Journal of Mathematical Psychology 55 (2011) 379–385

Contents lists available at SciVerse ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

Theory and tests of the conjoint commutativity axiom for additive conjoint measurement[☆]

R. Duncan Luce, Ragnar Steingrimsson [*]

Institute for Mathematical Behavioral Sciences, University of California, Irvine, CA 92697-5100, United States

ARTICLE INFO

Article history:
Received 17 January 2011
Received in revised form
25 May 2011
Available online 25 June 2011

Keywords:
Double cancellation
Thomsen condition
Commutativity rule
Conjoint additivity
Conjoint commutativity
Matching
Loudness
Brightness
Additivity
Measurement

ABSTRACT

The empirical study of the axioms underlying additive conjoint measurement initially focused mostly on the double cancellation axiom. That axiom was shown to exhibit redundant features that made its statistical evaluation a major challenge. The special case of double cancellation where inequalities are replaced by indifferences – the Thomsen condition – turned out in the full axiomatic context to be equivalent to the double cancellation property but without exhibiting the redundancies of double cancellation. However, it too has some undesirable features when it comes to its empirical evaluation, the chief among them being a certain statistical asymmetry in estimates used to evaluate it, namely two interlocked hypotheses and a single conclusion. Nevertheless, thinking we had no choice, we evaluated the Thomsen condition for both loudness and brightness and, in agreement with other lines of research, we found more support for conjoint additivity than not. However, we commented on the difficulties we had encountered in evaluating it. Thus we sought a more symmetric replacement, which as Gigerenzer and Strube (1983) first noted, is found in the conjoint commutativity axiom proposed by Falmagne (1976, who called it the "commutative rule"). It turns out that, in the presence of the usual structural and other necessary assumptions of additive conjoint measurement, we can show that conjoint commutativity is equivalent to the Thomsen condition, a result that seems to have been overlooked in the literature. We subjected this property to empirical evaluation for both loudness and brightness. In contrast to Gigerenzer and Strube (1983), our data show support for the conjoint commutativity in both domains and thus for conjoint additivity.

1. Background

Several recent studies (Luce, 2004, 2008; Steingrimsson, 2009; Steingrimsson & Luce, 2005a) have focused on whether or not subjective measures of intensity over the two ears or two eyes satisfy the axioms for additive conjoint measurement which lead to an additive numerical representation. A great deal of the relevant literature such as Luce and Tukey (1964) was summarized in the *Foundations of Measurement* (FofM) (Krantz, Luce, Suppes, & Tversky, 1971; Luce, Krantz, Suppes, & Tversky, 1990; Suppes, Krantz, Luce, & Tversky, 1989). The gist of that literature is that, in the presence of various axioms (summarized in Definition 7, p. 256, FofM I), additivity is forced if either of two properties is satisfied. Suppose that A and P are sets and that there is an ordering $\succsim$ over $A \times P$. For $a, b, c \in A$ and $p, q, s \in P$, the first property, called the axiom of *double cancellation*, is

$$(a, p) \succsim (b, q) \quad \text{and} \quad (c, q) \succsim (a, s) \tag{1}$$
$$\Rightarrow (c, p) \succsim (b, s). \tag{2}$$

The second axiom, called the *Thomsen condition*, is the special case of double cancellation where $\succsim$ is replaced by $\sim$:

$$(a, p) \sim (b, q) \quad \text{and} \quad (c, q) \sim (a, s) \tag{3}$$
$$\Rightarrow (c, p) \sim (b, s). \tag{4}$$

Gigerenzer and Strube (1983) pointed out why the considerable redundancy of double cancellation (see below) makes it a very unsatisfactory axiom to study empirically. Although the Thomsen condition avoids that redundancy, it necessarily exhibits a considerable statistical asymmetry (see below) which makes for an inference difficulty.

The article is structured as follows.

- We provide an overview of current tests of conjoint additivity and an expanded discussion of the problems with testing double cancellation and the Thomsen condition.
- We provide a presentation of the conjoint commutativity property and a demonstration of how it offers a possible solution to those problems.

[☆] Portions of this material have also appeared in Conference Proceedings of the Meeting of International Society for Psychophysics (Steingrimsson & Luce, 2010), in which we referred to Conjoint Commutativity as the Commutativity Rule.

[*] Corresponding author.
E-mail address: ragnar@uci.edu (R. Steingrimsson).

0022-2496/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmp.2011.05.004

- We provide empirical evaluations of the conjoint commutativity for both loudness and brightness.

A general comment: We are very aware that the measurement approach we take here is not currently fashionable, having been "replaced" by various process models. Unlike the measurement models for which the behavioral assumptions are directly testable, the process models are composed of unobservable, hypothetical mechanisms. We feel that the added flexibility of process models comes at the (usually unacknowledged) very high cost of unobservable mechanisms which, to this day, has not really been resolved by such imaging techniques as fMRI. And we feel that the very successful approach of four centuries of classical physics has not been given anything like a comparable effort in psychology. The first author has devoted the last 12 years of his career attempting to apply our knowledge of measurement to developing both psychophysical and utility measurement models, and collaborating with the second author and others he has focused on experimental studies suggested by these models.

2. Current tests of conjoint additivity

Additivity over sense organs has been studied in a variety of ways; here, we focus on its axiomatic evaluation. As summarized in Definition 7 of FofM I (p. 256), the test of additivity involves the evaluation of either double cancellation or the Thomsen condition. Double cancellation was explored by Levelt, Riemersma, and Bunt (1972), Falmagne (1976), Falmagne, Iverson, and Marcovici (1979), and Gigerenzer and Strube (1983) for loudness, by Schneider (1988) for loudness across critical bands, and by Legge and Rubin (1981) for perceived contrast; Ward (1990) evaluated double cancellation for cross-modal additivity of loudness and brightness. The Thomsen condition was evaluated by Steingrimsson and Luce (2005a) for loudness and by Steingrimsson (2009) for brightness. Of these studies, two rejected double cancellation (Falmagne, 1976; Gigerenzer & Strube, 1983).

Gigerenzer and Strube (1983) articulated clearly that double cancellation involves the following type of redundancy. Suppose the signals are a, b, c, p, q, s for which double cancellation holds:

$$(a, p) \succsim (b, q) \quad \text{and} \quad (c, q) \succsim (a, s) \Rightarrow (c, p) \succsim (b, s).$$

Then, for signals p' and s', with $p' \succ p$ and $s \succ s'$, we have, by monotonicity, another version of double cancellation:

$$(a, p') \succsim (b, q) \quad \text{and} \quad (c, q) \succsim (a, s') \Rightarrow (c, p') \succsim (b, s').$$

For example, they reported that "…no less than 82.64% of Levelt et al.'s data were defined a priori" in the sense that of a complete factorial design they selected a subdesign from which monotonicity and transitivity implied the rest of the data.

So, in our empirical studies (Steingrimsson, 2009; Steingrimsson & Luce, 2005a), we focused instead on testing the non-redundant Thomsen condition:

$$(a, p) \sim (b, q) \quad \text{and} \quad (c, q) \sim (a, s) \Rightarrow (c, p) \sim (b, s).$$

For both double cancellation and the Thomsen condition some of the signals are chosen by the experimenter and the rest by the respondent, as we next outline for the Thomsen condition.

In the experiments we have conducted, a stimulus pair (x, u) is taken to mean that an intensity x is presented to the left sensory organ, ear or eye, and intensity u is presented simultaneously to the right sensory organ, ear or eye. A match such as $(a, p) \sim (b, \mathbf{q})$ is one where (a, p) is judged to possess the same subjective intensity as does $(b, \mathbf{q})$, i.e., (a, p) seems equally loud or equally bright as $(b, \mathbf{q})$. An empirical match is one in which the experimenter selects a, b, and p and a respondent adjusts the intensity $\mathbf{q}$ in some way until a match is achieved. The $\mathbf{q}$ is in bold face to emphasize that it is produced by a respondent and so varies over repeated presentations.

In this fashion, the Thomsen condition can be empirically evaluated in several different ways. One is to start with a, b, p, s and first estimate $\mathbf{q}$ such that

$$(a, p) \sim (b, \mathbf{q}). \tag{5}$$

Next, estimate $\mathbf{c}$ such that

$$(\mathbf{c}, \mathbf{q}) \sim (a, s). \tag{6}$$

And finally, estimate $\mathbf{c}'$ such that

$$(\mathbf{c}', p) \sim (b, s). \tag{7}$$

The empirical test is whether or not

$$\mathbf{c} \sim \mathbf{c}'. \tag{8}$$

A source of difficulty is that, in any straightforward testing method, the estimate $\mathbf{c}$ rests upon the estimate $\mathbf{q}$, and so it has two sources of variability/bias, whereas $\mathbf{c}'$ has only one source of variability/bias, where a bias may, for example, be a time-order error (arising from sequential presentation of stimulus pairs which is inevitable in the auditory case).

Another possible source of difficulty relates to the fact that, in estimating $\mathbf{q}$, $\mathbf{c}$, and $\mathbf{c}'$, typically the stimulus in one sensory organ is adjusted while the stimulus in the other is held constant. In audition, this manipulation is experienced as a subjective change in the localization of a tone, whereas in brightness it is has no direct subjective correlate. The fact is that Steingrimsson and Luce (2005a) found for loudness matches and Steingrimsson (2009) for brightness matches that the respondents required up to three times the amount of practice with this task before their judgments stabilized as compared to matches with equal adjustments in both sensory organs. That may be attributable to the difference between the compound estimate and the simple one.

3. Equivalence of conjoint commutativity to the Thomsen condition

As first pointed out by Gigerenzer and Strube (1983), in an article on random conjoint measurement, Falmagne (1976) introduced the following property, which he called the *commutativity rule*, and which we make more specific by speaking of *conjoint commutativity*. With the function[1] $m_{p,q}$ defined by

$$b = m_{p,q}(a) \quad \text{iff } (a, p) \sim (b, q), \tag{9}$$

conjoint commutativity asserts that

$$m_{p,q}[m_{r,s}(a)] = m_{r,s}[m_{p,q}(a)]. \tag{10}$$

Note that, although Falmagne applied conjoint commutativity to random conjoint measurement, this property applies equally well to the algebraic case. As far as we know, the literature has, with the notable exception of Gigerenzer and Strube (1983), mostly overlooked the fact that conjoint commutativity can equally well play the role of the Thomsen condition, and has some advantages over it.

Consider the part of Definition 7 of a binary conjoint structure on p. 256 of Krantz et al. (1971) given by the following assumptions.

A1 Weak ordering.
A2 Monotonicity (= independence).
A3 The Thomsen condition.

[1] One can also use an operator notation instead of the function notation used by both Falmagne and us.

A4 Unrestricted solvability.[2]
A5 Archimedeaness.
A6 Each component is essential.

Our main focus is on showing that we may use conjoint commutativity, which is statistically symmetric, instead of the statistically asymmetric Thomsen condition. Assumptions A1, A2, and A5 are easily seen to be necessary properties that the ordering must satisfy if there is an additive representation, and Assumptions A4 and A6 are two non-necessary structural assumptions. Because the proof does not make explicit use of the Assumptions other than A3, we do not restate them here.

Theorem 1. *Under Assumptions* A1, A2, A4, A5 , *and* A6, *the following four properties are equivalent, and so either* (i), (ii), *or* (iii) *can play the role of* A3

(i) *Double cancellation.*
(ii) *Thomsen condition.*
(iii) *Conjoint commutativity (*Falmagne, 1976*).*
(iv) *An interval scale, additive conjoint representation.*

The proof is given in the Appendix.

A clear advantage of conjoint commutativity is that it is statistically symmetric in the sense that each side of (10) entails two respondent selected signals, which is not the case for either double cancellation or the Thomsen condition. It gains that at the expense of having successive estimates on each side of the null hypothesis. It does not, however, address the potential issue that may arise in empirically producing a match by adjusting the intensity in one sensory organ while keeping the intensity constant in the other organ. (Note: we will nevertheless attempt to address this final issue methodologically by varying the sensory organ in which the variable tone is presented).

Conjoint commutativity has, to the best of our knowledge, been empirically evaluated only by Gigerenzer and Strube (1983). They evaluated a probabilistic version and rejected it for loudness. Here, conjoint commutativity, (10), is evaluated in an algebraic form in two domains, loudness and brightness, presented as Experiment 1 and Experiment 2, respectively.

3.1. General method

The experiments have several common testing features, which now outline.

3.1.1. Respondents

A total of nine students at the University of California, Irvine, and one coauthor[3] participated in the two experiments. All respondents who provided loudness data reported normal hearing and those who provided brightness data reported corrected-to-normal vision. Except for the coauthor, each respondent received $12 per session. Each person provided written consent and was treated in accord with the "Ethical Principles of Psychologists and Code of Conduct" (American Psychological Association, 2002). Consent forms and procedures were approved by UC Irvine's Institutional Review Board.

3.1.2. Notational convention

Sound intensities are reported in dB SPL (dB for short) and light intensities in cd/m^2. The theory, however, is cast in terms of intensity increments above threshold intensity. Therefore, with a threshold of x_τ for the left ear/eye and a threshold of u_τ for the right ear/eye, the effective stimulus (x, u) has $x = x' - x_\tau$ and $u = u' - u_\tau$, where (x', u') are the actual intensities presented. However, because all signals were well above threshold and the respondents were selected for their normal hearing/vision, the error in reporting intensities (x', u'), using dB or cd/m^2, is negligible.

3.1.3. Statistical methods and presentation of results

The goal of our experiments is to evaluate our evidence for parameter-free null-hypotheses that have the generic form $L_{\text{side}} = R_{\text{side}}$. Since we have no a priori model of how individuals relate, all data analysis is done on individual data (e.g., Luce, 1995, p. 20). And, because we have no a priori model of the distribution of the data, we use the non-parametric Mann–Whitney U (M–W U) test at the 0.05 level. This is the choice of numerous other papers (e.g., Falmagne (1976); Gigerenzer and Strube (1983); Ellermeier and Faulhammer (2000); Zimmer, Luce, and Ellermeier (2001); Ellermeier, Narens, and Dielmann (2003), Zimmer (2005); Steingrimsson and Luce (2005a); Steingrimsson and Luce (2005b); Steingrimsson and Luce (2006); Steingrimsson and Luce (2007); Steingrimsson (2009); Luce, Steingrimsson, and Narens (2010); Steingrimsson (2011)).

As the estimation steps are made in discrete steps (see the Procedure section) and the estimates appear reasonably symmetric, the mean is known to be a good estimate for the median. In loudness, we report and collect data using dB values, whereas in brightness the stimuli and data are in LUT (the monitor's video card Look Up Table of integer values 0–255) values, but are reported in units of cd/m^2. This involves a power transform, and so we report the transformed mean LUT values as well as the normalized standard deviations (maintaining the relative magnitude vis-a-vis the mean in cd/m^2).[4]

3.1.4. Procedure

Empirical evaluation of the conjoint commutativity, (10), involves obtaining several matches of the generic form $(x, u) \sim (\mathbf{z}, v)$, where the $\mathbf{z}$ is under respondent control and x, u, and v are signals selected by the experimenter. The general procedure is a variation on the method of adjustment in which the respondent is free to adjust the intensity of z up or down as often as desired until satisfied with the match, which is called $\mathbf{z}$. Concretely, respondents could choose any of four changes in intensity described as extra-small, small, medium, and large. In the loudness case, they corresponded to .5, 1, 2, or 4 dB, and in the brightness case they corresponded to luminance steps of 1, 2, 4, or 8 LUT values. Following an adjustment, the stimuli were next presented with the requested adjustment included. The respondent repeated this process until satisfied with the match, which was indicated by pressing a different key, after which the next matching task commenced.

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

The following four matches were needed to evaluate the conjoint commutativity

$$m_{p,q}[m_{r,s}(a)] = m_{r,s}[m_{p,q}(a)].$$

[2] Definition 7, FofM I only needed restricted solvability, but to be sure that we can always deal with the constructions for conjoint commutativity, we need the unrestricted version of solvability.

[3] This we judged acceptable because knowledge of the experimental design does not change the sensations on which the behavioral tasks of matching are based. The coauthor is numbered R22.

[4] We thank Dr. J. Yellott for this suggestion.

1. $(a, r) \sim (\mathbf{b}, s)$: The respondent produces $\mathbf{b} = m_{r,s}(a)$.
2. $(\mathbf{b}, p) \sim (\mathbf{d}, q)$: The intensity $\mathbf{b}$ is obtained in step 1, the respondent produces $\mathbf{d} = m_{p,q}(\mathbf{b})$.
3. $(a, p) \sim (\mathbf{c}, q)$: The respondent produces $\mathbf{c} = m_{p,q}(a)$.
4. $(\mathbf{c}, r) \sim (\mathbf{e}, s)$: The intensity $\mathbf{c}$ is obtained in step 3, the respondent produces $\mathbf{e} = m_{r,s}(\mathbf{c})$.

The property is said to hold when the hypothesis that $\mathbf{d} \sim \mathbf{e}$ is not rejected.

In steps 1–4, note that all adjustments are made in the left sensory organ. The property can equally well be evaluated by its mirror image in which $\mathbf{b} = m'_{p,q}(a)$ iff $(p, a) \sim (q, \mathbf{b})$, and the conjoint commutativity is then given by $m'_{p,q}[m'_{r,s}(a)] = m'_{r,s}[m'_{p,q}(a)]$. The four matches required for these are referred to as steps 5–8, and are analogous to steps 1–4.

Note: If the sensations evoked from physically identical inputs to the two ears/eyes were identical, then they would be behaviorally interchangeable. However, such symmetry has been unequivocally rejected for both loudness (Steingrimsson & Luce, 2005a) and brightness (Steingrimsson, 2009). This means that $m'_{p,q}(a) \neq m_{p,q}(a)$, and thus m and m' constitute different experimental conditions. An additional empirical result of Steingrimsson and Luce (2005a) was that this non-symmetry of the ears was not behaviorally constant but could be affected by, for example, sustained matching in one of the two ears within a session, a result Steingrimsson and Luce (2006) described using a form of a filtering model.

One plausible consequence of Steingrimsson and Luce's (2006) filtering model is that a bias formed by consistently matching a tone in one ear may be counteracted by mixing within a block of trials the sensory organ to which the variable stimulus is presented. This can be accomplished by mixing the trials of steps 1–4 and 5–8 within a block, and hence we did that. This plausibly addresses, by experimental methodology, the remaining problem with evaluating additivity.

The eight matches (1–8) were run in a block of trials generating two tests of the conjoint commutativity. Respondents typically completed 10 blocks per (one-hour) session. Thus, in addition to a practice session, three experimental sessions were required to obtain the typical 30 estimates collected for each matching condition.

3.2. *Experiment 1: Loudness*

3.2.1. *Method*

Stimuli and equipment: The tones, x to the left ear and u to the right, were 1 kHz sinusoids of 100 ms duration which included 10 ms on and 10 ms off ramps. For a match such as $(x, u) \sim (\mathbf{z}, v)$, the two joint presentations were separated by 450 ms. The stimuli were generated digitally using a personal computer and played through a 24-bit digital-to-analog converter (RP2.1 Real-time processor, Tucker–Davis Technology). The intensity and frequency were controlled using a programmable interface for the RP2.1, and stimuli were presented over Sennheiser HD265L headphones to the respondent seated in an individual, single-walled IAC sound booth located in a quiet laboratory room. A safety ceiling of 90 dB was imposed in all experiments.

Loudness matching and the stimulus conditions: For the loudness match $(x, u) \sim (\mathbf{z}, v)$, the initial intensity for $\mathbf{z}$ is chosen by the experimenter at random in a 10 dB interval around a best guess for its final estimate. Listed in Table 1 are the stimulus values for the two conditions under which conjoint commutativity was evaluated in loudness.

3.2.2. *Results and discussion*

The results are summarized in Table 2. For each respondent, the columns are the means and standard deviations for $\mathbf{d}$ and $\mathbf{e}$, and the number of observations, n, for each sample. The statistics column reports the M–W U test of the null hypothesis $\mathbf{d} \sim \mathbf{e}$, given as $p_{\mathbf{d}\sim\mathbf{e}}$.

Table 1
The table lists the two stimulus instantiations under which the conjoint commutativity was tested, which, when applied to both m and m', resulted in four testing conditions.

Condition	Stimuli (dB)				
	a	r	s	p	q
C_1	64	70	67	66	58
C_2	60	66	67	68	62

For R81, the C'_1 condition fails. Otherwise, the property is found to be supported in 13/14 tests. We regard this to be reasonably strong evidence in favor of the property, and, as a consequence, evidence favoring an additive conjoint representation in loudness. This result mirrors the conclusion of Steingrimsson and Luce (2005a) about additivity, but is at odds with the results of Gigerenzer and Strube (1983), who evaluated double cancellation and conjoint commutativity and rejected both. Steingrimsson and Luce (2005a) discussed in some detail the discrepancy between the double cancellation and their favorably evaluated Thomsen condition. In comparing the test of conjoint commutativity, two major differences emerge (these were also discussed by Steingrimsson and Luce (2005a)). First, Gigerenzer and Strube (1983) used an experimental technique in which the estimated median of the observations in the first estimate, here $\mathbf{b}$, was used as an input in a subsequent evaluation in the second step (obtaining $\mathbf{d}$). This has three consequences: namely, any bias in the estimate of the median will pass on into the second estimation step, the first part of a session is devoted to the first step and the second part to the second step, risking inter-session bias (frequently observed), and the variance from the first step does not accumulate into the second step. Here, we avoid all three by using each estimated instance of $\mathbf{b}$ as input in the following step (see trial types 2 and 6). This methodology allows all trial types to appear within a block of trials, reduces the risk of step one being based on a bias in median estimation, and preserves variability through both estimation steps. Indeed, the second major difference is that Gigerenzer and Strube (1983) found a statistically significant bias such that $\mathbf{d} > \mathbf{e}$ (and similarly for double cancellation) whereas we do not observe such a bias in our results. Finally, it may be important that Gigerenzer and Strube (1983) collected fewer observations of $\mathbf{e}$ and $\mathbf{d}$ (20) than of $\mathbf{b}$ and $\mathbf{d}$ (41) – this they did to fit data collection within a session – whereas we collected equal numbers of both, and typically 30 or more, and, due the mixing of all trials in a block with a session consisting of typically 8–10 blocks, we could pool data across sessions, which, under the assumption of uniform variance within a session, is accounted for by the use of the M–W U test statistic. Finally, we used tones of 100 ms in duration all at 1 kHz, whereas Gigerenzer and Strube (1983) used 4 s presentation of reference pairs and tones of 200 Hz and 6 kHz, respectively. All these differences are both substantive and substantial enough plausibly to account for the difference in results.

3.3. *Experiment 2: Brightness*

3.3.1. *Method*

Stimuli and equipment: The experiment was conducted in a dark room, and each respondent received a minimum of 10 min of dark adaptation. The stimuli were generated by a personal computer and displayed on a monitor (Eizo RadiForce RX320) with an automatic luminance uniformity equalizer and black-light sensor to compensate for luminance fluctuation caused by ambient temperature and passage of time as well as build in gamma correction. The diagonal size is 54 cm, the maximum

Table 2
Results of Experiment 1: Test of the conjoint commutativity for loudness. Listed for each respondent are the conditions tested, means (M) and standard deviations (SD) of the results, number of observations (n) obtained for each condition, and finally the results of the statistical testing.

Condition	Respondent	Intensity (dB)					M–W-U
		d		**e**			
		M	SD	M	SD	n	$p_{\mathbf{d}\sim\mathbf{e}}$
C_1	R10	70.17	1.27	69.85	1.50	30	.429
C_1'		67.42	1.11	67.08	1.13		.473
C_1	R22	72.43	1.52	72.97	1.59	30	.195
C_1'		72.75	1.64	72.8	1.33		.929
C_1	R80	70.02	1.34	69.4	1.50	30	.151
C_1'		68.73	1.56	68.52	1.86		.542
C_1	R81	69.67	1.19	69.85	1.59	30	.743
C_1'		69.65	1.37	68.92	1.23		.012
C_1	R85	67.55	1.89	67.85	1.92	30	.562
C_1'		66.18	2.03	65.68	2.62		.364
C_2	R86	63.97	2.68	63.73	3.15	29	.557
C_2'		66.19	2.08	65.41	2.64		.127
C_1	R88	67.62	3.14	66.74	4.16	36	.384
C_1'		69.26	4.36	68.47	5.00		.484

Table 3
The table lists the three stimulus instantiations under which the conjoint commutativity was tested. This, when applied to both m and m', yields a total of six testing conditions.

Condition	Stimuli in cd/m^2				
	a	r	s	p	q
C_1	93.7	153.5	115.7	74.3	57.2
C_2	140.2	264.9	197.2	115.7	93.7
C_3	140.2	197.2	167.4	115.7	93.7

resolution is 1536 × 2048, and the maximum luminance is 742 cd/m^2. Luminance measures were taken using Photo Research's PR-670 SpectraScan Spectroradiometer, which verified the monitor calibration and determined the luminance of the stimuli. Information about the current block and trial number were displayed in small letters in the upper left corner of the screen.

Brightness matching and stimulus conditions: The stimulus (x, u) means a joint presentation of a light with intensity x in the left eye and a light with intensity u in the right eye. These lights are achromatic squares subtending 10° of visual angle presented on a uniform background of 4 cd/m^2.

To obtain the brightness match $(x, u) \sim (\mathbf{z}, v)$, the experimenter chooses x, u, and v, and the respondent produces the $\mathbf{z}$ using the method described in the Procedure section that makes (x, u) appear to be equally bright as $(\mathbf{z}, v)$. Fig. 1 describes the process. Panel A depicts what is displayed on the monitor, where the letters indicate the stimulus intensity. Panel B depicts the stereoscope through which the respondents view the monitor. Panel C depicts what the subject sees. Because the stereoscope creates a cyclopic image, a unitary percept, these are symbolically indicated as $\mathbf{z} \oplus v$ and $x \oplus u$; the symbol $\oplus$ stands for the unknown operation that combines images in the two eyes into a single percept. The initial intensity for $\mathbf{z}$ is chosen at random in a 40 LUT interval around a best guess for its final estimate.

In terms of the stimulus display, the goal is to find the $\mathbf{z}$ that will make the percept $\mathbf{z} \oplus v$ appear equally bright as the percept $x \oplus u$. Table 3 lists the stimulus values for the two conditions under which the conjoint commutativity was evaluated.

3.3.2. *Results and discussion*

The results are summarized in Table 4, which is organized as Table 2.

Data from two respondents were excluded for the following reason. Our extensive experience of collecting matching data made it clear that their inter-session variability was atypically large; hence we have excluded them. This should not be regarded as suspicious, because large variability generally makes it harder to reject a null hypothesis, although the M–W U test is largely unaffected by uniform changes in variability, a feature that is helpful in dealing statistically with inter-session variability.

The invariance property fails for one respondent in one condition. This means that the property is supported in 14/15 tests. This

Table 4
Results of Experiment 2: Test of the conjoint commutativity for brightness. Listed for each respondent are the conditions tested, mean (M) and standard deviations (SD) of the results, number of observations (n) obtained for each condition, and finally the results of the statistical testing.

Condition	Respondent	Intensity (cd/m^2)					M–W U
		d	SD	**e**	SD	n	$p_{\mathbf{d}\sim\mathbf{e}}$
		M		M			
C_1	R10	207.0	11.0	209.0	8.8	30	.722
C_1'		178.3	9.87	191.5	9.5		.008
C_2	R22	321.4	17.8	315.0	14.1	30	.636
C_2'		275.3	18.1	261.9	19.4		.424
C_1	R79	163.1	16.4	159.0	18.5	30	.327
C_1'				Technical error: Data N/A			
C_1	R81	164.7	20.5	178.6	13.4	30	.286
C_1'		160.8	15.7	154.8	16.3		.286
C_1	R86	157.8	18.5	153.8	16.7	51	.581
C_1'		172.3	19.1	179.8	25.0		.623
C_2		270	31.5	277.4	19.2	34	.922
C_2'		268.9	33.5	257.4	9.7		.418
C_3	R90	174.1	17.8	180.4	19.9	35	.523
C_3'		180.5	18.8	183.9	20.0		.977
C_3	R91	218.7	40.37	206.2	40.4	34	.525
C_3'		184	36.1	183.0	36.3		.822

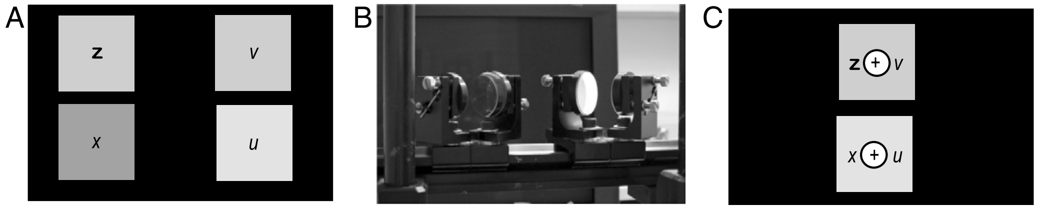

Fig. 1. Stimuli displayed on a monitor (A) viewed through a stereoscope (B) produce the subjective percept seen by the respondents (C). The *x*, *u*, and *z* values are luminance. (Figure reprinted with permission from Steingrimsson (2009)).

Table 5
The table summarizes the testing of conjoint commutativity. Listed for each domain are the number of tests, the testing result, and finally the conclusion for the property.

Domain	#Tests	#Hold	#Fail	%Hold	Hypothesis
Loudness	12	11	1	92	Supported
Brightness	15	14	1	93	Supported

we regard as reasonably strong evidence in favor of conjoint commutativity and, as a consequence, evidence favoring an additive representation in brightness.

Ours is, as far as we have determined, the first empirical evaluation of the algebraic version – or any other version – of conjoint commutativity reported in the literature. For loudness, Gigerenzer and Strube (1983) rejected a probabilistic version of the property, but a variety of differences in both methodology and pattern of results plausibly accounts for the difference in outcomes for that domain — see the discussion of Experiment 1 for details.

4. Discussion

We proved a theorem that shows that, under certain plausible structural assumptions of a binary conjoint structure, Assumptions A4 and A6 of Section 3, the conjoint commutativity property formulated by Falmagne (1976, which he called, somewhat ambiguously, the "commutative rule") can play the role of double cancellation or the Thomsen condition in arriving at the additive conjoint representation. Conjoint commutativity has certain symmetric advantages over the existing axioms when it comes to using matching procedures in evaluating additivity. This we coupled with empirical tests of the conjoint commutativity in both loudness and brightness. The results are summarized in Table 5.

Our tests of the conjoint commutativity property lead us to conclude that it has received good initial support for loudness and brightness. This conclusion is in sharp contrast to the loudness data of Gigerenzer and Strube (1983), who concluded that neither the double cancellation condition nor conjoint commutativity were sustained. As detailed in the discussion of Experiment 1, numerous differences between the two studies may plausibly account for the differences. The two most notable ones are that Gigerenzer and Strube (1983) used an estimate of the median from the first step when doing the second estimation, whereas we used individual estimates in a first estimation step as the input to the second one, a methodology that preserves the variance, avoids within-session bias, and facilitates pooling data from across several sessions; they observed a consistent bias in their results, a bias we do not observe in ours. In addition, we collected more observations for the second step than they did. Finally, there were some differences between the signals used which may have an unknown influence on the outcome. These several substantive differences seem enough to account for the different results.

In conclusion, our results are consistent with those of both Steingrimsson and Luce (2005a), who evaluated the Thomsen condition in loudness, and Steingrimsson (2009), who did the same for brightness. This separate axiomatic empirical evaluation thus strengthens the overall conclusion for an additive representation in these domains. Conjoint commutativity can readily be adapted to testing in other domains, and indeed is currently under investigation for perceived contrast. In a broader context, the results replicate and thus strengthen the additivity conclusion for these two domains, and in that sense support any model for these two domains that asserts additivity regardless of whether their "flavor" is axiomatic or based on any other approach.

Acknowledgments

This research was supported in part by National Science Foundation grant BCS-0720288 and by the Air Force Office of Scientific Research grant FA9550-08-1-0468. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or of the Air Force. We thank Dr. Bruce Berg for unfettered access to his auditory laboratory. And finally we thank the Editor and three reviewers for suggestions that led us to improve the article substantially.

Appendix. Proof of Theorem 1

Proof. It is well known that (iv) implies (i), (ii), and (iii). Theorem 1 on p. 257 of Krantz et al. (1971) shows that (i) implies (iv). The theorem on p. 490 of Holman (1971) (summarized as Theorem 2 on p. 257 of Krantz et al. (1971)) shows that (ii) implies (iv). To show that (iii) implies (iv), it is sufficient to show that (iii) implies (ii).

Suppose that A and P are sets with $a, b \in A$ and $p, q \in P$. Consider the Thomsen condition given by (3) and (4) above. In terms of m, (9), this is restated as

$$b = m_{p,q}(a) \quad \text{and} \quad c = m_{s,q}(a) \tag{A.1}$$

implies

$$c = m_{s,p}(b). \tag{A.2}$$

We wish to prove that (A.1) implies (A.2).

Assume (A.1), and let c' be the solution to

$$c' := m_{s,p}(b). \tag{A.3}$$

So we need to show that $c' = c$. From (A.3) and (A.1)

$$c' = m_{s,p}(b) = m_{s,p}[m_{p,q}(a)].$$

By conjoint commutativity (10),

$$c' = m_{p,q}[m_{s,p}(a)]. \tag{A.4}$$

Define

$$d := m_{s,p}(a) \Leftrightarrow (d, p) \sim (a, s). \tag{A.5}$$

By (3) and (A.5),

$$(d, p) \sim (c, q) \Leftrightarrow c = m_{p,q}(d)$$

but we know by (A.4) and (A.5)

$$c' = m_{p,q}(d) = c,$$

thus proving that $c' = c$, which is the Thomsen condition (ii). □

References

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

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, & E. Borg (Eds.), *Fechner Day 2003. Proceedings of the 19th annual meeting of the international society for psychophysics* (pp. 71–76). Stockholm, Sweden: International Society for Psychophysics.

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.

Gigerenzer, G., & Strube, G. (1983). Are there limits to binaural additivity of loudness? *Journal of Experimental Psychology: Human Perception and Performance*, *9*, 126–136.

Holman, E. W. (1971). A note on additive conjoint measurement. *Journal of Mathematical Psychology*, *8*, 489–494.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971–2007). *Foundations of measurement, Vol. I*. Academic Press, Reprinted 2007 by Dover Publications.

Legge, G. E., & Rubin, G. S. (1981). Binocular interactions in suprathreshold contrast perception. *Perception & Psychophysics*, *30*, 49–61.

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. (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. (1990–2007). *Foundations of measurement, Vol. III*. Academic Press, Reprinted 2007 by Dover Publications.

Luce, R. D., Steingrimsson, R., & Narens, L. (2010). Are psychophysical scales of intensities the same or different when stimuli vary on other dimensions? Theory with experiments varying loudness and pitch. *Psychological Review*, *117*, 1247–1258.

Luce, R. D., & Tukey, J. (1964). Simultaneous conjoint measurement: a new type of fundamental measurement. *Journal of Mathematical Psychology*, *1*, 1–27.

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

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. (2011). Evaluating a model of global psychophysical judgments for brightness: II. Behavioral Properties Linking Summations and Productions. *Attention, Perception & Psychophysics*, *73*, 872–885. doi:10.3758/s13414-010-0067-5.

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, Italy: International Society for Psychophysics.

Suppes, P., Krantz, D. H., Luce, R. D., & Tversky, A. (1989–2007). *Foundations of measurement, Vol. II*. Academic Press, Reprinted 2007 by Dover Publications.

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

Zimmer, K., Luce, R.D., & Ellermeier, W. (2001). Testing a new theory of psychophysical scaling: Temporal loudness integration. In G. Sommerfeld, R. Kompass and T. Lachmann (eds.), *Fechner day 2001, Proceedings of the 17th annual meeting of the international society for psychophysics*.

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