THE COMMUTATIVE RULE AS NEW TEST FOR ADDITIVE CONJOINT MEASURMENT: THEORY AND DATA

Ragnar Steingrimsson and R. Duncan Luce
Institute of Mathematical Behavioral Sciences, University of California, Irvine, Irvine, CA 92697
ragnar@uci.edu and rdluce@uci.edu

Abstract

Empirical evaluation of the key axiom underlying additive conjoint representation first studied the double cancellation axiom. That was shown to have considerable redundancy that made the statistical problems formidable. The special case called the Thomsen condition was shown to suffice and not to be redundant. However, it has the undesirable feature, for empirical purposes, of a statistical asymmetry in estimation. This led us to seek a symmetric replacement, which we have found in the commutative rule proposed by Falmagne (1976). In the presence of the usual assumptions, we show that the commutative rule is equivalent to the Thomsen condition, a result that appears to have been overlooked in the literature. We subject this property to empirical evaluation in both loudness and brightness. Current data show support for the commutative rule in both domains and thus for additivity.

Additivity over sensory organs, e.g., loudness over the two ears or brightness over the two eyes has been studied in a variety of ways. In the context of axiomatic psychophysics, this involves evaluating whether subjective intensity measures satisfy the axioms of additive conjoint measurement, in which case an additive numerical representation may be concluded. As summarized by Krantz, Luce, Suppes, & Tversky, 1971/2007 (Def. 8, p. 76-77), this effectively involves evaluating either double cancellation or the special case of it, the Thomsen condition.

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 $p,q,r \in P$ then the axiom of double cancelation can be expressed as, if $(a,p) \succsim (b,q)$ and $(b,r) \succsim (c,p)$, then $(a,r) \succsim (c,q)$. The Thomsen condition is the special case where *)* $\succsim$ is replaced by ~. Both of these have been evaluated in a variety of settings and domains, including loudness, brightness, their cross-modal case, and perceived contrast (Levelt, Riemersma, & Bunt,1972; Falmagne, 1976; Falmagne, Iverson, & Marcovici, 1979; Gigerenzer & Strube, 1983; Schneider, 1988; Legge & Rubin, 1981; Ward, 1990; Steingrimsson & Luce, 2005; Steingrimsson, 2009).

Although the results have been somewhat mixed, more studies have favored additivity than have not. Some of the discussions in the literature have focused on the actual difficulty involved in the empirical testing itself. For instance, Giegerenzer & Strube (1983) uncovered a great deal of redundancy in the tests of double cancelation, e.g., that no less than ~83% of Levelt et al.'s data were a priori defined. Steingrimsson and Luce (2005) and Steingrimsson (2009) therefore focused on the testing of the Thomsen condition, which does not include such redundancy. We may formulate this in terms of Falmagne's (1976) defined the function m:

$$b = m_{p,q}(a) \text{ iff } (a,p) \sim (b,q). \quad (1)$$

In this notation, Thomsen condition may be expressed as:

$$m_{r,p}(m_{p,q}(a)) \Rightarrow m_{r,q}(a). \quad (2)$$

In this form, it becomes immediately obvious that the left side of (2) consists of two steps, whereas the right side consists of only one step. This feature introduces a statistically undesirable consequence, which is possibly the source of some or all of the difficulty in the empirical evaluation of the property.

Falmagne (1976) introduced the property he called the *commutative rule*, which asserts that

$$m_{p,q}[m_{r,s}(a)] \Leftrightarrow m_{r,s}[m_{p,q}(a)]. \quad (3)$$

As far as we have been able to ascertain, the fact that this commutative rule can equally well play the role of the Thomsen condition has been overlooked until Luce and Steingrimsson (in preparation) proved that result. It is immediately obvious that the commutative rule is balanced in terms of steps on the left and the right side of the equivalence, which addresses at least one potential problem in the empirical evaluation of additivity. In the following, we report tests of the commutative rule for loudness and brightness.

Method

Respondents

A total of seven students at the University of California, Irvine, and one coauthor participated in the two experiments. All respondents who provided loudness data reported normal hearing and those providing brightness data reported corrected-to-normal vision. UC Irvine's Institutional Review Board approved consent forms and procedures.

Statistical methods

All analysis was done on individual data. Lack of parametric information led us to use the non-parametric Mann-Whitney test at the customary significance level of .05. Equality of medians (a test holding) required three criteria to be met—if one or more failed, the equality was rejected (see Luce, Steingrimsson, & Narens, in press, for details).

1. A *p*-value is generated by the Mann-Whitney test.
2. A Monte-Carlo simulation was used to evaluate the adequacy (power) of the samples to reject the null hypothesis.
3. The two medians were required to be within Weber's fraction of each other, a form of evaluation of effect size.

Equipment, Stimuli, and Intensity reporting

In studies of loudness, it has long been customary to measure, collect, analyze, and report stimuli intensities in dB. This we do as well using the SPL reference. In studies of brightness, the tradition is less clear. In early studies by. e.g., S. S. Stevens, intensities were reported in dB, whereas in recent years, cd/m^2 have become a more dominant measure. However, studies using computer monitors by their technical design use a variable taking the integer values of 0-255 called LUTs that varies as a power function with cd/m^2. This power relation requires some care when reporting stimuli, results, and when analyzing data. For instance, if standard deviations were calculated by first converting LTUs to cd/m^2 then these would vary exponentially with the mean LUT values, rather proportional and thus appear far larger than they were in the actual experimental situation. Conversely, if dB's were used, and standard deviations of these were calculated in the same manner, these would appear smaller. For monitors with small effective range (up to say 100 cd/m^2) these errors are minor. Clearly, this issue needs some attention in future publications; in the meantime we report brightness intensity in a fashion analogous to the loudness ones, i.e., using LUTs. The calibration of our equipment provides the following conversion formula:

$$I = 742\left(\frac{LUT}{255}\right)^{2.2101} \quad (4)$$

where I is intensity in cd/m^2.

Loudness: 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. These 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.

Brightness: 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. The experiment was conducted in a dark room and each respondent received a minimum of 10 minutes of dark adaptation. Information about the current block and trial number were displayed in small letters in the upper left corner of the screen. The stimuli were generated by a personal computer and displayed on a 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.

Procedure

Empirical evaluation of the commutative rule, (3), involves obtaining several matches of the generic form $(x,u)\sim(\mathbf{z},v)$ where the 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 z up and down in intensity as often as desired until s/he is satisfied with the match.

The following four matches were needed to evaluate the commutative rule:

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 $\mathbf{d} = m_{p,q}$(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 $\mathbf{e} = m_{r,s}$(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 $b = m'_{r,s}(a)$ iff $(p,a)\sim(q,b)$, and the commutative rule is then given by $m'_{p,q}[(m'_{r,s}(a)] = m'_{r,s}[(m'_{p,q}(a)]$. 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 (Steingrimssn & Luce, 2005) and brightness (Steingrimsson, 2009). This means that $m_{p,q}$(a) $\neq m'_{p,q}$(a) and thus constitute different experimental conditions.

Obtaining the matches for m' requires 4 steps, all of which are analogous to the steps 1-4 (effectively, the order of stimuli is switched). Together steps 1-4 along with those for obtaining t m' require eight matches in all. These eight matches were run in a block of trials generating two tests of the commutative rule. An additional benefit is that in these blocks of trials which ear/ear in which the light/tone appears, is randomized.

Loudness: The task of obtain the ***b*** such that *(a,p)* is experienced equally in loud to *(**b**,q)* involved the first playing signal *(a,p)*, experienced as a single tone, followed after 450 ms by *(**b**,q)*. The respondent then indicated either a desire to increase or decrease the loudness of the second pair, i.e., ***b***, after which the tone sequenced was repeated with the with the indicated change. The respondent could also repeat the sequence unchanged. The respondent continued until s/he was satisfied with the loudness match between the first and second tone. The initial intensity for ***b*** is randomly chosen.

Brightness: The task is to obtain the ***b*** so that *(a,p)* is experienced to be equally in brightness to *(**b**,q)*. The realization of this method is depicted in Figure 1. 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 $a \oplus p$ and $\boldsymbol{b} \oplus q$, where the symbol ⊕ stands for the unknown operation that combines images in the two eyes into a single percept. The initial intensity for ***b*** is randomly chosen.

Fig. 1: Stimuli displayed on a monitor (A) viewed through a stereoscope (B), produce the subjective percept seen by the respondents (C). The a, b, p, and q values are luminance.

Respondents typically completed 10 blocks per 1-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. Listed in Table 1 are the 2 stimulus sets under which the commutativity rule was tested.

Table 1: Stimuli and conditions used for testing the commutativity rule.

Condition			Stimuli dB		
Loudness	a	r	S	p	q
C1	64	70	67	66	58
C2	60	66	67	68	62
Brightness			Stimuli in LUT		
C1	100	125	110	90	80
C2	120	160	140	110	110

Results and Discussion

Listed in Tables 2 and 3 are the results of the test of the commutativity rule in loudness and brightness, respectively. Listed by respondent are the conditions tested, the mean and the standard deviations of final estimates of ***d*** and ***e***, the number of observations collected for each and the *p* resulting from the test of the equality of medians. Each result is also evaluated by the two additional criteria listed under *Statistics Methods.* Conversion from LUT values to cm/m^2 is provided by Equation (4). Table 4 summarizes these results.

Table 2: Results of testing the commutative rule in Loudness

		Intensity Level in dB					
		d		***e***			
Condition	Respondent	M	*SD*	*M*	*SD*	*n*	$p_{d\sim e}$
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: Results of testing the commutative rule in Brightness

		Intensity Level in LUT					
		d		***e***			
Condition	Respondent	M	*SD*	*M*	*SD*	*n*	$p_{d\sim e}$
C_1	R10	143.1	6.0	143.7	4.2	30	.722
C_1'		133.8	7.3	138.2	6.9		.008
C_2	R22	174.6	9.7	173.1	7.7	30	.636
C_2'		162.4	10.7	158.7	11.8		.424
C_1	R80	153.8	35.4	156.5	26.7	29	.529
C_1'		138.6	20.0	144.8	13.5		.301
C_1	R81	129.1	16.1	133.9	10.0	30	.286
C_1'		127.7	12.5	125.5	13.2		.286
C_1	R86	126.6	14.8	125.1	13.6	51	.581
C_1'		131.7	14.6	134.3	18.7		.623
C_2		161.5	18.8	163.4	18.5	34	.922
C_2'		161.1	20.1	157.9	6.0		.418

Table 4: Summary of results of testing the commutative rule.

Domain	#Tests	#Hold	#Fail	%Hold	Hypothesis
Loudness	14	13	1	93	Supported
Brightness	12	11	1	92	Supported

The overall conclusion is that the commutativity rule is well sustained in both loudness and brightness. This is well in line with previous data confirming additivity in these two domains and as such is not a great surprise (see Steingrimsson & Luce, 2005; Steingrimsson, 2009 for details). However, the results and the testing is considerably easier than previous axiomatic efforts that relied on double cancelation or the Thomsen condition.

Acknowledgements

This research was supported in part by National Science Foundation grant BCS-0720288 to the University of California, Irvine---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.

References

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.

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., Steingrimsson, R., & Narens, L. (in press). Are Psychophysical Scales of Intensities the Same or Different When Stimuli Vary on Other Dimensions? Theory with Experiments Varying Loudness and Pitch. *Psychological Review*.

Luce, R. D., R. Steingrimsson (in preparation). Testing for Additive Conjoint Measurement.

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., & Luce, R. D. (2005). Evaluating a model of global psychophysical judgments: II. Behavioral properties linking summations and productions. *Journal of Mathematical Psychology*, 49, 308-319.

Ward, L. M. (1990). Cross-Modal Additive Conjoint Structures and Psychophysical Scale Convergence. *Journal of Experimental Psychology: General*. 119, 161-175.