Dear Cognition and Color Reading Group

The following is a submitted manuscript, currently undergoing a major revision that includes additional data. Current results do not alter the conclusions reached in the manuscript although methodology is refined substantially.

Please do not distribute or cite.

Sincerely,

Ragnar Steingrimsson.

Evaluating a Model of Global Psychophysical Judgments for Brightness: I. Behavioral Properties of Summations and Productions

Ragnar Steingrimsson
October, 14, 2008

Send correspondence to:
R. Steingrimsson
Social Science Plaza
University of California
Irvine, CA 92697-5100
FAX: 949-824-3733
Phone: 917.670.8082
E-mail: ragnar@nyu.edu

Evaluating a Model of Global Psychophysical Judgments for Brightness:
I. Behavioral Properties of Summations and Productions

Abstract

Steingrimsson and Luce (2005a,b, 2006, 2007) evaluated Luce's (2002, 2004) proposed psychophysical theory in loudness and found it substantially supported. The aim of the current research is to begin an extension of this research to brightness. Luce's theory deals with the global percept of subjective intensity, in which there is a psychophysical function Ψ that maps pairs of physical intensities onto the positive real numbers and represents, in an explicit mathematical way, subjective summation and a form of ratio production. These representations derive from a number behavioral properties including certain plausible background assumptions. In three experiments involving the subjective perception of luminance, brightness, the key behavioral properties of summation over the two eyes and a form of generalized ratio production are empirically evaluated. Considerable support is reported for particular forms of Ψ for summations and ratio productions separately.

Keywords: Binocular brightness; Brightness summation; Ratio production; Magnitude production; Magnitude estimation; Psychophysics; Matching; Production commutativity; Thomsen condition; Double cancellation;

In a series of, currently, four, papers Steingrimsson and Luce (2005a,b, 2006, 2007) evaluated Luce's (2002, 2004) theory of global psychophysics in loudness. In sum, the result provided broad support for the branch of the theory in which the left and the right ears are not assumed behaviorally completely alike (the *biased* or *asymmetric case*). The main aim of this paper is to begin an analogous series of evaluations of the theory in the brightness domain. Specifically, the current paper is patterned after Steingrimsson and Luce (2005a) and since there is no compelling reason to do otherwise, borrows much of its structure and presentation from that paper. However, apart from studying a new domain, this paper differs from Steingrimsson and Luce (2005a) in two notable ways. First, their Experiment 1 is not carried out here—this experiment concerned single-ear matches but due to Fechner's paradox (Fechner, 1861) in brightness perception, these types of matches are not accounted for in the present theoretical context (more about this later). Second. the exposition is, by editorial advice, aimed at an audience not necessarily familiar with the methods of axiomatic psychophysics. The latter is reflected in the following overview and "reading guide" for the paper:

- Summary of relevant theory and interpretation in brightness: Summarizes the relevant portions of Luce's (2002, 2004) theory of global psychophysical judgments. In particular, two representations, (5, 6), are presented. It is the aim of this paper to explore support for those two, separately. In the tradition of the axiomatic approach, the interpretation of the theoretical primitives is also developed. In an effort to set this paper in the context of an ongoing research program, that experimental program is briefly outlined.
- Experiments: Three experiments are presented. The first examines whether the eyes can be treated as identical with respect to brightness perception. The outcome of this experiment determines which of two possible avenues subsequent testing should take. The avenue indicated by the first experiment is followed in the subsequent two experiments—the alternative is not commented on as it is well presented elsewhere (Luce, 2002, 2004). The second and third experiments seek support for the two representation of the theory (5, 6).
- Summary, conclusion, and further work: The paper's subject matter and conclusions are summarized and the consequence of those for subsequent work is outlined. [Review of this section now may be helpful to get an overview of the research]

Some readers may not be well versed in historical background for the current reader, the axiomatic psychophysical approach, or how this model fits in with previous an realated work. Three Appendeces address these three points (the first two may be helpful to some before proceeding further)

- Historical context: The psychophysical approach used in this paper set in a historical context along with developments directly relevant to the current paper.
- Theoretical background: The paper grows out of theoretical developments in axiomatic psychophysics. While this is a rich theoretical literature, only recently have empirical papers in psychophysics begun to appear that make use of this theory building approach.
- Relations the prevous work: A number of models for binocular combination of brightness have been forwarded. The fit of the current model in this body of literature is briefly addressed.

1 Summary of theory and interpretation in brightness

1.1 Primitives

The first step in the axiomatic approach is the specification of the theory's primitives. Interpreting these in the context of brightness perception will determine our stimuli as well as how they may be manipulated (i.e., methods).

1.1.1 Joint presentations

The first primitive is the set of ordered pairs (x, u), where x and u correspond to physical intensities. Our interpretation of this primitive in the visual domain is that of squares of white (all three light projectors set to the same intensity) lights with intensities x and u presented simultaneously to the left and right eyes, respectively.[1]

Thus (x, u) is our stimulus. An associated behavioral task is for respondents to produce a light (z, z) that (in some to-be-specified sense) is perceived as equal in brightness to the stimulus (x, u). Here, this process is termed *brightness matching* or just *matching* for short.

1.1.2 Notational convention

Let ϵ_l and ϵ_r denote thresholds for the left and the right eye respectively and let x' and u' be intensities actually presented in the left and the right eye respectively; then our notation is $x = x' - \epsilon_l$ and $u = u' - \epsilon_r$. Thus, $x = 0$ denotes the threshold intensity (or less) in the left eye and $u = 0$, denotes the same in the right eye. For signals well above threshold, the difference $x - x'$ is negligible. Intensities for x, u are given in cd/m^2.

1.1.3 Ordering

The second primitive, $\succsim$, is the ordering of stimuli by, in our case, brightness: $(x, u) \succsim (y, v)$ means that the stimulus (x, u) is judged to be at least as bright as (y, v). The indifference relation of matching $\sim$ is defined by: $(x, u) \sim (y, v)$ if and only if both $(x, u) \succsim (y, v)$ and $(y, v) \succsim (x, u)$ hold. Importantly, the symbols $\succsim, \sim$ are used rather than $\geqslant, =$, since the latter refer to ordering of real numbers, while the former refers to psychological judgments. These psychological judgments form what in the axiomatic literature is called a weak ordering, [2]. This mean that $\succsim$ behaves similarly to the ordering $\geq$ of the real numbers. Moreover, it is assumed that it agrees with physical intensity in the sense that if the intensity is held constant in one eye, the binocular brightness varies monotonically with intensity changes in the other eye.

This monotonicity assumption in binocluar brightness (fused images) needs addressing in the context of the so-called Fechner's paradox. Fechner's paradox arises when the ratio when $x/u < k$,

---

[1] This is by no means the only possible interpretation of this primitive in brightness. For instance, another one is to project these lights to the two hemifields, which is possibly a more natural, albeit not more or less "correct" interpretation given the ocular-to-cortical wiring. However, this is technically harder to accomplish than what was done here, but certainly worthy of future efforts.

[2] It is not difficult to show from the assumptions formulated by Luce (2002, 2004) that $\succsim$ is a weak order, i.e., transitive and connected, on the stimulus conjoint structure (Proposition 1 of Luce, 2002).

$x < u$ (Fechner, 1861). The paradox is that as the weaker stimulus is decreased in intensity, at some point the subjective sense of brightness increases. This violates Luce's (2002, 2004) monotonicity assumption. For this reason, the testing of brightness has been restricted to the space in which Fechner's paradox does not obtain. According to the isoluminance data of Levelt (1965), Fig. 5, this is the area outside of where the luminance of one eye is about 13% or less than the luminance in the other eye,i.e., $k \approx .13$—Levelt's (1965) stimuli were in structure similar to those used here. Although it is desirable to remove this restriction, it is not serious in practice: contrary to sounds, natural scenes rarely present the visual system with stimuli that vary substantially in intensity to the two eyes—this fact could lead to speculations such that the environment provided little evolutionary pressure to extend behavior into the region where Fechner's paradox manifests itself, but such discussion is well beyond the scope of this paper.

Each joint presentation (x, u) can be matched by (z, z) or formally,

$$(x, u) \sim (z, z) \tag{1}$$

which is referred to as a *symmetric* match. It is convenient to make the dependence of z upon x and u by the notation

$$x \oplus u := z, \tag{2}$$

where $A := B$ means A is defined by B. Given our assumptions below, $\oplus$ can be proven to be a mathematical operator thus allowing us to call it a *summation operator*. Obviously, (1) and (2) imply $(x, u) \sim (x \oplus u, x \oplus u)$.

Luce's theory (2002, 2004) permits what is equivalent to single-eye matches, or *asymmetric* matches; however, Fechner's paradox limits the testing to symmetric matches. In practice, this is not a limitation as the properties of the theory may be tested using symmetric matches only (Luce, 2004).

Note that (x, u) refers to a joint presentation of a signal pair, whereas $x \oplus u$ refers to a subjective *summation* of two signals. In practice, the notation is used somewhat interchangeably.

The ordering primitive allows us to ask one natural question about brightness matching, namely about the *symmetry* of joint presentations:

$$x \oplus u \sim u \oplus x, \tag{3}$$

which is abbreviated as *jp-symmetry.* This test amounts to asking whether the two eyes are behaviorally alike with respect to brightness perception—intuitively, this is akin to testing whether the psychological operator $\oplus$ is commutative. Whether or not this holds matters considerably for the nature of the theory. Therefore, our first experiment focuses on this property.[3]

---

[3] For the curious reader, the original impetus for Luce's (2002) model of global psychophysical judgments were results originally developed within the context of utility theory (Luce, 2000). They were based on an assumption of no bias, which in the psychophysical context could be translated into saying that, e.g., the two eyes or ear are behaviorally identical. This hypothesis has been unambiguously rejected in audition (Steingrimsson & Luce, 2005a) and, as we shall see, for brightness as well (Experiment 1). The auditory data led Luce to generalize the result for the psychophysical context (Luce, 2002) to incorporate such bias. This formulation was subsequently improved and generalized further in Luce (2004).

1.1.4 Generalized ratio production

The third primitive is a generalization of ratio production. Suppose that $x > y \geq 0$ and let $p > 0$ be a positive number. Let (z, z) denote a signal pair that the respondent says makes the brightness "interval"[4] from (y, y) to (z, z) stand in the ratio p to the brightness interval from (y, y) to (x, x). Clearly z is a function of x, y, p. It is convenient to write this function as a mathematical operator of the following form:$(x, x) \circ_p (y, y) := (z, z)$,[5] The *generalization* part of the ratio production can seen by the fact that it agrees with ordinary *ratio production* when $(y, y) = (0, 0)$.

1.2 Representations of $\oplus$ and $\circ_p$

Luce (2002, 2004) gave a set of sufficient behavioral axioms with the key testable one also necessary, formulated in terms of the presented primitives, that allowed him to construct a numerical mapping Ψ, the psychophysical function, over the stimulus pairs that preserves the order $\succsim$, i.e.,

$$\Psi(x, u) \geq \Psi(y, v) \text{ iff } (x, u) \succsim (y, v), \tag{4}$$

and for which there exists a constant $\delta \geq 0$ such that

$$\Psi(x, u) = \Psi(x, 0) + \Psi(0, u) + \delta\Psi(x, 0)\Psi(0, u), \tag{5}$$

and there is a strictly increasing numerical function W from the positive real numbers onto itself such that

$$W(p) = \frac{\Psi\left[(x, x) \circ_p (y, y)\right] - \Psi(y, y)}{\Psi(x, x) - \Psi(y, y)} \quad (x > y \geq 0). \tag{6}$$

In words, the order preserving condition (4) simply states that the brightness ordering judgment of the physical stimuli agrees with the physical intensity ordering. Property (5) is referred to as the *summation representation* (also referred to in the literature as a *p-additive representation*) as it describes what subjects do when subjectively combining the inputs to the two eyes. It is obvious that with with $\delta = 0$, the representation stipulates that binocular brightness is the addition of the subjective inputs to each eye, this is less obvious with $\delta > 0$, but is nevertheless true (Luce, 2002).[6] Property (6) is referred to as the *production representation* as it describes what respondents do when performing the generalized ratio production operation.

These representations encompass the common abstraction where x, u, and y are any non-negative real numbers. Of course, there is in practice a maximum luminance level exposure that is

---

[4] The term "interval" is being used figuratively to refer to the difference in brightness that respondents experience between two intensity pairs.

[5] Using the several background assumptions (listed in Luce, 2002, 2004), one can replace any symmetric pairs, (x, x), by non-symmetric pairs to get the more general expression

$$(x, u) \circ_p (y, v) = (z, w)$$

and vice-versa. Thus, there is no loss of generality in studying $\circ_p$ in the symmetric case.

[6] In short, using the identifications $\Phi(x, u) := \Psi(x, u), \Phi_l(x) := \Psi(x, 0), \Phi_r(u) := \Psi(0, u)$ and the logarithmic transformation $\Phi(x, u) = \ln\left[1 + \delta\Psi(x, u)\right] \quad (\delta > 0)$.the additive representation (5) can be transformed into a binary additive conjoint representation of summation of the form: $\Phi(x, u) = \Phi_l(x) + \Phi_r(u)$.(Luce, 2002).

safe for the human eye. One can speculate on the perceptual experience as this limit is approached, but, again, in practice, it certainly is not one that could ethically be empirically evaluated. In this sense, this abstraction is of minor concern. For the brightness domain, another abstraction worth noting is that, due to Fechner's paradox, the signal corresponding to, e.g., $(x, 0)$ is really equivalent to some (x', u'), but again, this abstraction is not of particular concern in that the research domain is restricted to the stimulus space outside of where Fechner's paradox obtains. This issue was addressed to some degree in Section 1.1.3 and will be considered in context of other modeling efforts in Section C.

An important point is that since the representations consist of two unspecified functions Ψ and W plus the one constant, they allow for great freedom for individual variety. Yet, the functions and the representation are guaranteed to exist, provided that the parameter-free behavioral properties are satisfied. This situation is typical of axiomatic derivations of representations: no free parameters in the axioms and considerable freedom in the representation. Although many methodological issues are involved in experimental tests of such behavioral properties, this approach seems more definitive than trying to fit to the data representations that have many degrees of freedom. Naturally, one is interested in the actual forms of the unknown functions. In the axiomatic context this takes the form of invariance properties that are equivalent to certain functional forms. Steingrimsson and Luce (2006, 2007) have done so in loudness for the psychophysical and the weighting function respectively and found support for power functions—or a Prelec function that is a generalization of a power function—to obtain for most respondents. Future work aims at a parallel investigation for brightness. What we do know is that brightness data from magnitude estimation/production, averaged over respondents, fits reasonably well to a power function form (e.g., Stevens, 1975), a situation similar to loudness. Hence it would not be surprising that the same form would be concluded using axiomatic techniques. Given that W is a cognitive function, one does not expect it to be domain specific. However, for the present investigation, the form of the functions is of no concern.

Previous theoretical work in axiomatic psychophysics has used operations analogous to either the summation or production operations (e.g., Levelt et al., 1972; Narens, 1996). However, the conceptual novelty in the psychological context (although typical in physics) is the linking of these two operations together. This linking proved critical for establishing that a common psychophysical function can be used to represent both summation and production (currently being worked on in brightness).

1.3 Two behavioral properties of the representations

Now stated are the two behavioral properties that are implied individually by the representations (5) and (6) which, in the context of background assumptions, are sufficient for these representations (listed in Luce, 2002, 2004). "Individually" refers to the fact that a priori the Ψ appearing in both representations is not guaranteed to be the same function, call them $\Psi_\oplus$ and $\Psi_{\circ_p}$—it is the subject of ongoing research whether $\Psi_\oplus = \Psi_{\circ_p} = \Psi$ (see Section 1.4). These properties are tested in Experiments 2 and 3.

1.3.1 Subjective summation

From the results of Krantz et al. (1971, p. 250) the key necessary condition of binary additive conjoint measurement, the *Thomsen condition*

$$\left.\begin{array}{l} x \oplus t \sim v \oplus w \\ v \oplus u \sim z \oplus t \end{array}\right\} \Longrightarrow x \oplus u \sim z \oplus w, \tag{7}$$

must hold. In a qualitative sense, the Thomsen condition describes the "additive cancellation" of t and z.

In the presence of some general background assumptions[7] the Thomsen condition, (7), implies the summation representation, (5). This in turn implies a stronger property called *double cancelation*, which is the same as (7) but with each $\sim$ replaced by $\succsim$. Obviously, double cancelation implies the Thomsen condition, but not conversely except when solvability and monotonicity obtains (see Krantz et al., 1971, for details). See Section C for discussion on binocular summation.

1.3.2 Production commutativity

The basic idea embodied in the representation of generalized ratio production, (6), is that the respondents perform the task as they are told to, using the distortion Ψ of intensities and the distortion W of numbers. An important, easily demonstrated consequence of (6) is the behavioral property called *(subjective) production commutativity:* For $p > 0, q > 0$,

$$[(x, x) \circ_p (y, y)] \circ_q (y, y) \sim [(x, x) \circ_q (y, y)] \circ_p (y, y) \text{ where } y \geq 0 \tag{8}$$

Observe that the two sides differ only in the order of applying p, q, which is the reason for the term "commutativity." Proportion commutativity with $y = 0$ also arose in Narens' (1996) theory. That hypothesis was sustained for brightness by Peißner (1999) using $p, q > 1$. The present work treats the general case of (8) for which $y > 0$.

1.4 The experimental program

The current experimental program is patterned on that carried out for loudness (Steingrimsson & Luce, 2005a,b; 2006; 2007). For various reasons, the work was broken into four parts. The first paper deals with testing the representations (5,6) separately. The second part deals with the question of whether $\Psi_\oplus = \Psi_{\circ_p} = \Psi$, i.e., whether the psychophysical function, Ψ, appearing in both representations is the same. The third explores the functional form of the unknown functions, Ψ and W, respectively.

---

[7] Namely, monotonicity, solvability, and Archimedeanness (a way of stating that subjectively measured intensities are commensurable).

2 Experiments

Three experiments are presented: Test of jp-symmetry (3) (Exp. 1), the Thomsen condition (7) (Exp. 2), and production commutativity (8) (Exp. 3).

2.1 Respondents

A total of 13 students—graduate and undergraduate—from New York University and University of California, Irvine, plus the author[8], participated in the three experiments of this article; although desireble, for practical reasons, not all respondents particiapted in all the experiments. All respondents reported normal or corrected-to-normal vision. All respondents, except the author, received compensation of $10 per session. Each person provided written consent and was treated in accordance with the "Ethical Principles of Psychologists and Code of Conduct" (American Psychological Association, 2002). Consent forms and procedures were approved by the Institutional Review Boards of New York University and UC Irvine.

2.2 Stimuli

The experiments were carried out in using squares, 10 degrees of visual angle, of monochrome light (RBG channels set to the same DAC value).

2.3 Experimental methods

The experiments reported have a number of testing strategies in common that are now outlined. Other aspects are described later as relevant.

2.3.1 The summation operation, $\oplus$, and matches

The joint presentation (x, u), which can also be written as $x \oplus u$, means that the stimulus x is presented in the left eye and u in the right eye. The goal is to obtain estimates of the subjective brightness match of joint presentation, i.e., finding that stimulus $z \oplus z$ that is perceived equal in brightness to $x \oplus u$, which can be written as $x \oplus u \sim z \oplus z$. Figure 1 describes the process: the four intensities are displayed on a monitor on a zero luminance background; the respondents views these stimuli through a stereoscope that uses mirrors to "fuse" the two halves of the computer monitor resulting in the percepts of $z \oplus z$ above that of the $x \oplus u$. Thus, the summation operation is "implicit" in this method. To accomplish the matching, respondents adjust the intensity of z until they are satisfied that the two percepts are equal in brightness

---

[8] I judged it acceptable to include me, the author, (R8), because the behavioral measures of matching and ratio production are not determined by the knowledge of the experimental design.

Figure 1: Stimulus in brightness matching task: from what is displayed on the monitor to what respondents see. The x, u, z are luminance values. The task: respondents adjust the luminance z until they are satisfied that $z \oplus z$ is equivalent to $x \oplus u$ in brightness.

After the initial presentation, respondents used key presses either to adjust the luminance of z or to indicate satisfaction with the brightness match. Respondents could chose any of four adjustment steps in luminance levels, which were named and presented to respondents as extra-small, small, medium, and large. These corresponded to 1, 2, 4, or 8 steps in the intensity units of the monitor. The increments (positive or negative) were tied to the keyboard keys “a”, “s”, “d”, and “f” for increasing and “;”, “l”, “k”, and “j” for decreasing of luminance. After an adjustment, the screen was set to zero luminance background for 100 ms. and the next trial was presented—subjectively, this was experienced as a blinking and evolved out of pilot studies showing that respondents did not always notice if a new trial had been displayed. This process was repeated until respondents were satisfied with the match, indicated by pressing “b”, at which time the trial ended and z was recorded as the response.

Information about the current block and trial number were displayed in small letters in the upper left corner of the screen.

In verbal instructions to respondents, the task was explained as that of making the upper stimulus equal in brightness to the lower one.

2.3.2 Ratio productions, $\circ_p$

The goal is to obtain an estimate $z \oplus z = (x \oplus x) \circ_p (y \oplus y)$, which in shorthand is $z = x \circ_p y$. In that form the respondent’s task is finding the intensity z such that the respondent perceives the brightness “interval” from the intensity y to z to stand in proportion p to the brightness reference “interval” from y to x.

Figure 2 illustrates the stimulus situation: the reference interval is on the left and the respondent adjusts the intensity of z. Note: here mirrors are not needed as each of the squares of intensity x, z, and y’s are viewed by both eyes simultaneously

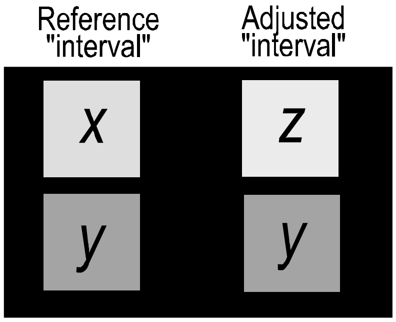

Figure 2: Stimulus in brightness ratio production. The z, y, z are luminance levels and p is a proportion. Respondents adjust the luminance z until they are satisfied that the brightness "interval" between y and z is perceived p times that between y and x..

Adjustments were done as in matching. The value of the proportion p was displayed on upper left side of the monitor.

Instructions to respondents consisted of a description of the task coupled with graphical examples of the form given in Figure 2 and the task was described as making the difference between the brightness of the lower and higher squares in the adjusted interval, e.g., twice ($p = 2$) that between the reference squares on the left. The task was explained using $p = 1/3$ and $p = 2$.

2.3.3 Background luminance

In retrospect, the choice of zero luminance background was likely not the best one in that it mixes scotopic and photopic visual conditions. This has been corrected in experiments currently under way and was incorporated into a few conditions reported here. However, due to the nature of the experimental paradigm of the current experiment, there is little reason to suspect an overly negatively effect on the results.

2.3.4 Procedure

Experiments were conducted in sessions lasting no more than one hour. The initial session was devoted to obtaining written consent, explaining the practice task, and running the practice blocks. The practice task was with a matching task as outlined in Section 1 or a ratio production task as outlined in Section 2. Respondents typically ran two to four sessions per week and no more than one session per day. Because some respondents participated in multiple experiments, the total practice that individuals had prior to any one experiment varied substantially. Depending on the experiment, practiced respondents typically completed around 60 estimates per session, organized into blocks of six or eight estimates. Rest periods were encouraged but their frequency and duration were entirely under the respondents' control.

2.3.5 Equipment

Stimuli were generated with an Apple G4 using PsychToolbox extensions in MATLAB (Pelli, 1997; Brainard, 1997). At New York University, stimuli were presented on a 17" ViewSonic P810 CRT and at the University of California, Irvine, on an 18" Nec Multisync FE 950+ both with a resolution of 1024 × 768 pixels and refresh rate of 75 Hz. The monitors were calibrated by taking repeated measures of luminance at every 5th of the 255 possible luminance increments. Then, using polynomial curve fitting, luminance mapping function for the monitor's luminance steps to actual luminance was established. The calibration was done on both sides of the monitors where stimuli were presented; the luminance disparity between sides was not appreciable. Experiments were conducted in small dark and light-insolated room. Since most of the data were collected at NYU, when not otherwise noted, the data were collected there.

2.3.6 Statistical method and presentation of results

Studies testing behavioral axioms typically examine parameter-free null hypotheses of the form $L_{\text{side}} = R_{\text{side}}$. This reflects the nature of the empirical axioms being tested. If the hypothesis $L_{\text{side}} = R_{\text{side}}$ is correct, it is equivalent to asserting that both L_{side} and R_{side} are drawn from the same distribution. Yet, because there is no theory that predicts the distributions of the estimates, a nonparametric Mann-Whitney U test is chosen for the statistical evaluation, with a significance level of .05.—a practice that has now become common in this situation (e.g., Ellermeier & Faulhammer, 2000; Zimmer, et al., 2001; Ellermeier, et al., 2003; Zimmer, 2005; Steingrimsson & Luce 2005a, 2005b, 2006, 2007).

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

Since we do not have an a priori model of how individuals relate, all data analysis is done on individual data.

2.3.7 Multi-step testing

Testing in Experiments 2 and 3 requires one estimate to be used as stimulus in a subsequent trial. Steingrimsson and Luce (2005a) concluded that in such cases the best result was obtained when all estimates were collected within a session such that each individual estimate was used as the input in the subsequent estimation step. The main advantage of this method is that it allows the variance in the first estimate to propagate through the testing process and it avoids any issue with inter-session variability. See Appendices *A.1,2*, & *4* in Steingrimsson and Luce (2005a) for further details.

2.4 Experiment 1: JP-Symmetry

Steingrimsson and Luce (2005a) found that, in general, the two ears were not symmetric in the sense that jp-symmetry, (3), $x \oplus u \sim u \oplus x$, did not hold in general. At the time, the theory being tested was reinterpretation of Luce's (2000) work in utility theory, in which jp-symmetry was assumed to hold. The psychophysical results led Luce to expand it to include the biased case (Luce, 2002; and further in 2004). Whether jp-symmetry holds or not determines which of somewhat different sets of behavioral properties must be tested. Thus, this is the natural first test to conduct.

2.4.1 Method

Testing the property involves obtaining two types of matches: $x \oplus u = z \oplus z$ and $u \oplus x = z' \oplus z'$ and statically testing whether $z = z'$. These matches are obtained as described in Section 2.3.1.

Three luminances were used: $a = 29.62\ cd/m^2$, $b = 56.25\ cd/m^2$, $c = 91.39\ cd/m^2$. These three intensities gave rise to six ordered stimulus pairs: (a, b), (a, c), and (b, c) corresponding to the left side of (3) and (b, a), (c, a), and (b, c) corresponding to the right side of (3) and three tests of the property. These six matching conditions were all run randomized within a block of trials.

2.4.2 Results

Eight individuals participated and their data are presented graphically in Figure 3. Each graph shows results of the six matching conditions, where, e.g., the matching of (x, u) is labeled xu (note: multiplication is not implied, this is a label only) and the rest analogously.

The statistical hypothesis is that $x, u \sim u, x \Leftrightarrow xu = ux$. Hence, the three statistical hypotheses to be tested are $ab = ba$, $ac = ca$, and $bc = cb$, which are marked on the abscissa. Average luminance level is marked on the ordinate. Sample size is indicated in the upper left portion of each graph.

The result of the statistical test is indicated on the abscissa, above the label of the relevant conditions, with no asterisk denoting failure to reject at the 0.05 level, $\star$ denoting rejection at the 0.05 level, and $\star\star$ at the 0.01 level.

Five respondents (Rs 1, 3, 5, 6, 8) rejected all three of the testing conditions, two (Rs 7, 9) rejected one of the three, and one rejected none (R2). Overall, the property was rejected in 17/24 tests.

For 7 of 8 respondents, one or more of the three conditions were found to be statistically different, with the trend for the remaining conditions consistent with this statistical difference. In all seven cases the pattern of results was that of $(x, u) \succ (u, x)$, which in terms developed by Steingrimsson and Luce (2005a), called a *right bias.* For the remaining respondent the hypothesis of no bias, i.e., $(x, u) \sim (u, x)$, was not rejected.

2.4.3 Discussion

The results echo those obtained in loudness, namely that, in general, jp-symmetry does not hold, but differ in the sense that only right bias was observed here in contrast to a majority exhibiting left bias in the auditory context (Steingrimsson & Luce, 2005a).

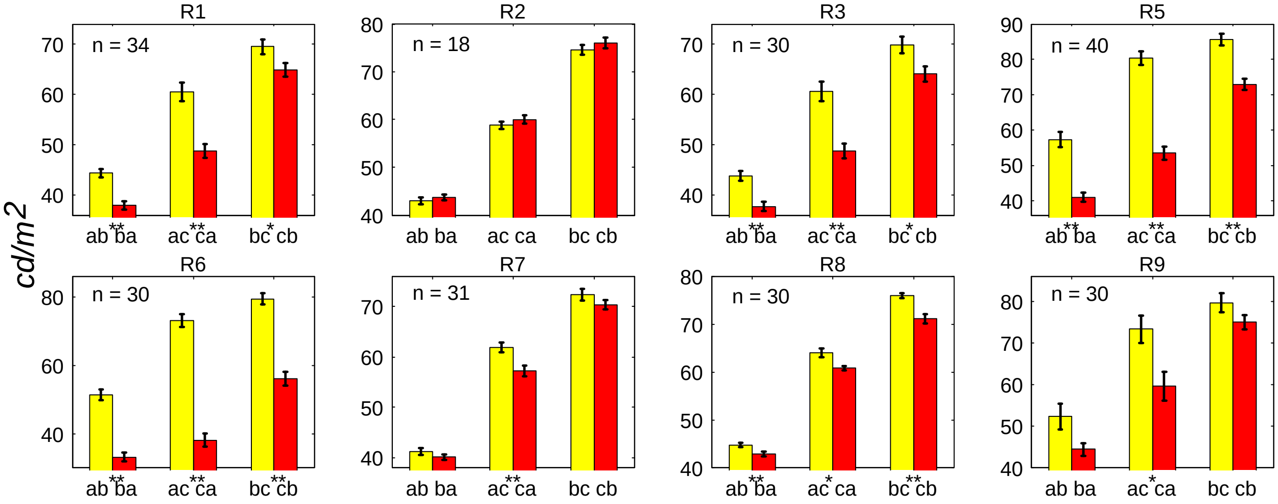

Figure 3: Experiment 1: Result from testing the jp-symmetry property, Eq. (3).

Luce's (2002; 2004) theory admits bias in either direction, but the theory makes no attempt to explain the proportions of people who are left and right biased. Dominance of side (left vs. right) is common human feature. Beyond the familiar handedness, most people exhibit eye dominance but whether such dominance plays a part in the bias observed is an open question. Steingrimsson and Luce (2005a) explored one explanation suggested to them, namely that differences in sensory thresholds of the two ears could explain the bias behavior. They rejected that explanation based on the observation that in normal hearing people, very small changes in energy were involved at the level different sensory threshold compared to the energy in the well-above threshold stimuli used in their experiments. The disproportional energy differences are consequence of loudness growing approximately as the power of intensity. Since the situation is analogous for brightness, it would seem that the same argument obtains her. While understanding the reasons for the bias is of interest, it is not the topic of this paper, nor necessary for any of our conclusions. Thus no further speculation is made with regards to this issue.

As pointed out by Steingrimsson and Luce (2005b), in the unbiased case, one can easily show from (5) that the following two conditions are met:

Commutativity (or symmetry) of $\oplus$:

$$x \oplus u = u \oplus x. \tag{9}$$

Associativity of $\oplus$:

$$x \oplus (y \oplus z) = (x \oplus y) \oplus z. \tag{10}$$

Zimmer et al. (2001) found that associativity was rejected in those cases when commutativity was not rejected. Thus, associativity is an important further test to carry out when trying to decide if a person is actually unbiased. Here, we failed to reject the hypothesis of jp-symmetry in the case of only one subject: R2.. We did not have access to R2 for a follow-up experiment with associativity, but in any future replication of this work, such followup test is recommended for those who do not reject jp-symmetry.

In conclusion, jp-symmetry (3) is not, as a general rule, found to hold for brightness.

2.5 Experiment 2: Thomsen condition

The goal of this experiment is to test the well-known necessary condition of binary additive conjoint representation (Krantz et al., 1971; Michell, 1990, pp. 68–73), the Thomsen condition, (7),

$$\left.\begin{array}{l} x \oplus t \sim v \oplus w \\ v \oplus u \sim z \oplus t \end{array}\right\} \Longrightarrow x \oplus u \sim z \oplus w.$$

To the author's knowledge, no test of the Thomsen condition or the related double cancellation property have been published for brightness. In the auditory literature there are several studies, detailed in Steingrimsson and Luce (2005a), who found the property reasonably sustained. Steingrimsson and Luce (2005a) used a testing method that involved binaural stimuli and three estimation steps in which the intensity was adjusted in one ear only. This feature initially produced somewhat larger than expected variance in respondents' judgment, a feature that went away with additional training. Pilot studies with brightness seemed to replicate this phenomenon from loudness and in response a four step process was devised where the judgments are more balanced in the sense that in the trials feature condition of adjusting luminance in the left eye only, right eye only, and both eyes.

2.5.1 Method

With reference to (7) and the notation of (1), the testing involved obtaining the estimates z', w', y' and y'' using

$$\begin{aligned} (x,t) &\sim (v,w') \\ (v,u) &\sim (z',t) \\ (x,u) &\sim (y',y') \\ (z',w') &\sim (y'',y'') \end{aligned}$$

where the Thomsen condition is said to hold if y' and y'' are found to be statistically equivalent.

All four trial types were run twice within a block in a pseudo-randomized order. Individual estimates of w' and z' were used for subsequent estimates of y''. The stimuli in cd/m^2 were:

Location	x	t	v	u	Background
NYU	44.58	25.31	11.48	69.29	0
UCI	43.11	24.43	10.88	66.91	2.44

Two observations for each estimate were collected within a block of trials in accordance with the method described in Section 2.3.7.

2.5.2 Results

Results for six respondents are displayed in Table 1. In the table, averages, standard deviations, number of observations, and the results of the hypothesis tests are listed for each respondent. Luminance levels are in cd/m^2. For R8, marked with *, the data were collected at UC Irvine.

Resp.	y' (s.d.)	y'' (s.d)	n	p_{stat}	Statistical Conclusion
R3	55.62 (3.66)	57.83 (8.07)	34	.189	$y' = y''$
R4	56.71 (2.30)	57.40 (4.31)	30	.917	$y' = y''$
R6	56.17 (2.15)	55.85 (2.63)	30	.557	$y' = y''$
R8*	52.91 (2.75)	53.97 (4.53)	30	.445	$y' = y''$
R9	51.89 (2.76)	52.13 (5.45)	30	.976	$y' = y''$
R14	53.63 (4.20)	56.88 (9.49)	30	.242	$y' = y''$

Table 1: Experiment 2: Results from testing the Thomsen condition, (7)

The property is found to hold for six of six participants. Note: for five of six respondents $y' < y''$

2.5.3 Discussion

This property, crucial to establish the summation representation, (5), has to the best of our knowledge, never been tested in brightness. However, the question of summation has been discussed a fair bit (see Section C) and the conclusions have varied. The results here suggest that brightness sums over the two eyes in the sense of axiomatic psychophysics. Additivity implies that the brightness percept increases monotonically with luminance. As has often been noted, if one closes one eye, the resulting brightness changes little. This is perhaps a misleading way of intuiting brightness summation of the two eyes. The main problem is that closing one eye puts the visual system in the region where Fechner's Paradox obtains, so perhaps a better way is to think of luminance being held constant in one eye and then steadily increased in the other. In this situation it is vivid that as luminance increases, the resulting brightness percept increases monotonically with luminance—such luminance curves were produced by, e.g., Levelt (1965) showing just such monotonic increase except in the region where Fechner's paradox obtains. What the current results suggest is that this brightness increase is captured by the summation representation (5).

As mentioned, the testing procedure here differed somewhat from that of Steingrimsson and Luce (2005a) with the aim of making it easier to test the property. Yet, in a curious parallel to loudness, Steingrimsson and Luce (2005a) noted that the testing of the Thomsen condition required more training trials than the standard one session for respondents before within-session variability became stable. They offered the possible explanation that, contrary to other experiments, here the intensity is varied in only one eye rather than both. Whatever the explanation may be, pilot studies revealed exactly the same pattern. For this reason, all respondents trained for two sessions before experimental data were collected. Also notable, Gigerenzer and Strube (1983) studied double cancellation and argued that their data showed a bias pattern. Steingrimssion and Luce (2005a) found this bias pattern to disappear with increased training. Here we see $y' < y''$ in 5/6 cases, a coincidence perhaps, but also a suggestion that even more training might be in order. While perhaps an interesting issue for future investigation, this matter is left as a curiosity.

In conclusion, the Thomsen condition appears a reasonably sustained hypothesis in brightness.

2.6 Experiment 3: Production commutativity

For convenience: production commutativity, (8), is given by

$$[(x,x) \circ_p (y,y)] \circ_q (y,y) \sim [(x,x) \circ_q (y,y)] \circ_p (y,y), \text{ where } y < x.$$

In principle, one would want to test the property for a variety of values of p and q. In practice, the results of Steingrimsson and Luce (2007) suggest that the numerical distortion function, W, (6) differs for numbers above and below 1 and that algebraically, nothing simple emerges from looking at the mixed case, e.g, where $p < 1$, $q > 1$. Consequently, in practice the case of $p < 1$, $q < 1$ and $p > 1$, $q > 1$ are consider separately.

2.6.1 Method

The testing requires four estimates in two steps. The first step consists of estimating v and then w in

$$\begin{aligned} (x,x) \circ_p (y,y) &\sim (v,v), \\ (v,v) \circ_q (y,y) &\sim (w,w), \end{aligned}$$

and the second of estimating v' and then w' in

$$\begin{aligned} (x,x) \circ_q (y,y) &\sim (v',v'), \\ (v',v') \circ_p (y,y) &\sim (w',w'). \end{aligned}$$

Production commutativity is considered to hold if w and w' are found to be statistically equivalent. These matches are obtained as described in Section 2.3.2. Two observations for each estimate were collected within a block of trials in accordance with the method described in Section 2.3.7.

The stimuli in cd/m^2 were:

Condition	y (UCI)	x (UCI)	p	q	Background
C_1: $p > 1, q > 1$	11.48 (10.88)	29.62 (28.62)	2	3	0 (2.44)
C_2: $p < 1, q < 1$	11.48 (10.88)	56.25 (54.37)	2/3	1/3	0 (2.44)

2.6.2 Results

Six respondents participated in this experiment. One respondent, who only participated in condition C_1 and R20 sought repeatedly to adjust their response beyond the upper luminance limit of the monitor, making the respondents desired adjustment inaccessible. The remaining data are presented in Table 2

The table lists respondent, condition, the means and standard deviations of w and w', respectively, and results of the statistical tests. For conditions marked with *, the data were collected at UC Irvine.

Resp.	Cond.	w (s.d.)	w'(s.d.)	p_{stat}	n	Statistical Conclusion
R3	C_1	67.05 (18.74)	69.16 (15.00)	.643	32	$w = w'$
	C_2	23.36 (3.41)	22.28 (5.03)	.130	30	$w = w'$
R4	C_1	72.65 (12.56)	69.92 (14.28)	.226	32	$w = w'$
	C_2^*	17.82 (1.57)	19.38 (1.78)	<.001	30	$w \neq w'$
R8*	C_1	59.71 (5.84)	61.78 (5.83)	.189	28	$w = w'$
	C_2	17.74 (3.62)	17.03 (3.07)	.490	28	$w = w'$
R11	C_2	17.39 (6.27)	16.80 (3.73)	.979	32	$w = w'$
R20*	C_2	19.80 (1.35)	20.27 (1.82)	.176	40	$w = w'$
R22*	C_1	70.29 (8.60)	(68.99) (7.24)	.505	30	$w = w'$
	C_2	29.17 (2.620)	29.55 (2.56)	.553	30	$w = w'$

Table 2: Experiment 3: Results from testing the proportion commutativity property, (8).

The property was accepted in 9/10 tasks.[9]

2.6.3 Discussion

Peißner (1999) investigated the related property, threshold-production commutativity, namely,

$$[(x, x) \circ_p (0, 0)] \circ_q (0, 0) \sim [(x, x) \circ_q (0, 0)] \circ_p (0, 0),$$

which is the special case of (8) in which $y = 0$. He used an experimental paradigm and stimuli similar to those employed here, except he used only $p, q > 1$ and found it to hold. The property remains to be tested for $p, q < 1$. Here the generalized proportion commutativity is tested. Given that it is sustained in 9/10 cases, it is concluded that it as reasonably sustained hypothesis for brightness. This conclusion establishes the production representation, (6).

3 Summary, conclusions, discussion, and further work

3.1 Summary and conclusions

The test results are summarized in Table 3.

---

[9] R4's rejection for C_2 seems surprising: the individual passed the same property in audition and also in pilot studies for brightness (different equipment and evolving conditions). Yet, there is no objective reason to collect more experimental data so the result stands.

Exp. #	Name	#R	#Tests	#Fail
1	JP-Symmetry	8	24	17
2	Thomsen condition	6	6	0
3	Proportion. Commutativity.	6	10	1

Table 3: Summary of experimental results

The topic has been a theory of global psychophysical judgments leading to the two representation classes. With a failure of jp-symmetry (3), the asymmetric case, the theory leads to the two representations:

$$\Psi(x,u) = \Psi(x,0) + \Psi(0,u) + \delta\Psi(x,0)\Psi(0,u) \quad (\delta \geq 0), \tag{5}$$

$$W(p) = \frac{\Psi\left[(x,u) \circ_p (y,v)\right] - \Psi(y,v)}{\Psi(x,u) - \Psi(y,v)} \quad [(x,u) \succ (y,v) \succsim (0,0)]. \tag{6}$$

The aim of this article has been the separate testing of the properties derived from the first and second expression. Note, the above Eqs. (5, 6) are reproduced as presented by Luce (2004). However, the results of the present paper can, at best, support (5) and (6) with different psychophysical functions, $\Psi_{\oplus}$ and $\Psi_{\circ_p}$(possibly but not necesserly the same). Although ongoing work, not reported here, strongly supports the hypothesis that $\Psi_{\oplus} = \Psi_{\circ_p}$, the overall conclusion from the three experiments reported is that the summation and production forms of Luce's (2002) theory are separately supported in the brightness domain. This conclusion is identical to that found by Steingrimsson and Luce (2005a) for loudness. Thus, Luce's (2002, 2004) theory seems to offer a certain unification on the level of description of at least these two domains. However, more work is needed before this statement can be made without reservation.

3.2 Further work

As mentioned in Section 1.4, the overarching goal is to replicate, insofar as possible, the work of Steingrimsson and Luce (2005a,b; 2006; 2007) in brightness. The present paper parallels, both in structure and result, Steingrimsson and Luce (2005a). This favorable outcome paves the way for a second paper to parallel Steingrimsson and Luce (2005b). This work, to see if $\Psi_{\oplus} = \Psi_{\circ_p}$in the brightness domain, is currently underway.

Another worthwhile avenue is to extend the work done here to other types of stimuli. Even monitors suited for vision research tend to have a small luminance range compared to that found in natural environments. Yet, as supported by, e.g., Anstis and Ho (1998), isobrightness curves change shape at higher luminances. Also, as shown by Bolonowski (1987) and Bourassa & Rule (1994), summation of Ganzfeld type stimuli appears to produce increased perceived summation than bounded or contoured stimuli. These results do not invalidate the conclusions reached in the present study.

Acknowledgments

This research was supported in part by National Science Foundation grant SBR-9808057 (R. D. Luce, PI) to the University of California, Irvine—any opinion, finding, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation. Additional financial support was provided by the School of Social Sciences and the Department of Cognitive Sciences at UC Irvine. I am especially grateful to Dr. Marisa Carrasco for unfettered access to her laboratory and equipment to carry out most of the research reported here. I thank Dr. Randolph Blake for valuable input on technical matters. And I appreciate many helpful comments of Dr. Joetta Gobell and Dr. R. Duncan Luce on earlier versions.

References

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

Anstis, S., & Ho, A. (1998). Nonlinear combination of luminance excursions during flicker, simultaneous contrast, afterimages and binocular fusion. *Vision Research, 38*, 523–539.

Bolanowski, Jr., S. J.(1987). Contourless stimuli produce binocular brightness summation. *Vision Research, 27*, 1943-1951.

Bourassa, S. J., & Rule, C. M (1994). Binocular brightness: a suppression-summation trade off. *Canadian Journal of Experimental Psychology,.48*, 418–434.

Brainard, D.H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 433-436.

Cliff, N. (1992). Abstract measurement theory and the revolution that never happened. *Psychological Science, 3*, 186-190.

Cohn, T. E., & Lasley, D. J. (1976) Binocular vision: Two possible central interactions between signals from two eyes. *Science, 192*, 561–563.

Curtis, D. W., & Rule, S. J. (1978). Binocular processing of brightness information: A vector-sum model. *Journal of Experimental Psychology: Human Perception and Performance, 4*, 132–143.

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

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

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

Engel, G. R. (1969). The autocorrelation function and binocular brightness mixing *Vision Research, 9*, 1111-1130.

Falmagne, J.-C. (1976). Random conjoint measurement and loudness summation. *Psychological Review, 83*, 65–79.

Falmagne, J.-C., Iverson, G., & Marcovici, S. (1979). Binaural "loudness" summation: Probabilistic theory and data. *Psychological Review, 86*, 25–43.

Fechner, G. T. (1861) Über einige Verhältnisse des binocularen Sehens. *Abhandlungen der mathematisch-physischen Classe der königlich sächsischen Gesellschaft der Wissenschaften, 5*, 337–564.

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. (1997). Cortical dynamics of three-dimensional figureground perception of two-dimensional pictures. *Psychological Review, 104*, 618–658.

Grossberg, S., & Kelly, F. (1999). Neural dynamics of binocular brightness perception. *Vision Research, 39*, 3796–3816.

Irtel, H. (1998). Binocular brightness combination: A mechanism for combining two sources of rather similar information. In W. G. K. Backhaus, R. Kliegl & J. S. Werner (Eds.) *Color Vision: Perspectives from different disciplines*. Berlin: Walter de Gruyterp, 267–274.

Irtel, H. (1986). Experimente zu Fechners Paradoxon der binokularen helligkeit. *Zeitschrift für Experimentelle und Angewandte Psychologie, 33*, 413–422.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement.* Vol. 1. Academic Press, New York.

Lehky, S. R. (1983). A model of binocular brightness and binaural loudness perception in humans with general applications to nonlinear summation of sensory inputs. *Biological Cybernetics, 49*, 89–97.

Legge, G. E. (1984). Binocular contrast summation II. Quadratic summation. *Vision Research, 24*, 385–394.

Levelt, W. J. M. (1965). Binocular brightness averaging and contour information. *British Journal of Psychology, 56*, 1–13.

Levelt, W. J. M., Riemersma, J. B., & Bunt, A. A. (1972). Binaural additivity of loudness. *British Journal of Mathematical and Statistical Psychology, 25*, 51–68.

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

Luce, R. D. (2000).*Utility of gains and losses: Measurement theoretical and experimental approaches.* Erlbaum, Mahwah, N.J., errata: see Luce's web page at http://www.imbs.uci.edu/personnel/luce.

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 asymmmetric matching of joint presentations. *Psychological Review, 111*, 446–454.

Luce, R. D., & Krumhansl, C. L. (1988). Measurement, scaling, and psychophysics. In: Atkinson, R. C., Herrnstein, R. J., Lindzey, G., Luce, R. D. (Eds.), *Stevens' handbook of experimental psychology*, 2nd Edition. Vol. 1 and 2. Wiley, New York, pp. 3–74.

Luce, R. D., & Narens, L.(1993) Comments on the "non-revolution" in the representational theory of measurement. *Psychological Science, 4*, 127-130.

MacLeod, D. I. A. (1972). The Schrodinger equation in binocular brightness combination. *Perception, 1*, 321–324.

Michell, J. (1990). *An Introduction to the Logic of Measurement.* Erlbaum, Hillsdale, N. J.

Narens, L. (1985). *Abstract measurement theory.* Cambridge, MA: MIT.

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

A meaningful justification for the representational theory of measurement. J*ournal of Mathematical Psychology, 46*, 2002, 746—768.

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

Peißner, M. (1999). Experimente zur direkten Skalierbarkeit von gesehenen Helligkeiten. *Unpublished master's thesis*, Universität Regensburg.

Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437–442.

Roberts, F. S. (1979). *Measurement theory.* Vol. 7 of Encyclopedia of Mathematics and Its Applications.Reading, MA: Addison-Wesley.

Schneider, B. (1988). The additivity of loudness across critical bands: A conjoint measurement approach. *Perception & Psychophysics, 43*, 211–222.

Schrodinger, E. (1926). Die Gesichtsempfindungen. In *Mueller–Pouillets Lehrbuch der Physik, Book 2, Part 1* (11th ed., pp. 456–560). Vieweg: Braunschweig.

de Silva, H. R., & Bartley, S. H. (1930). Summation and subtraction of brightness in binocular perception. *British Journal of Psychology, 20*, 242–252.

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. (1967). Brightness function: Binocular versus monocular stimulation. *Perception & Psychophysics, 2*, 451–454

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

Wade, N. J., & Ono, H. (2005). From dichoptic to dichotic: historical contrasts between binocular vision and binaural hearing. *Perception & Psychophysics, 34*, 645–668.

de Weert, C. M. M., & Levelt, W. J. M. (1974). Binocular brightness combinations: Additive and nonadditive aspects. *Perception & Psychophysics, 15*, 551–562.

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

Zimmer, K., Luce, R. D., & Ellermeier, W. (2001). Testing a new theory of psychophysical scaling: Temporal loudness integration. Fechner Day 2001. *Proceedings of the 17th Annual Meeting of the International Society for Psychophysics.* Lengerirch, Germany: Pabst.

Appendices

A Historical context

Psychophysicists typically study the relationship between subjective attributes arising from physical intensity, in particular how these change as a function of changes in intensity. This measurement approach is not only concerned with the assignment of numbers to sensations, but also with the formal properties of the number system into which the observed psychological measures are being mapped. If the enterprize is successful, general laws that systematically relate sensations to physical attributes emerge. This is, e.g., embodied in the famous laws of Fechner and Weber as well as the power law of Stevens.

Central to Stevens' approach was his method of magnitude estimation/production (often called direct measurement) where respondents are instructed to either give numbers in response to stimuli or to produced stimuli that invoked a sensation that are in some prescribed proportion to a standard one. The data collected using these methods suggest that for intensive continua (brightness, loudness, pain, heat, cold, etc.) sensation grows approximately as a power function of physical intensity (Stevens, 1975, for summary and references). These conclusions have almost exclusively involved averaging data over respondents and then fitting the results to function forms, and by evaluating the goodness of fit, the power function form emerged as a well-fitting one.

Also studied is how intensity summates when, e.g., signals are administered independently to the two ears (Levelt, Riemersma, & Bunt, 1972; Falmagne, 1976; Falmagne, Iverson, & Marcovici, 1979; Gigerenzer & Strube, 1983; Schneider, 1988;) or the two eyes (De Silva & Bartley, 1930; Levelt, 1965; Stevens, 1967; Engel, 1969; de Weert & Levelt, 1974; Cohn & Lasley, 1976; Curtis & Rule, 1978; Bolanowski, 1987; Irtel, 1998; Bourassa & Rule,1994; Grossberg & Kelly, 1999; Steingrimsson & Luce, 2005a,b) and how the two modalities compare (Lehky, 1983; Wade & Ono, 2005). Summation has been studied in a variety of ways, which can be said to have in common that some comparison is done between sensation magnitudes when a signal intensity is varied in the two sensory organs. An example is *sensation matching*, which is the special case of ratio production where the proportion is one. Respondents are typically instructed to produce a stimulus equal in subjective intensity to that of a standard. In these studies the standard involves, e.g., projecting lights of different intensities to the left and the right eye respectively and the respondent then *matches* the resulting sensation by adjusting the intensity of two other lights.

Few would deny the tremendous contribution that Stevens made to psychophysics within the area just described and other important activities (e.g., cataloging possible scale types), however, his approach has not been spared criticism either. For present purposes, it suffices to mention that Stevens never seemed to have articulated or tested fundamental assumptions inherent in his magnitude estimation/production tasks. He appeared aware of some oddities that defied clear explanation, e.g., he discussed what he termed a "regression effect," namely that magnitude estimations produced fit to power functions that had slightly different exponents than did magnitude productions. He appeared to have assumed that respondents indeed did give responses that preserved ratios as well as treating numbers in a veridical fashion. And as a final example, he largely ignored individual differences taking them to be "noise" in what he must have assumed was a universal mechanism (this is evident throughout his work, summarized in Stevens, 1975).

While this and similar approaches by those who may be called *scalers*, have produced enormously useful information in a rather simple fashion, the problems just highlighted as well as others not mentioned can be seen as motivation for the approach taken by the so-called *axiomatizers*, in whose form the theory being evaluated here is forged (Luce and Krumhansl, 1988). These methods are detailed in the following section.

B Theoretical background

B.1 The axiomatic measurement approach

The following is a much abbreviated account of the the basic steps of axiomatic psychophysics (for a much more extensive treatment see, e.g., Krantz, Luce, Suppes & Tversky, 1971; Narens, 1985; & Roberts, 1979). The axiomatizers tend to treat the following type of problem:

> If a body of (potential) qualitative observations satisfies certain primitive laws—axioms that capture properties of these observations—then is it possible to find a numerical structure that accurately summarizes these observations? In technical terms, the question is: To which numerical structures is the set of qualitative observations isomorphic? An isomorphism is a one-to-one mapping between structures under which the structure of the one maps into that of the other. It is also desirable to have an explicit process whereby the numerical structure can be constructed from the qualitative one. (Luce & Krumhansl, 1988, p. 5)

The axiomatic approach to this problem can be outlined as follows:

> [M]easurement theory proceeds in the deductive fashion of mathematics: certain formal properties are defined and theorems are proved. These theorems take the form of assertions that, if certain properties are true of the structure in question, then certain conclusions follow as a matter of pure logic. First, there are the primitive relations among attributes that determine both the measurement representations...Most results in measurement theory come in pairs. The first specifies conditions (axioms) under which it is possible to find numerical representation of the qualitative information. In other words, it formulates properties of a qualitative set of observations that are adequate for a certain kind of measurement system or scale to be appropriate. Such a result is called a representation theorem. The second type of result, called a uniqueness theorem, determines how unique the resulting measure or scale is. (Luce & Krumhansl, 1988, p. 7)

It should be noted that while this literature is purely mathematical, the choice of structures to be studied is much influenced by the intended application, a point that is amply exemplified in the current paper.

All measurement, whether within the axiomatic or scaling tradition, begins with a method to study aspects of internal states such as brightness, but one has at present few other options than to ask the respondents for an overt response, e.g., which of two signals produces more or less of the psychological attribute in question (e.g., which is louder, brighter, etc.).[10] For a variety of reasons, repeated questions of these kinds tend to result in variability in responses, or what we take to be errors of measurement which must be dealt with statistically.

---

[10] Recent development in functional magnetic resonance imaging and other such techniques may at some future date make it practical to observe neural activity in such a direct fashion that it becomes a meaningful direct measure of an internal responses; for now we must rely on overt responses.

It should be noted that this approach has not escaped criticism. Perhaps the most serious of these is simply that despite decades of work, it has produced little in terms of practical results (e.g., Cliff, 1992), although Luce & Narens (1993) counter this criticism, pointing to applications in utility theory and elsewhere. What seems clear is that in the field of psychophysics, practical applications of the axiomatic measurement approach have begun to emerge. The progress is both in theoretical papers that set forward directly testable behavioral axioms (e.g., Narens, 1996, 2002, 2006; Luce, 2002, 2004) as well as empirical papers that report results of such tests (e.g., Peißner, 1999; Ellermeier & Faulhammer, 2000; Zimmer, Luce, & Ellermeier, 2001; Ellermeier, Narens, & Dielmann, 2003; Zimmer, 2005; Steingrimsson & Luce 2005a, 2005b, 2006, 2007; current paper).

B.2 Recent directly relevant developments in axiomatic measurement

Narens (1996) set out what he thought had to be the implicit assumptions behind Stevens' magnitude estimation/production methods. The first was that respondents treated numbers in a veridical fashion. That is, if W denotes a function that describes a respondent's interpretation of numbers p, then Narens showed that Stevens must have assumed $W(p) = p$. He formulated a behavioral axiom, which was equivalent to a slightly weaker demand, namely the power form $W(p) = p^b$. This property has been unambiguously rejected in audition (Ellermeier & Faulhammer, 2000; Zimmer, 2005; Steingrimsson & Luce, 2007) and in brightness (Peißner, 1999).

A second result of Narens (1996) was a behavioral axiom equivalent to measurements done using ratio productions being on a subscale of a ratio scale. This property has been tested for both loudness and brightness and has been generally sustained (Peißner, 1999; Ellermeier & Faulhammer, 2000; Zimmer, 2005). Luce (2002) extended this result of Narens (1996) to what he called generalized ratio production and formulated the equivalent (subjective) production commutativity, (8), which was sustained in loudness by Steingrimsson and Luce (2005a). It is tested here in brightness as Experiment 3.

These results made clear that there are two unknown functions in play. First, the psychophysical one, which is called Ψ, which maps stimuli into sensations and the second one, W, which maps numerical instructions into their cognitive equivalents. The following section summarizes relevant theoretical developments of Luce (2002, 2004), which led to separate representations for each of these two functions, (5,6).

C Relation to previous work

The task of comparing current results to the existing literature would be a rather arduous one were it not for the paper of Grossberg and Kelley (1999). There is no compelling reason to more than summarize their main conclusions; the reader interested in more details is advised to start with Grossberg and Kelly (1999).

Grossberg and Kelly (1999) evaluated and created a taxonomy of all the binocular summation models they found to date (see their Table 1) and evaluate them against existing data. In addition they present an updated version of the FACADE theory (Grossberg, 1997).

The history of brightness summation has ranged from whether it summates at all to how different stimuli produce different percepts. In a 1930 paper, De Silva and Bartley wrote:

> Numerous results gathered from observations of stereoscopic phenomena such as retinal rivalry support the view that alteration of conditions upon one retina always exerts some influence on the functioning of the other retina. Consequently it is rather surprising that Fechner, Sherrington, Abney and Watson, Dawson and others have claimed to have proved by experiment that there is no binocular summation of brightness...as the weight of argument from the standpoint of numbers of investigations, prestige of investigators and use of refined apparatus is decidedly against it, binocular integration as regards brightness has come to be a generally discredited fact in the literature.
>
> It was the purpose of this investigation to attempt to put this problem of binocular summation of brightness to a critical test, and in particular to determine if possible why the experimental results should be so flatly contradictory. (pp. 242–243)

The author cannot help but find it somewhat ironic, some 77 years later, to be bringing up this question of brightness summation anew. Suffice it to say that the conclusion of De Silva and Barley, the aggregate conclusions of Grossberg and Kelly's (1999) review of additional literature, and the conclusion of the current paper, is that outside the bounds of the Fechner paradox, brightness summates but that the effect is generally small.

Large summation effects have only been asserted for stimuli consisting of Ganzfelds (stimuli of uniform luminance that cover the entire visual field). Bolonowski (1987) reported, using magnitude estimation in flashed Ganzfeld conditions that "complete binocular brightness summation occurs." Bourassa & Rule (1994) reported two experiments, where the first used Ganzfeld conditions and the second used smaller targets with very low spatial frequencies. They found Ganzfeld stimuli "produced a large amount of binocular brightness summation and very little Fechner's paradox" whereas their smaller low-frequency stimuli "produced greater Fechner's paradox than the Ganzfelds, but more binocular summation and less Fechner's paradox than what is usually reported for small targets with abrupt contours."

Grossberg and Kelly (1999) identified 13 models, including the one they proposed. Four of those they classified as "eye-weighting", three as "vector summation", and six as "neural networks". Grossberg and Kelly (1999, Section 5) meticulously dissected each of the proposed models and found a number of limitations which are applicable to the current discussion. Grossberg and Kelly (1999) argued that the eye-weighting models of Levelt (1965) and Engel (1969) suffered from using weights that do not allow for binocular summation. They further argued that a model introduced by de Weert and Levelt (1974) made predictions that are at odds with data. Their main problem with Irtel's (1986) model was that it did not seem to extend to Ganzfelds. This last argument seems too weak to reject the model outright. However, from the point of view of the current results, Irtel's (1986)[11] model relies on an invariance condition which effectively assumes the two eyes are identical, an assumption rejected here in Experiment 1. Whether that is an easily fixable problem is not something that is addressed here.

Grossberg and Kelly (1999) discussed several vector summation models (Shcrodinger, 1926; MacLeod, 1972; Curtis and Rule, 1978; Legge, 1984) and found flaws with them all, mostly related

---

[11] For a reference in English, see Irtel (1998). However, note a typo where $\varepsilon(tx)$ and $m(tx)$ at threshold should equal $\varepsilon(x)$ and $m(x)$, respectively (Irtel, 2007, personal communication).

to their extendability to various stimulus conditions such as Fechner's paradox. The latter criticism could also be leveled at the model tested here. Conversely, it seems going too far to reject these models on the grounds that they have limitations, especially when it is untested whether those limitations may be overcome as well as considering that stimulus conditions under which Fechner's paradox obtains is not likely in natural conditions. In contrast, from the point of view of the current work, it seems that none of the models explicitly treat the biased case, i.e., that the two eyes are different. It is not clear how hard it would be to amend these models to admit this empirical result, but for now this appears as a flaw in the current context.

The rest of the models identified and discussed by Grossberg and Kelly (1999) are neural network models. While Grossberg and Kelly (1999) also found limitations in all the neural network models they evaluated, the main issue is, again, that these models treat the eyes as unbiased. It is not the intent here to reject every model proposed that reasonably accounts for data on the basis that they do not admit the biased case, a situation that may well be easily remedied. Rather, the contrast drawn here is in the modeling approach. Luce (2002, 2004) takes an approach that in Grossberg and Kelly's (1999) framework would define a new category, namely, an axiomatic model. Steingrimsson and Luce (2005a) articulated the main conceptual difference as follows:

> Two further points should be stressed. First, [Luce's (2002, 2004]...theory is not domain specific in the sense that it can, in principle, apply to any intensive dimensions (e.g., loudness, brightness, or [heaviness]). Second, although neuronal activity ultimately underlies perception, the approach taken here is entirely behavioral. The important abstraction is that the results we obtain are valid answers to our questions regardless of what the neural machinery may be, only the behavior matters. In effect, we could use the very same approach to an alien life form or a robot. Consequently, we do not make any attempt to draw conclusions about the biological workings of the perceptual system from our results. (p.291).

For this reason, comparing neural network models to the model discussed here is not entirely apt.