Theory And Tests Of The Conjoint Commutativity Axiom For Additive Conjoint Measurement☆

Duncan Luce, Ragnar Steingrimsson*

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

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 without exhibiting the redundancies of double cancellation. However, it too has some undesirable features when it comes to its empirical evaluation, chief among them being a certain statistical asymmetry in estimates used to evaluate it caused by two interlocked hypotheses and one conclusion. Nevertheless, thinking we had no choice, we evaluated the Thomsen condition for both loudness and brightness and in conformability with other lines of research found more support for additivity than not. However, we commented on the difficulties we had encountered in evaluating it. Thus we sought a more symmetric replacement, which we have found in the conjoint commutativity axiom proposed by Falmagne (1976, who referred to it as 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 the conjoint commutativity is equivalent to the

☆Portions of this material has also appeared in Conference Proceedings of the Meeting of the Fechner Society (Steingrimsson & Luce, 2010), in which we referred to Conjoint Commutativity as the Commutativity Rule.

*Corresponding author

Email address: ragnar@uci.edu (Ragnar Steingrimsson)

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. Current data show support for the conjoint commutativity in both domains and thus for conjoint additivity.

---

1. Background

In a variety of recent studies (Luce, 2004, 2008; Steingrimsson & Luce, 2005a; Steingrimsson, 2009) we have been confronted with deciding whether the two ears or two eyes under subjective intensity measures satisfy the axioms for additive conjoint measurement leading to an additive numerical representation. A great deal of the relevant literature was summarized in the *Foundations of Measurement* (FofM) (Vol. I Krantz, Luce, Suppes, & Tversky, 1971/2007; Vol. II Suppes, Krantz, Luce, & Tversky, 1989/2007; Vol. III Luce, Krantz, Suppes, & Tversky, 1990/2007). The gist of that literature is that in the presence of various axioms (summarized in Def. 8, p. 76–77, FofM I) we have additivity under either of two properties. Suppose that A and P are sets and there is an ordering $\succsim$ over $A \times P$. Suppose that for $a, b, c \in A$ and $p, q, s \in P$ we have the axiom of *double cancellation*, i.e.,

$$(a,p) \succsim (b,q) \;\text{ and }\; (c,q) \succsim (a,s) \tag{1}$$

$$\Rightarrow (c,p) \succsim (b,s), \tag{2}$$

which is key to the additive conjoint representation. Its special case

$$(a,p) \sim (b,q) \;\text{ and }\; (c,q) \sim (a,s) \tag{3}$$

$$\Rightarrow (c,p) \sim (b,s), \tag{4}$$

is called the *Thomsen condition* which in the same context also suffices for the additive representation.

Gigerenzer & Strube (1983) pointed out very clearly 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) that makes for inference problems.

The article is structured as follows:

- An overview of current tests of conjoint additivity and an expanded discussion of the problems with testing double cancellation and the Thomsen condition.
- Presentation of the conjoint commutativity and a demonstration of how it offers a possible solution to those problems.
- An empirical evaluation of the conjoint commutativity for both loudness and brightness.

2. Current tests of conjoint additivity

Additivity over sense organs has been studied in a variety of ways; here we focus on the axiomatic evaluation. As summarized in Def. 8 of FofM I (p. 76–77) the test of additivity involves the evaluation of either double cancellation or the Thomsen condition. Double cancellation was explored for loudness by Levelt, Riemersma, & Bunt (1972), Falmagne (1976), Falmagne, Iverson, & Marcovici (1979), Gigerenzer & Strube (1983), and Schneider (1988); for perceived contrast by Legge & Rubin (1981); and for cross-modal additivity of loudness and brightness of double cancellation by Ward (1990). Of these studies, two rejected double cancellation (Falmagne, 1976; Gigerenzer & Strube, 1983).

Giegerenzer & Strube (1983) made clear why one should not attempt to test double cancellation because of the following type redundancy: Suppose the signals are a, b, c, p, q, s for which double cancellation holds

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

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

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

For example, they reported that "...no less than 82.64% of Levelt et al.'s data were defined a priori."

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

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

Because these are indifferences they have to be constructed empirically. In the experiments that we 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 fashion until a match is archived. 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 ways. One is to start with a, b, p, s and estimate first $\mathbf{q}$ 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 straight forward testing method, the estimate $\mathbf{c}$ rests upon the estimate of $\mathbf{q}$, and so it has two sources of variability/bias whereas that of $\mathbf{c}'$ has only one source of variability/bias, where a bias, e.g., may be a time-order error (arising from sequential presentation of stimulus pairs).

The empirical record for ests of both double cancellation and the Thomsen condition are mixed and attributable, respectively, to redundancy and statistical asymmetry.

Another possible source of difficulty relates to the respondents who produce these estimates. 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 location change of a tone, whereas in brightness it is has no direct subjective correlate. The fact is that Steingrimsson and Luce (2005a) and Steingrimsson (2009) found, for both loudness and 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 in which the adjusted stimulus was changes equally in both sensory organs.

3. Equivalence of Conjoint Commutativity to the Thomsen Condition

In an article on random conjoint measurement, Falmagne (1976) introduced the following property that he called the *commutativity rule* and we make more specific by speaking of *conjoint commutativity.* With the function[1] $m_{p,q}$ defined by:

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

then 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 the conjoint commutativity to random conjoint measurement, this property applies equally well to the algebraic case. As far as we have been able to ascertain, the literature has overlooked the

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

fact that conjoint commutativity can equally well play the role of the Thomsen condition and with some advantages.

Consider the part of the Definition 7 of a binary conjoint structure on p. 256 of Krantz et al. (1971) given by the following assumptions[2]:

A1 Weak ordering.

A2 Monotonicity (= independence).

A4 Unrestricted solvability.

A5 Archimedeaness.

(6) Each component is essential.

Assumptions A1 and A2 are easily seen to be necessary properties that ordering must satisfy if there is an additive representation and assumptions A4-A6 are three structural assumptions.

Theorem 1. *Under assumptions A1, A2, A4, A5, and A6, then the following four properties that can play the role of A3 are equivalent:*

(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 the 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 that both sides of the hypotheses have

[2]The misssing Assumption A3 is what we are studying.

double estimates. It does not, however, address the potential issue that may arise in empirically producing a match by adjusting intensity in one sensory organ while keeping the intensity constant in the other organ.

Conjoint commutativity, (10), is evaluated in two domains, loudness and brightness, presented as Experiment 1 and 2, respectively.

3.1. General Method

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

3.1.1. Respondents

A total of 9 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 accordance with the "Ethical Principles of Psychologists and Code of Conduct" (American Psychological Association, 2002). Consent forms and procedures were approved by UC Irvine's Institutional Review Board.

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, therefor for a left ear/eye with a threshold of x_τ and a right ear/eye one of u_τ, the effective stimulus (x, u) consists of $x = x' - x_\tau$ 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 normal hearing/vision, the error in reporting intensities (x', u'), using dB or cd/m^2 is negligible.

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

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 $L_{\text{side}} = R_{\text{side}}$. The Social Sciences, in contrast to other sciences, tend to focus on statistical inferences based on rejecting null hypotheses. In physics our type of testing is quite familiar and tends to take the form of articulating a criteria and a level of accuracy with which a null hypothesis — an invariance of the theory — is said to be supported by the data. Here a similar a similar approach is pursued through the formulation of a criterion consisting of multiple interlocked components which has to hold for accepting the data as supportive of the hypothesis. Furthermore, 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).

The consequent three component analysis is the criterion that has resulted (see Steingrimsson, 2009, for extensive details).

1. Because we have no a priori model of the distribution of the data, we use a non-parametric statistical test, namely, the Mann-Whitney U test at the 0.05 level. This is the choice of numerous other papers (e.g., Falmagne, 1976; Gigerenzer & Strube, 1983; Ellermeier & Faulhammer, 2000; Zimmer, Luce, & Ellermeier, 2001; Ellermeier, Narens, & Dielmann, 2003; Zimmer, 2005; Steingrimsson & Luce 2005a, 2005b, 2006, 2007; Steingrimsson, 2009, 2010)
2. Test of statistical power: Using our estimation method, our sample is known to converge on the median of the distribution (Theorem 2 of Falmagne, 1976), and according to Pratt (1964) this means ensuring that the sample is adequate to accept or to reject the null. We use a Monte Carlo simulation (details in Steingrimsson & Luce, 2005a; Steingrimsson, 2009) which is suggested an appropriate method for this purpose (Mumby, 2002). Briefly, the simulation involves pooling the L_{side} and the R_{side} data, take two random samples from the pool and repeat the test many times and observe the percentage satisfying the M-WU condition. Using samples of

size $1,000$ we also report whether or not the results deviates from the test statistic.

3. Effect size evaluation: There is no clear way to evaluate the effect size when using a non-parametric statistic. Instead we adapt the somewhat novel approach of Steingrimsson (2009) of using Weber's fraction as the rough evaluation of the effect size. We reject the null when the means differ by more than Weber's fraction. Steingrimsson (2009) finds that a value of between .05 and .08 is appropriate in the current testing situation. We use the lower of the two.

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

Because the estimation steps are made in discrete steps and the estimates appear reasonably symmetric, the mean is known to be the best estimate for the median. In loudness, we report and collect data using dB's, whereas in brightness stimuli and data are in LUT (the monitor's video card **L**ook**U**p **T**able of integer values 0-255) values, but reported in cd/m^2. This involves a power transform wherefore we report the transformed mean LUT values as well as the normalized standard deviations (maintain relative magnitude vis-a-vis the mean in cd/m^2—we thank Dr. J. Yellott for this suggestion).

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 the respondents control and x, u, and v are constants provided 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 $\mathbf{z}$ up and down in intensity as often as desired until he or she is satisfied with the match. Concretely, respondents could choose any of four changes in intensity described as extra-small, small, medium, large. in the loudness case, they corresponded to .5, 1, 2, or 4 dB whereas the brightness case these corresponded to luminance steps of 1, 2, 4, or 8 LUT values. Following an adjustment the stimuli were re-presented with the requested adjustment included. The respondent repeated this process until satisfied with the match, which was indicated by another key press, after which the next matching task commenced.

The following four matches were needed to evaluate the conjoint commutativity

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

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 $\mathbf{b}$ 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 $\mathbf{c}$ is obtained in step 3; the respondent produces $\mathbf{e} = m_{r,s}(\mathbf{c})$.

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

In steps 1-4, all adjustments are made in the left sensory organ. The property can equally well be evaluated by its mirror image in which $b = m'_{p,q}(a)$ iff $(p, a) \sim (q, b)$, 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 labeled steps 5-8 and are analogous to steps 1-4.

If sensation evoked from physically identical inputs to the two ears/eyes were identical, they would be behaviorally interchangeable. However, such symmetry has unequivocally been rejected in both loudness (Steingrimsson & Luce, 2005a) and brightness (Steingrimsson, 2009). This means that $m'_{p,q}(a) \neq m_{p,q}(a)$ and thus 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 effected by, e.g., sustained matching in one ear, e.g., within a session, a result Steingrimsson and Luce (2006) described using a form of a filtering model.

A feature of the experimental steps 1-4 and 5-8, respectively, is that while the matching stimulus is presented to the two sensory organs, the variable stimulus is in either the left or the right ear only. One plausible consequence of

Steingrimsson and Luce's (2006) filtering model is that it may be desirable to mix the sensory organ receiving the variable stimulus within a block of trials. Therefore, by mixing the steps 1-4 and 5-8 within a block of trials, two different conditions are evaluated while maintaining a plausibly important balance in the testing situation.

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 eight matches (1-8) were run in a block of trials generating two tests of the conjoint commutativity. Respondents typically completed 10 blocks per 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 stimulus (x, u) means a joint presentation of a tone with intensity x in the left ear and a tone with intensity u in the right ear. These tones were sinusoids of 100 ms duration that included a 10 ms on and 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). Intensity and frequency was controlled through 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 lab-room. A safety ceiling of 90 dB was imposed in all experiments.

Loudness matching and the stimulus conditions: In addition to the general description of matching provided in the Procedure section of the General

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

Table 1: In the table are listed the two stimulus instantiations under which the the conjoint commutativity was tested, which, when applied to both m and m', makes for a total of four testing conditions.

Methods, specific to achieving the loudness match $(x, u) \sim (\mathbf{z}, v)$, the experimenter chooses x, u, and and v, and the respondent produces, the $\mathbf{z}$ that makes (x, u) sound equally loud as $(\mathbf{z}, v)$. 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 t and t', the number of observations, n, for each sample. The statistics columns have three components. The first is the the Mann-Whitney, where we report the result of the null hypothesis test $t \sim t'$, given as $p_{t \sim t'}$. This is followed by the second component, the results of the simulation for evaluating the adequacy of the samples to detect the true failure of the hypothesis. The third component is the evaluation of the effect size columns which asserts that the samples may not differ by more than .05 ($\sim$Weber's fraction). For the data to be said to support the hypothesis, all three components must support it. That conclusion is reported in the last column of the table.

For R81, the C_1' condition fails on 2/3 components of the statistical criteria. Otherwise, Thus the property is found to be supported in 13/14 tests. This we regard as a reasonably strong evidence in favor of the property and as a consequence, evidence favoring an additive conjoint representation in loudness.

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

Table 2: Results of Experiment 1: Test of the conjoint commutativity for loudness. Listed for each respondent are the conditions tested, mean and standard deviations of the results, number of observation obtained for each conditions, and finally the results of the statistical testing.

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 minutes of dark adaptation. The stimuli were generated by a personal computer and displayed on an monitor (Eizo RadiForce RX320) with automatic luminance uniformity equalizer and backlight sensor to compensation for luminance fluctuation caused by ambient temperature and passage of time as well as build in gamma correction. The diagonal size is 54 cm, maximum resolution is 1536 x 2048, and 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 to determined luminance of 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 degrees 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 and v, and the respondent produces the $\mathbf{z}$, using the method described in the Procedure section, that makes (x, u) appear equally bright as $(\mathbf{z}, v)$. Figure 1 describes the process: Panel A depicts what is displayed on the monitor, where the letters indicate stimulus intensity. Panel B depicts the stereoscope through which the respondents view the monitor. Panel C depicts what the subject sees. Since the stereoscope creates a cyclopic image, a unitary percepts, these are symbolically indicated as $\mathbf{z} \oplus v$ and $x \oplus u$, the 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 30 cd/m^2 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 $x\oplus u$. Listed in Table 3 are the

	Stimuli in cd/m^2				
Condition	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

Table 3: In the table are listed the three stimulus instantiations under which the the conjoint commutativity was tested, which, when applied to both m and m', makes for a total of six testing conditions.

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 on account of very unusually large inter-session variability. We have conducted such matching data in a variety of experiments and can say such inter-session variability is quite uncharacteristic of the task, hence we have excluded them. Of course, such high inter-session variability leads to higher variances in data than usual, which makes it harder to reject the propriety under investigation. So these excluded data tended to favor accepting the null hypothesis the property than not.

The property does not fail out-right for any of the respondent, but it fails in 1 of 2 tests for 3 respondents. That is, the property is supported in 12/15 tests. This we regard as a reasonably strong evidence in favor of the property and as

Figure 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).

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

Table 4: Results of Experiment 2: Test of the conjoint commutativity for brightness. Listed for each respondent are the conditions tested, mean and standard deviations of the results, number of observation obtained for each conditions, and finally the results of the statistical testing.

Domain	#Tests	#Hold	#Fail	%Hold	Hypothesis
Loudness	12	11	1	92	Supported
Brightness	15	12	3	80	Supported

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

a consequence, evidence favoring an additive representation in brightness.

4. Discussion

We have proved a theorem that shows that under certain plausible structural assumptions of a binary conjoint structure, Assumptions A4, A5, and A6 of Section 3, that a certain conjoint commutativity formulated by Falmagne (1976, which he referred to simply as the commutative rule) can play the role of double cancellation or the Thomsen condition in arriving at the additive conjoint representation. This 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.

The conclusion of the testing is that the conjoint commutativity has received a good initial support in these domains. These results are consistent with both Steingrimsson and Luce (2005a) who evaluated the Thomsen condition in loudness as well as 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. The conjoint commutativity can readily be adapted to testing in other domains and indeed is currently under investigation for perceived contrast.

Acknowledgments This research was supported in part by National Science Foundation grant BCS-0720288 and from the Air Force Office of Research grant FA9550-08-1-0468—any opinion, finding, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or of the

Air Force. We thank Dr. Bruce Berg for unfettered access to his auditory laboratory.

References

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

Augustin, T., & Maier, K. (2008). Empirical evaluation of the axioms of multiplicativity, commutativity, and monotonicity in ratio production of area. *Acta Psychologica, 129*, 208–216.

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.

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. (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.

Mumby, P. J. (2002). Statistical power of non-parametric tests: A quick guide for designing sampling strategies. *Marine Pollution Bulletin, 44*, 85–87.

Pratt, J. W. (1964). Robustness of some procedures for the two-sample location problem. *Journal of the American Statistical Association, 59*, 665–680.

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. (2010). Evaluating a model of global psychophysical judgments for brightness: II. Behavioral Properties Linking Summations and Productions. *Attention, Perception, & Psychophysics.* Online First: DOI 10.3758/s13414-010-0067-5.

Steingrimsson, R., & 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 Bastianelli, A.. & Vidotto, G. (Eds.) *Fechner Day 2010. Proceedings of the 19th annual meeting of the International Society for Psychophysics.* Padua, Italy: International Society for Psychophysics. Pp. 21–26.

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, & T. Lachmann (Eds.), *Fechner Day 2001: Proceedings of the 17th Annual Meeting of the International Society for Psychophysics.* Lengerirch, Germany: Pabst.

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

Proof of the Theorem

It is well known that (iv) implies (i), (ii) and (iii). Theorem 1 on p. 257 of Krantz et al. (1971) shows (i) implies (iii). 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 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 restates as

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

implies

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

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

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

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

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

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

By conjoint commutativity (10)

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

Define

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

By (3) and (.5)

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

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

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

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