi_{\oplus}$ for summations and $\Psi_{\circ_p}$ for productions. At present, it is not known whether $\Psi_{\oplus} = \Psi_{\circ_p} = \Psi$. That question is the topic of the present article as we shall see, we are warranted in writing just Ψ in Eqs. (3) and (4).

Behavioral properties linking summations and productions

The behavioral properties are presented using the mathematical operator notations. To test them requires transforming the mathematical expressions into actual stimuli, of which there may be many; thus, this presentation preserves generality.

Bisymmetry Luce (2004) initially held that under the assumptions of the theory, $\delta = 0$ was equivalent to the property of bisymmetry:

$$(x \oplus y) \oplus (u \oplus v) = (x \oplus u) \oplus (y \oplus v). \quad (5)$$

However, in a correction reported by Luce (2008), bisymmetry is shown to simply be a direct prediction of the summation representation (Eq. 3). Therefore its holding pertains to that representation only and, thus, should logically have been evaluated by Steingrimsón (2009). Because this result became clear only well into the review process of that article, its evaluation is presented here.

Simple joint-presentation decomposition One of the two key linking properties is simple joint-presentation decomposition. Using the shorthand $x \circ_p y = (x, x) \circ_p (y, y)$, then, for positive real numbers p ,

$$(x \oplus u) \circ_{p_s} 0 = (x \circ_{p_s} 0) \oplus (u \circ_{p_s} 0). \quad (6)$$

By replacing $\oplus$ and $\circ_p$ with $+$ and $\times$, it is easy to see the arithmetic parallel to Eq. (6).

Distributivity A second key linking property for the representations of Eqs. (3) and (4) is the property of distributivity. Distributivity comes in two forms:

Left distributivity:

$$z \oplus (x \circ_{p_s} u) = (z \oplus x) \circ_p (z \oplus u) (u > 0). \quad (7)$$

Right distributivity:

$$(x \circ_{p_s} u) \oplus z = (x \oplus z) \circ_p (u \oplus z) (u > 0). \quad (8)$$

The difference between the two forms of distributivity captures the fact that if the input order to the left and right eyes is switched on one side of the equality, it must also be switched on the other side. Because of the failure of joint-presentation symmetry (Steingrímsson, 2009), left and right distributivity cannot be assumed to be equivalent.³

As with Eq. (6), by replacing $\oplus$ and $\circ_p$ with $+$ and $\times$, it is easy to see that Eqs. (7) and (8) are forms of distributivity of $\oplus$ over $\circ_p$, and if, in Eq. (6), 0 is replaced by z , the relation of Eqs. (6) to (7) and (8) becomes evident.

Experiments

Three experiments are presented: The first is a test of bisymmetry (Exp. 1), which is an induced property of the summation representation (Eq. 3). The remaining two, the simple joint-presentation decomposition (Exp. 2) and distributivity (Exp. 3), are properties that link the summation and proportion representations and, if found to hold, support the hypothesis that $\Psi_{\oplus} = \Psi_{\circ_p} = \Psi$.

General method

The experiments have several common testing features, which are outlined in the following.

Respondents A total of 15 students from New York University (NYU) and the University of California, Irvine (UCI), and the author⁴ participated in the three experiments. For practical reasons, not all the respondents participated in all of the experiments. All the respondents reported normal or corrected-to-normal vision and, except for the author, received compensation of \$10 per session. Every 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 the Institutional Review Boards of NYU and UCI.

Stimuli The stimuli consisted of squares, subtending 10° of visual angle (except in Exp. 3, in which 5° of visual angle was used to accommodate all the stimuli on the screen

within the view of the mirrors; cf. Fig. 2), of achromatic light (the RGB channels set to the same value) displayed on a computer monitor (e.g., Fig. 1A).

Apparatus Stimuli were presented using PsychToolbox extensions in MATLAB (Pelli, 1997) generated by an Apple G4. At UCI, stimuli were presented on an 18-in. NEC Multisync FE 950 + 17, except for the second condition of Exp. 1, where stimuli were collected at a reviewer’s request and at time when the monitor had been upgraded to an Eizo RadiForce RX320 with an automatic luminance uniformity equalizer and backlight sensor to compensation for luminance fluctuation caused by ambient temperature and passage of time, as well as built-in gamma correction. The diagonal size was 54 cm, maximum resolution was 1,536 × 2,048 pixels, and maximum luminance was 742 cd/m². At NYU, a ViewSonic P810 CRT, at a resolution of 1,027 × 768 pixels and a refresh rate of 75 Hz was used. Experiments were conducted in a dark, light-insulated room.

Luminance calibration: Equipment calibration and background conditions were of two kinds. A photometer, PhotoResearch PR-650, was used to measure luminance. In both cases, a gamma function was estimated by averaging five repeated measures of luminance at every 5th of the 255 RGB values (starting from 1). Luminance intensities are reported in candela per square meter.

NYU Calibration: Measures were taken from both the left and right sides of the monitor, equidistant from the monitor’s center. The luminance disparity between the two sides was not appreciable; hence, a single gamma function was fit to the luminance measures.

UCI Calibration: The NYU calibration was improved with the aim to better counteract any possible spatial luminance inhomogeneity of the monitor.⁵ A gamma function was determined for each stimulus location, and using a reverse lookup procedure, the RGB value was determined that produced, as closely as possible, the desired luminance.

Data collected at UCI will be identified with an *; absent that marker, the data were collected at NYU.

Background and luminance range/steps: The monitors achieved an upper luminance of ~100 cd/m², with the lowest stimulus level used being ~8 cd/m². Initially, stimuli were displayed on a no-luminance-level (zeros for all RGB channels) background. In order to

³ Distributivity is a generalization of the segregation property. The generalization is that it extends segregation to the case where $y > 0$; that is, segregation is the special case of distributivity when $y = 0$ (Luce, 2004).

⁴ This I judged acceptable because knowledge of the experimental design does not change the sensations on which the behavioral tasks of matching and magnitude/ratio productions are based, and the variability trial forms, conditions, and block structure do not readily suggest a biased behavior in favor of a particular result. The author is numbered R8.

⁵ Particular thanks go to M. Rudd, who, as a reviewer on Steingrímsson (2009), was responsible for this improvement in procedure.

Fig. 2 Stimuli displayed on a monitor (**A**) and viewed through a stereoscope (**B**) produce the subjective percept seen by the respondents (**C**). The x , u , z , and t' values are luminance. The stimuli in **C** creates two stimulus “intervals”; on the left, one from $z\oplus u$ to $z\oplus x$, and on the right side, $z\oplus u$ to $t'\oplus t'$. Respondents adjust the luminance of t'

until they are satisfied that the brightness interval between $z\oplus u$ and $z\oplus x$ is perceived as p time the brightness interval between $z\oplus u$ to $t'\oplus t'$. (Format and Panel B reprinted with permission from Steingrimsson, 2009)

minimize the mixing of scotopic and photopic conditions, later experiments used 3.4-cd/m^2 background luminance, a level at which photopic vision is dominant (R. Blake, personal communication, September 12, 2007). This change largely coincides with the location change from NYU to UCI. To maximize available adjustment options, all available stimulus values were used.⁶

Stereoscope: A stereoscope (Fig. 1B) aided in the generation of some of the stimuli. A stereoscope is a mirror system that accomplishes the projection of left (right) half of the monitor to the left (right) eye. Thus, a stimulus of intensity x on the left and u on the right side, viewed through a stereoscope, creates the cyclopic image of the stimulus primitive (x, u) (see joint presentations for details).

Procedure An experimental session lasted no more than an hour. An initial session was used to obtain written consent, explain the task, answer questions, and run practice trials. When practiced, respondents typically completed around 60 estimates per session, organized into blocks of 6 or 8 estimates. The block structure allowed for frequent rest periods, which were encouraged, but their frequency and duration were under respondents' control. Information about the current block and trial number were displayed in small letters in the upper left corner of the screen. Respondents received a minimum of 10 min of dark adaption prior to each session. Training was provided for each of the two tasks (matching or magnitude/ratio production, as applicable).

Statistical method and presentation of results The goal is to evaluate parameter-free null hypotheses that have the generic form $L_{\text{side}} = R_{\text{side}}$. The social sciences, in contrast to other sciences, tend to focus on statistical inferences

based on rejecting null hypotheses. This fact has raised numerous questions from reviewers, and this is reflected in the number and range of statistical testing performed. In physics, this type of testing is quite familiar and tends to take the form of articulating a criterion and a level of accuracy with which a null hypothesis is said to be supported by the data. Here, a similar approach was pursued through the formulation of a criterion consisting of multiple interlocked components that had to hold for accepting the data as supportive of the hypothesis. Furthermore, since we had no a priori model of how individuals relate, all data analysis was done on individual data (e.g., Luce, 1995, p. 20).

- **Component 1:** Historically, researchers evaluating behavioral axioms have, on the basis of not possessing a theory that predicts the distributions of the estimates, used a nonparametric test (e.g., Ellermeier & Faulhammer, 2000; Ellermeier, Narens, & Dielmann, 2003; Falmagne, 1976; Gigerenzer & Strube, 1983; Steingrimsson, 2009; Steingrimsson & Luce, 2005a, 2005b, 2006, 2007; Zimmer, 2005; Zimmer, Luce, & Ellermeier, 2001).

- **Part 1:** The most commonly used test in the articles cited has been the Mann–Whitney U with a significance level of .05. This is also the choice here. This method effectively evaluates whether two samples can be said to be drawn from the same distribution—hence, having the same median. By theorem 2 of Falmagne (1976), the testing method (described later) will produce response distributions that converge to the median of the sample as long as the sample size is adequate (Pratt, 1964).

- **Part 2:** To evaluate the power of the test, it is instructive to note that while in the present article, all predictions are of properties holding, prediction of failures are also made. The overall pattern for these predictions and the pattern of test results demonstrate that the test is capable of both rejecting and failing to

⁶ Some researchers use linearized luminance steps. Since I seek subjective judgments from respondents and this is not a linear function of luminance, physical linearization does not clearly provide an advantage over finer adjustments. Subjective linearization is problematic, due to individual differences.

reject hypotheses in quite systematic fashion.⁷ Pratt's (1964) theorem rests on having a sample adequate to detect a true failure of the null hypothesis. Mumby (2002) suggested that for a nonparametric test, the best method for this purpose is a Monte Carlo simulation. The simulation devised consisted of drawing, without replacement, two samples from the combined data and carrying out the Mann–Whitney test. This was repeated 1,000 times, and if the test statistic from component 1 agreed with the distribution of the tests of the simulation at the .05 level, the test's power was deemed adequate. This simulation was carried out for all samples (for additional details, see Steingrímsson, 2009; Steingrímsson & Luce, 2005a).

- **Component 2:** No agreed method exists for calculating the effect size for nonparametric tests. A novel approach to this issue⁸ makes use of the simple observation that should two medians (means) differ by less than Weber's fraction, they are arguably not noticeably different to an observer. Weber's fraction can vary considerably by testing condition. By way of direct assessment, the lower bound for the fraction for the experimental situation in the present article is .05.⁹ Teghtsoonian (1971) reported the mean Weber's fraction for brightness from five, reportedly conservative, studies, to be .08, but these used some variation of sequential presentation, which is known to introduce temporal bias. The Weber's fraction for present purposes is taken as the lower of the two—that is, .05.
- **Component 3:** In recent years, some effort has been directed at the use of Bayesian methods to evaluate point hypothesis (e.g., Gallistel, 2009; Rouder, Speckman, Sun, Morey, & Iverson, 2009; and see Wagenmakers, 2007, for a review). To complement the testing, as well as to respond to reviewers' unease about the Mann–Whitney test, all results are also analyzed using a Bayesian test (Gallistel, 2009).¹⁰

⁷ As a concrete example, in Luce, Steingrímsson, and Narens (2010), of 50 tests predicted to hold, 45 were found to be supported, and, out of 16 conditions predicted to fail, 15 failed statistically. Where it the case that the test had very low power, this pattern of results would constitute a staggering coincidence, and more so since many similar results may be cited.

⁸ I thank J. Yellott for this simple and elegant observation.

⁹ Respondents matched a static standard to a static variable (stimuli identical to those used in the experiments). Several intensities for the standard were used. The Weber's fraction is $\Delta\psi/\psi$ where ψ is the psychophysical function. Using the well-supported power form for ψ and the population exponent $\beta = .33$, the average for four individuals was .025. With each of the three experiments having the feature that an estimate from a first step is used in a second, the applicable fraction is .05.

¹⁰ Thanks go to Dr. Gallistel for providing both computer code and full assistance with adapting it and his method to the present testing strategy.

Gallistel (personal communication, September, 2010) recommended a two-part approach:

- **Part 1:** Carry out the test and establish the odds for and against the null hypothesis. This involves first establishing a boundary to constrain the prior. In evaluating effect size (component 2), Weber's fraction was chosen as the outside limit of by how much the medians can be accepted to differ before the null hypothesis is rejected. It seemed, therefore, natural to use a similar limit to constrain the prior. In practice, a slightly larger limit was used, .1, because of the second part of the test. It is required that the odds for the null be at least 2:1.
- **Part 2:** There is a certain arbitrariness in the choice of the prior and, thus, the evidence obtained for and against the null hypothesis. Therefore, say, in the case of the evidence being in favor of the null hypothesis, the test is run with the prior approaching 1, and it is recorded whether the alternative ever becomes favored. If so, the test is not said to provide sufficient evidence for the null hypothesis.

The statistical criterion established is that all five parts of the three components must favor the invariance relationship for it to be said to be supported by the data.

An additional test was requested by a reviewer: All the samples were evaluated for possible evidence of bias evidenced by the maximal and minimal values of the samples over the samples in an experiment. No such bias was observed in any of the three experiments; thus, the outcome of this additional test is not stated explicitly, apart from a summary in the final discussion.

Since intensity steps are discrete and estimates appear reasonably Gaussian, medians are known to be best estimated by the mean, and variability thus to be indicated by standard deviations. Therefore, these are the central tendency indicators reported. The intensity variable that is actually manipulated is the monitor's lookup table value (LUT). Converting these to candela per square meter involves a power transform (i.e., the gamma function that is commonly fit LUTs). If standard deviations are taken of these transformed values, then they grow as a power of the intensity, rather than proportionally, and, as a consequence, can be quite misleading. To counter this issue, I follow the suggestion of J. Yellott (personal communication, September, 2010) and report normalized standard deviations incandela per square meter.

Experiment 1: bisymmetry

Method Empirical testing of Eq. (5) requires obtaining several respondent-generated matches.

The summation operation, $\oplus$, and matches: The task is to find (z, z') that is perceived as equally bright as (x, u) . Figure 1 describes the process: Fig. 1A depicts what is displayed on the monitor, where the letters indicate stimulus intensity. Figure 1B depicts the stereoscope through which the respondents view the monitor. Figure 1C depicts what the subject sees. Since the stereoscope creates a cyclopic image, the unitary percepts are those of $z \oplus z'$ and $x \oplus u$.

To produce a brightness match, respondents adjust the intensity of z until they are satisfied that the two percepts—the upper and lower squares in Fig. 1C—are experienced as equal in brightness. Respondents used keypresses either to adjust the luminance of z or to indicate satisfaction with the brightness match. Respondents could choose any of four luminance steps of 1, 2, 4, or 8 RGB values—with equal change on the three channels—described as extra-small, small, medium, and large. After an adjustment, the screen was set to a uniform background luminance for 100 ms, and then the next stimulus was presented; subjectively, this was experienced as a very brief blink that signaled that the adjustment had been made. This process was repeated until respondents were satisfied with the match, which they indicated by a keypress, at which time the trial ended and z was recorded as the response. In verbal instructions to respondents, the task was explained as that of making the upper stimulus equal in brightness to the lower one.

Empirical testing of Eq. (5) requires obtaining six matches made in two steps. The left side of Eq. (5) is reduced to a single estimate through three (matches) estimates:

$$w = x \oplus y, w' = u \oplus v, \text{ and then } t = w \oplus w'.$$

And similarly, for the right side of Eq. 5,

$$z = x \oplus u, z' = y \oplus v, \text{ and then } t' = z \oplus z'.$$

The bisymmetry property is said to hold if t and t' are found to be statistically equivalent.

These six estimates were made within a block of trials.¹¹

Listed in Table 1 are the stimulus values for the two conditions under which the bisymmetry was evaluated.

Results The data for 10 respondents are presented in Table 2. Listed for each respondent are the means and standard deviations for t and t' and the number of observations, n , for each sample. The statistics portion

Table 1 The two stimulus conditions under which the Bisymmetry was tested

Condition	Stimuli (cd/m ²)			
	x	y	u	v
C ₁	13.01	29.84	53.78	84.93
C ₂	12.37	42.61	115.72	140.25

consists of the three components. The first is the Mann–Whitney, where what is reported is the result of the hypothesis test $t \sim t'$, given as $p_{t \sim t'}$, as well as the results of the simulation for evaluating the adequacy of the samples to detect the true failure of the hypothesis. The evaluation of the effect size is the second component, which asserts that the samples may not differ by more than .05 (~Weber's fraction). The third component is a Bayesian test, where first are reported are the odds for (OF) the null hypothesis, as compared with the odds against (OA) the null hypothesis, which are required to be a minimum of 2:1, followed by the question as to whether OF ever exceeds OA as the prior approaches 1 (OF/OA). For the data to be said to support the hypothesis, all five elements must support it. That conclusion is reported in the last column of the table.

R9 failed both the first and the second components of the statistical criterion, whereas the other subjects passed the test. R8, who was the only one available to run both stimulus conditions, passed both. That is, the bisymmetry property was not rejected for 10 of 10 respondents in 11 tests.

Discussion The bisymmetry property is an induced property of the summation representation (Eq. 3). Its being supported in 10/11 tests provides favorable initial support for that representation.

Experiment 2: simple joint-presentation decomposition

Method The goal is to evaluate simple joint-presentation decomposition (Eq. 6). This evaluation requires obtaining respondent-generated matches and several magnitude productions. The matching procedure was outlined in Experiment 1; here, the magnitude production procedure is described.

Magnitude production: The task $(z', z') = (x', u') \circ_p (0, 0)$ is to produce a stimulus (z', z') that is perceived as a proportion p of the standard (x', u') . Note that when $p = 1$, this task amounts to matching. Thus, a magnitude production may be obtained using a procedure and a stimulus identical to those in the matching task, with only the addition of a proportion instruction (see the Method section for Exp. 1 for details).

¹¹ Steingrimsón and Luce (2005b) showed that in experiments requiring multiple estimation steps, it was important that all estimates be collected within the same session (to avoid intersession variability) and that individual estimates from one step be used as input into the following step (as opposed to using medians or averages of estimates), in order to preserve accumulation of variance in multistep estimations.

Table 2 Results of Experiment 1: Bisymmetry. Listed for each respondent, the conditions tested, means and normalized standard deviations of the results, number of observation obtained for eachcondition, and the results of the statistical testing. Under statistics, OF stands for odds for and OA for odds against and OF $\nearrow$ OA indicates whether the OF ever exceeds OA as the prior approaches 1

Respondent	Condition	Luminance Level					Statistics					
		t		t'		n	Mann–Whitney		Effect Size	Bayesian Test		Conclusion
		M	SD	M	SD		$p_{t \sim t'}$	Simulation		OF:OA	OF $\nearrow$ OA	
R3	C ₁	61.21	3.42	61.53	3.94	30	1.000	Pass	Pass	5.1:0.20	No	Pass
R4	C ₁	29.59	1.17	30.27	1.21	30	.285	Pass	Pass	4.3:0.23	No	Pass
R5	C ₁	54.23	3.47	53.31	5.53	30	.414	Pass	Pass	3.2:0.31	No	Pass
R6	C ₁	59.83	4.36	59.83	5.45	35	.925	Pass	Pass	4:0.25	No	Pass
R7	C ₁	42.71	2.38	42.03	3.12	30	.620	Pass	Pass	4.1:0.24	No	Pass
R8	C ₁	25.83	0.90	25.80	1.04	30	.847	Pass	Pass	8.2:0.12	No	Pass
	C ₂ [*]	94.60	6.18	96.32	5.42	30	.523	Pass	Pass	4.3:0.23	No	Pass
R9	C ₁	44.10	6.39	46.74	4.88	30	.241	Pass	Fail	1.5:0.68	Yes	Fail
R83	C ₂ [*]	76.75	2.68	78.61	1.85	30	.150	Pass	Pass	3.5:0.29	No	Pass
R86	C ₂ [*]	58.06	4.35	57.08	4.42	30	.756	Pass	Pass	3.7:0.27	No	Pass
R89	C ₂ [*]	77.34	5.40	78.94	4.76	37	.505	Pass	Pass	4.2:0.24	No	Pass

^{*}R6 used the same stimulus value for $p = 2$ as for $p = 2/3$

Recalling that with $x \circ_p y = (x, x) \circ_p (y, y)$, Eq. (6) may be written as

$$[(x \oplus u), (x \oplus u)] \circ_p (0, 0) = [(x, x) \circ_p (0, 0)] \oplus [(u, u) \circ_p (0, 0)]. \quad (9)$$

Testing the property required five trial types:

- A: $(z, z) \sim (x, u) \circ_1 (0, 0)$
- B: $(t, t) \sim (z, z) \circ_p (0, 0)$
- C: $(v, v) \sim (x, x) \circ_p (0, 0)$
- D: $(w, w) \sim (u, u) \circ_p (0, 0)$
- E: $(t', t') \sim (v, w) \circ_1 (0, 0)$

Together, the trial types A and B reduce the left side of Eq. (9) to a single estimate t , and in the subsequent trials, C–E, the right side of Eq. (9) is reduced to the estimate t' (Note that trial types A and E, where $p = 1$, are equivalent to matching). The property is said to be supported if the hypothesis $t \sim t'$ is not rejected.

A theoretical prediction is that the property holds for both $p < 1$ and $p \geq 1$. Two production conditions meeting these constraints were chosen—namely, $p = 2/3$ and $p = 2$. The luminance levels used in the case of $p = 2$ were $x = 12.98$ cd/m² and $u = 29.80$ cd/m², and for $p = 2/3$, the luminance values were $x = 29.80$ cd/m² and $u = 53.74$ cd/m².

With two proportion instructions, there were a total of 10 trial forms. These were run randomized¹² within a block of

trials. The p instruction was displayed in the upper left corner of the monitor.

The task was described to respondents as making the upper square appear, for example, twice ($p = 2$) or two thirds ($p = 2/3$) that of the lower square. For the matching trials, the instruction was $p = 1$, and the task was explained as equivalent to matching. Respondents were initially observed making the adjustments to help ensure complete understanding of the task.

Results The data for 6 respondents are presented in Table 3. Listed for each respondent are the means and standard deviations for t and t' , the proportion p , and the number of observations, n . The statistics portion consists of the three components. The first is the Mann–Whitney, where what is reported is the result of the hypothesis test $t \sim t'$, given as $p_{t \sim t'}$ as well as the results of the simulation for evaluating the adequacy of the samples to detect the true failure of the hypothesis. The evaluation of the effect size is the second component, which asserts that the samples may not differ by more than .05 (~Weber's fraction). The third component is a Bayesian test where first reported are the odds for the null hypothesis (OF) compared with the odds against (OA) the null hypothesis, which are required to be a minimum of 2:1, followed by the question as to whether OF ever exceed those of the OA as the prior approaches 1. For the data to be said to support the hypothesis, all five elements must support it. That conclusion is reported in the last column of the table.

Respondent R4 failed the second part of the Bayesian test in one test. Even though R4 passed all the others, the

¹² The randomization was constrained by, for example, needing an estimate for z prior to estimating a t .

Table 3 Results of Experiment 6: Simple joint-presentation decomposition. Listed for each respondent, the conditions tested, means and normalized standard deviations of the results, number of observation obtained for each condition, and the results of the statistical testing.

Under statistics, OF stands for odds for and OA for odds against and OF/OA indicates whether the OF ever exceeds OA as the prior approaches 1

Respondent	Luminance Level						Statistics									
	t		t'		p	n	Mann–Whitney		Effect Size	Bayesian Test		Conclusion				
	M	SD	M	SD			$p_{t \sim t'}$	Simulation		OF:OA	OF $\nearrow$ OA					
R3	46.20	4.47	48.12	4.20	2	30	.155	Pass	Pass	2.3:0.43	No	Pass				
	21.39	0.90	21.75	1.17	$_{2/3}$	30	.941	Pass	Pass	5:2.19	No	Pass				
R4	40.55	1.95	41.83	1.49	2	32	.075	Pass	Pass	2.5:0.40	Yes	Fail				
	14.33	2.54	14.67	1.96	$_{2/3}$	34	.279	Pass	Pass	2.0:0.51	No	Pass				
R6†	63.23	4.80	64.15	4.85	2	30	.574	Pass	Pass	3.8:0.26	No	Pass				
	36.41	1.72	35.57	1.57	$_{2/3}$	30	.192	Pass	Pass	4.4:0.23	No	Pass				
R8	33.4	4.60	32.36	3.86	2	30	.670	Pass	Pass	3.2:0.45	No	Pass				
R9	57.96	5.19	55.90	5.74	2	30	.329	Pass	Pass	2.4:0.42	No	Pass				
	28.44	2.32	28.16	2.48	$_{2/3}$	30	.722	Pass	Pass	3.5:0.28	No	Pass				
R15	78.93	4.55	78.06	2.90	2	30	.283	Pass	Pass	6.1:0.16	No	Pass				
	16.14	1.37	15.66	1.36	$_{2/3}$	30	.988	Pass	Pass	2.9:0.35	No	Pass				

†R6 used the same stimulus value for $p = 2$ as for $p = 2/3$

rejection of the test is also somewhat supported by a marginally low $p_{t \sim t'}$ value. Thus, after evaluating the results by all three components of the statistical criterion, the property was found to hold for both proportion conditions in 11/12 tests.

Discussion With the property not rejected for any of the respondents for either the $p < 1$ or the $p > 1$ condition in 12 tests, the property is taken as having received reasonable initial support for brightness.

Experiment 3: distributivity

Method Distributivity comes in two forms, left (Eq. 7) and right (Eq. 8) distributivity. The method for testing left distributivity will be outlined; the method for right distributivity is analogous. The testing required matching and several ratio productions. The matching procedure was outlined in Exp. 1. Here, the ratio production procedure is described.

Ratio Production: The task $(t', t') = (z, x) \circ_p (z, u)$ is that of producing a stimulus (t', t') that makes the brightness “interval” from (z, u) to (t', t') be a proportion p of the “interval” from (z, u) to (z, x) . Figure 2 illustrates how this is accomplished. Figure 2A shows the stimuli as presented on the screen. Figure 2C illustrates the percept the respondent sees when the display in Fig. 1A is viewed through the stereoscope shown in Fig. 1B. Recalling that the percept of the stimulus (z, x) is that of $z \oplus x$, then by a method analogous to that of matching (Fig. 1), the respondent need

only adjust the luminance of t' to arrive at an estimate of the ratio production.

The adjustment procedure is the same as that for matching; the proportion p was displayed on the upper left side of the monitor (as was noted earlier; and to accommodate the number of stimuli on the monitor, each square subtended 5° of visual angle).

Recalling that $x \circ_p y = (x, x) \circ_p (y, y)$, then Eq. (7) may be written as

$$[z, (x, x) \circ_p (u, u)] = (z, x) \circ_p (z, u). \quad (10)$$

The testing required three trial forms

- A: $(v, v) \sim (x, x) \circ_p (u, u)$
- B: $(t, t) \sim (z, v)$
- C: $(t', t') \sim (z, x) \circ_p (z, u)$

The left side of Eq. (10) is reduced to a single estimate t using the trial types A and B; the right side is reduced to a single estimate t' using the trial type C.

Left distributivity is found to be supported if the hypothesis $t \sim t'$ is not rejected.

The property is predicted to hold for both $p < 1$ and $p \geq 1$; $p = 2$ and $p = 2/3$ were the proportions used. The RGB values for u and z were fixed in all conditions, and thus the corresponding luminance values varied slightly between the two monitors used. At NYU (UCI), these were $u = 10.50$ (7.46) cd/m^2 and $z = 18.90$ (15.06) cd/m^2 . Three values for x were used and mixed over proportion conditions; these are specified in Table 4.

Table 4 Results of Experiment 3: Distributivity. Bisymmetry. Listed for each respondent, the conditions tested, means and normalized standard deviations of the results, number of observation obtained for

each conditions, and the results of the statistical testing. Under statistics, OF stands for odds for and OA for odds against and OF/OA indicates whether the OF ever exceeds OA as the prior approaches 1

Respondent	Version	Luminance Level								Statistics					
		<i>t</i>				<i>t'</i>				Mann–Whitney		Effect Size	Bayesian Test		Conclusion
		<i>x</i>	<i>M</i>		<i>SD</i>		<i>p</i>	<i>n</i>	OA:OF				OF $\nearrow$ OA		
			<i>p_{t~t'}</i>	Sim.											
R4	Left	25.87	38.26	0.80	38.04	0.70	2	30	.221	Pass	Pass	14:0.07	No	Pass	
*		35.22	20.31	0.42	20.72	0.52	2/3	30	.536	Pass	Pass	5.8:0.17	No	Pass	
*		21.93	38.36	3.96	37.87	2.66	2	24	.090	Pass	Pass	3.2:0.31	No	Pass	
*		65.34	29.05	2.28	32.61	0.89	2/3	24	<.001	Fail	Fail	.02:58.8.2	Yes	Fail	
R8	Left	25.87	37.45	2.18	37.22	2.31	2	30	.711	Pass	Pass	5:0.20	No	Pass	
*		35.22	20.53	0.42	20.64	0.97	2/3	30	.112	Pass	Pass	9.1:0.11	No	Pass	
*		21.93	26.67	1.02	27.09	1.32	2	30	.894	Pass	Pass	6.2:0.16	No	Pass	
*		35.22	21.57	0.52	21.34	0.74	2/3	30	.929	Pass	Pass	8.6:12	No	Pass	
R10	Left	25.87	37.79	2.18	37.88	1.28	2	30	.976	Pass	Pass	6.6:0.15	No	Pass	
		25.87	20.16	0.60	20.52	0.50	2/3	30	.225	Pass	Pass	5.1:0.19	No	Pass	
R12	Left	38.51	41.70	1.73	43.49	1.67	2	30	.080	Pass	Pass	1.1:0.94	Yes	Fail	
		38.51	22.17	0.59	22.70	0.52	2/3	30	.130	Pass	Pass	2.3:0.43	No	Pass	
R13	Left	38.51	39.82	1.65	38.20	1.8	2	30	.131	Pass	Pass	1.4:0.71	Yes	Fail	
		38.51	22.29	0.43	22.05	1.40	2/3	30	.422	Pass	Pass	14:0.07	No	Pass	
R35	Left	21.93	37.78	1.86	39.09	1.07	2	30	.134	Pass	Pass	2.5:0.40	No	Pass	

* Data collected at University of California, Irvine

With two proportion instructions and three trial types, there was a total of six trial forms. The four trials (two each of type A and C) were randomized with a block and run within a session. A second session comprised the other two trials (type B) run in a single block consisting of each individual estimate of v (from A) matched with z . The p instruction was displayed in the upper left corner of the monitor.

The task was described to respondents as adjusting the brightness of the upper right square such that the adjusted interval between it and the lower right square be twice ($p = 2$) or two thirds ($p = 2/3$) that of the reference interval, the lower left to the upper left squares. Respondents were initially observed making the adjustments to help ensure complete understanding of the task.

Results The data for 6 respondents are presented in Table 4. Listed for each respondent are the means and standard deviations for t and t' , the proportion p , and the number of observations, n . The statistics portion consists of the three components. The first is the the Mann–Whitney, where what is reported is the result of the hypothesis test $t \sim t'$, given as $p_{t \sim t'}$ as well as the results of the simulation for evaluating the adequacy of the samples to detect the true failure of the hypothesis. The evaluation of the effect size is the second component, which asserts that the samples may not differ by more than .05 (~Weber's fraction). The third

component is a Bayesian test where reported first are the odds for (OF) the null hypothesis, compared with the odds against (OA) the null hypothesis, which are required to be a minimum of 2:1, followed by the question as to whether OF ever exceed those of the OA as the prior approached 1. For the data to be said to support the hypothesis, all five elements must support it. That conclusion is reported in the last column of the table.

R4 failed outright the property for one test (left, $p = 2/3$). R12 and R13 failed both parts of the Bayesian test for one condition (left, $p = 2$), respectively. Therefore, the property is found to be supported in 12 of 15 tests.

Discussion Evaluating distributivity involves arguably the most complex psychophysical task, and it is technically the most complex of the tasks in the present three experiments, as well as of those evaluated by Steingrimsdottir (2009). Therefore, its being supported in 12 of 15 tests is a good initial support for the property.

General discussion

The results of the tests of bisymmetry and the two linking properties are summarized in Table 5.

Table 5 Summary of experimental results

Property (Exp. #)	#Respondents	#Tests	#Fail	% Pass	Conclusion
Bisymmetry (Exp. 1)	10	11	1	91	Supported
JP decomposition (Exp. 2)	6	12	1	92	Supported
Distributivity (Exp. 3)	6	15	3	80	Supported

By a reviewer's request, all samples were evaluated for possible evidence of bias evidenced by the maximal and minimal values that obtained in them. No such bias was observed in any of the three experiments.

The topic has been the evaluation in brightness of three behavioral properties that arise in a theory of global psychophysical judgments, which leads to the two representations:

$$\Psi(x, u) = \Psi(x, 0) + \Psi(0, u) + \delta \Psi(x, 0) \Psi(0, u) \quad (\delta = 0, 1), \quad (3)$$

$$W(p) = \frac{\Psi[(x, u) \circ_p (y, v)] - \Psi(y, v)}{\Psi(x, u) - \Psi(y, v)} [(x, u) \succ (y, v) \sim (0, 0)], \quad (4)$$

These representations have a number of necessary consequences (behavioral properties) that, in turn, are sufficient under certain structural conditions to give rise to the representations. Those that underlie the summation representation (3) and proportion (4) were examined separately and sustained in Steingrímsson (2009). Those

that link the two representations and force a common psychophysical function were examined here.

The result of Steingrímsson (2009) and the present article are summarized in Fig. 3, as well as how the two relate and how, together, they form the conclusion that establishes the representations (3) and (4).

The main conclusion from these empirical evaluations is that the theory of global psychophysics (Luce, 2002, 2004, 2008 [erratum]) has received reasonable (initial) support in the brightness domain for achromatic stimuli of intensity well above threshold.

The representations (3, 4) have been found to capture behavior in two separate domains, loudness and brightness, which to my knowledge no model has accomplished before. The results suggest exploring the extension of the work not only to additional domains (work on perceived contrast is under way), but also to the level at which the description of behavior is unified. This question is the topic of Appendix C of Steingrímsson (2009). In the appendix, the meaning of the theoretical result is discussed in a broader context of invariances. These results readily open the way for a third article in the series, one whose topic is the functional form

Fig. 3 The diagram shows, on the left, the properties tested by Steingrímsson (2009); on the right, those tested in this article; and in the lower middle, their results, their implications, and how they all come together to establish the representations (3) and (4). The dotted line into Bisymmetry indicates that the property should most logically have been tested in Steingrímsson (2009). At the bottom, the topic of the third article Steingrímsson (in preparation) the functional form of Ψ and W emerges as the next article in the series

of Ψ and W . This work has been completed for loudness (Steingrímsson & Luce, 2006, 2007) and is in progress for brightness.

Ever since Newton's remarkable feat of deriving a description of the mechanical world from three laws and one assumption, physics has favored his combining of inductive and deductive methodology, because it has provided both incremental growth in theory development and the making of novel predictions. In following the same approach here, the same strength has been realized in both an ever-expanding scope of Luce's (2002, 2004) original theory and by a number of derived and empirically testable/ tested predictions. For instance, Luce, Steingrímsson, and Narens (2010) extended the theory to the evaluation of the commonality of intensity scales while a second attribute such as frequency was varied. On the basis of Luce's (2002, 2004) theory's holding for loudness, they successfully evaluated the resulting axioms for loudness, and the result was that the scales are the same as long as different reference points for increasing and decreasing intensities are allowed to differ; thus, individuals can be said to rely on a single scale of loudness.

One direct benefit of establishing support for the theory for brightness (Steingrímsson, 2009; present article) is that it allows the evaluation of the commonality of brightness scales. The resulting article (Steingrímsson, Luce, & Narens, submitted) reports data that, in pattern, are identical to those for loudness. That result, in turn, opens the avenue for asking whether a single scale of intensity exists for brightness and loudness and, if so, potentially for other intensive continua. Numerous other predictions and avenues are being explored. Of these, I mention one, since it further shores up the conclusion of Steingrímsson (2009). The result is that a certain commutative rule first formulated by Falmagne (1976) is, in fact, equivalent to the Thomsen condition (Luce & Steingrímsson, submitted) and, arguably, solves certain problems with empirically evaluating that property. The data for loudness and brightness were reported by Steingrímsson & Luce (2010), with further elaboration of these planned for Luce & Steingrímsson (submitted).

Author Note This research was supported in part by National Science Foundation grants SBR-9808057 and BCS-0720288 (R. D. Luce, PI) to UCI. Any opinion, finding, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation. Additional financial support was provided by the School of Social Sciences and the Department of Cognitive Sciences at UCI. I am especially grateful to Marisa Carrasco for unfettered access to her laboratory and equipment at NYU to carry out much of the research reported here. Thanks go to Charles Gallistel for providing the Bayesian analysis and advice for its use. And I appreciate the many helpful comments of Joetta Gobell and R. Duncan Luce on earlier versions.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *The American Psychologist*, 57, 1060–1073.
- Ding, J., & Sperling, G. (2006). A gain-control theory of binocular combination. *Proceedings of the National Academy of Sciences*, 103, 1141–1146.
- Ellermeier, W., & Faulhammer, G. (2000). Empirical evaluation of axioms fundamental to Stevens's ratio-scaling approach: I. Loudness production. *Perception & Psychophysics*, 62, 1505–1511.
- Ellermeier, W., Narens, L., & Dielmann, B. (2003). Perceptual ratios, differences, and the underlying scale. In B. Berglund & E. Borg (Eds.), *Fechner Day 2003: Proceedings of the 19th Annual Meeting of the International Society for Psychophysics* (pp. 71–76). Stockholm: International Society for Psychophysics.
- Falmagne, J.-C. (1976). Random conjoint measurement and loudness summation. *Psychological Review*, 83, 65–79.
- Fechner, G. T. (1861). Über einige Verhältnisse des binocularen Sehens. *Abhandlungen der mathematisch-physischen Classe der königlich sächsischen Gesellschaft der Wissenschaften*, 5, 337–564.
- Gallistel, C. R. (2009). The importance of proving the null. *Psychological Review*, 116, 439–453.
- Gigerenzer, G., & Strube, G. (1983). Are there limits to binaural additivity of loudness? *Journal of Experimental Psychology: Human Perception and Performance*, 9, 126–136.
- Grossberg, S., & Kelly, F. (1999). Neural dynamics of binocular brightness perception. *Vision Research*, 39, 3796–3816.
- 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.
- Levelt, W. J. M. (1965). Binocular brightness averaging and contour information. *British Journal of Psychology*, 56, 1–13.
- Luce, R. D. (1995). Four tensions concerning mathematical modeling in psychology. *Annual Reviews of Psychology*, 46, 1–26.
- Luce, R. D. (2002). A psychophysical theory of intensity proportions, joint presentations, and matches. *Psychological Review*, 109, 520–532.
- 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., Krantz, D. H., Suppes, P., & Tversky, A. (1990). *Foundations of measurement, vol III*. New York: Academic Press. Reprinted 2007 by Dover Publications.
- Luce, R. D., & Steingrímsson, R. (2008). Note on a changed empirical inference in several Steingrímsson 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., Steingrímsson, 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.
- Luce, R. D., & Steingrímsson, R. (submitted). *Testing for additive conjoint measurement*.

- Mumby, P. J. (2002). Statistical power of non-parametric tests: A quick guide for designing sampling strategies. *Marine Pollution Bulletin*, 44, 85–87.
- Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442.
- Pratt, J. W. (1964). Robustness of some procedures for the two-sample location problem. *Journal of the American Statistical Association*, 59, 665–680.
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16, 225–237.
- Steingrimsson, R. (2009). Evaluating a model of global psychophysical judgments in 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: III. Forms for the psychophysical and the weighting function*. Manuscript in preparation.
- 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.
- Steingrimsson, R., & Luce, R. D. (2010). The commutative rule as new test for additive conjoint measurement: Theory and data. In A. Bastianelli & G. Vidotto (Eds.), *Fechner Day 2010. Proceedings of the 19th Annual Meeting of the International Society for Psychophysics* (pp. 21–26). Padua: International Society for Psychophysics.
- Steingrimsson, R., Luce, R. D., & Narens, L. (submitted). *Brightness of different hues is a single psychophysical ratio scale of intensity*.
- Stevens, S. S. (1975). *Psychophysics: Introduction to its perceptual, neural, and social prospects*. New York: Wiley.
- Teghtsoonian, R. (1971). On the exponents in Stevens' law and the constant in Ekman's law. *Psychological Review*, 78, 71–80.
- Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, 14, 779–804.
- Zimmer, K. (2005). Examining the validity of numerical ratios in loudness fractionation. *Perception & Psychophysics*, 67, 569–579.
- Zimmer, K., Luce, R. D., & Ellermeier, W. (2001). Testing a new theory of psychophysical scaling: Temporal loudness integration. In G. Sommerfeld, R. Kompass, & T. Lachmann (Eds.), *Fechner Day 2001: Proceedings of the 17th Annual Meeting of the International Society for Psychophysics*. Lengerich: Pabst.