AFRL-OSR-VA-TR-2013-0122

Empirical and Theoretical Studies of Psychophysical Phenomena

R. Duncan Luce, Ragnar Steingrimsson

The Regents of the University of California

March 2013
Final Report

DISTRIBUTION A: Approved for public release.

AIR FORCE RESEARCH LABORATORY
AF OFFICE OF SCIENTIFIC RESEARCH (AFOSR)
ARLINGTON, VIRGINIA 22203
AIR FORCE MATERIEL COMMAND

REPORT DOCUMENTATION PAGE		*Form Approved* *OMB No. 0704-0188*

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to the Department of Defense, Executive Services and Communications Directorate (0704-0188). Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
02-14-2013	Final Performance Report	09/29/2010-12/29/2012

4. TITLE AND SUBTITLE Empirical and Theoretical Studies of Psychophysical Phenomena	5a. CONTRACT NUMBER
	5b. GRANT NUMBER FA9550-10-1-0468
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) R. Duncan Luce; Ragnar Steingrimsson	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) The Regents of the University of California 5171 California Avenue, Suite 150 Irvine, CA 92697-7600	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) AF Office of Scientific Research 875 N. Randolph Street Room 3112 Arlington, VA 22203	10. SPONSOR/MONITOR'S ACRONYM(S) AFOSR
	11. SPONSOR/MONITOR'S REPORT NUMBER(S) AFRL-OSR-VA-TR-2013-0122

12. DISTRIBUTION/AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A:Approved for public release

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This project is an extensive exploration of ratio scales of subjective measures on the basis of an empirically tested theory of fundamental psychophysical measurement, mimicking classical, static physical approach to measurement. The theory has been favorably evaluated in loudness, brightness, and perceived contrast The theory, formulated for binary senses, is extended to unary senses, giving a complete theory for intensive dimensions. Cross-modal and dimensional predictions have been developed to answer questions such as: to what degree are all subjective intensities a special cases of a single ratio scale of subjective intensity? How do binary and unary dimension relate? A commutativity property extends to asking whether a single ratio scale obtains when a stimulus varies on more than one dimension. Data support a single ratio scale of loudness/brightness when pitch/hue is varied. Various implications of the behavioral theory have been have been derived. Among the areas studied are the time-order error, regression effect, Torgerson's conjecture. The theory, based on a theory of magnitude productions, is extended to ratings.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					Jay Myung, Program Manager
U	U	U	UU	15	19b. TELEPHONE NUMBER *(Include area code)* 703-696-8478

Reset

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Empirical and Theoretical Studies of Psychophysical Phenomena
Grant Award No.FA9550-10-1-0468
Program Manager: Dr. Jay Myung
PI: R. Duncan Luce
University of California, Irvine

1 Preamble & Introduction

Co-PI's note: The PI, R. Duncan Luce, passed away in early August of this year (2012). Dr. Luce's long and distinguished scientific career remained active until close to his death. The volumes of scientific papers, books, and other published materials that he leaves behind speak of unusually broad, productive, and important scientific accomplishments. The many honors he received in his life-time, including the National Medal of Science, show the esteem in which his colleagues held him, and the many memorial events being held to honor him and his work being held at the meetings of several scientific societies demonstrate the indelible a mark Dr. Luce's work has made on the scientific areas to which he contributed. To say that Dr. Luce will be missed as scientist and as a person is an understatement.

It has befallen me, his last Co-PI, to compile this report. Despite our 15 year long collaboration, it feels unsettling to put into words the last scientific contributions Dr. Luce made, but it also speaks of how strong his intellect remained to the end that there is more to report than can easily be fit into this report. Some of the threads of the research have, as a consequence of his passing, loose ends. A no-cost extension of the current grant was sought to, in part, tie some of these ends, but due to changes in operating procedures at the AFOSR, such extension could not be obtained. It is nevertheless the case that all stated aims of the grant were achieved. Additionally, in an on-going collaboration colleagues important research paths on which Dr. Luce was working will continued to be pursued, which include several extending from the work reported here..

This final report of the Grant describes what has been predicted and what has been accomplished empirically during the two years of its duration. Some of the articles resulting from this grant were begun during a no-cost extension of NSF grant BCS-0720288. The relevant articles are denote with an * in the in the list of article resulting from this grant.

2 Scientific Objectives of Research

An abiding problem of scientific psychological is how to measure the not directly observable sensation that intervenes between physical stimuli and behavioral responses. The trust of past efforts has involved collecting numerical responses from respondents and fitting these data to certain mathematical functions—e.g., logarithm or power. This approach to measurement, while of much value, has some well-documented limits of generalization within and across domains and for being lacking predictive power beyond specialized situations within a domain. The PI, independently and in collaborations, focussed substantial part of his career on theoretical foundations for measurement. This foundational work progressed considerably in the latter half of the last century, but remained unmatched by empirical applications. Such empirical applications became a focus of the PI's research in the last decade, where he applied measurement theoretical results to utility theory on based on that work, to psychophysics. The work has progressed to develop behavioral equivalents to certain mathematical forms of the psychophysical function and of how respondents behave in tasks aimed at measuring the sensations of interest. The overacting objective of this grant was to continue this effort. The grant proposal listed eight main objectives, all of which have been achieved. Conceptually, the plan that has been executed had three main thrusts: expanding the existing theory, derive from the theory accounts for several outstanding psychophysical problems, and to carry out empirical evaluations. More specifically:

Theoretical development and expansion: Extend the theory from binary (e.g., two ears) senses to all intensive dimensions (e.g., unary taste). Incorporate two or more dimensions of a perceptual domain in a single account (e.g., brightness when hue or saturation vary). Establish a cross modal description that can a) establish the degree to which all subjective intensities are special cases of a single ratio scale of subjective intensity; b) and, to the degree to which that is not the case, to establish a single unifying framework for the interrelation of all intensive perceptions. Individual perceptions differ, wherefore the theory establishes on the level of modeling precisely how they differ and relate.

Derived predictions from theory: Studied are: The distorting perceptual effects of temporal and/or spatial order of stimuli presentation; the so-called regression effect; the Torgerson's conjecture that people do not distinguish subjective ratios and differences; scale properties obtained by the ratings paradigm.

Empirical evaluations: The extensive (favorable) evaluation of the theory for loudness and pitch has been extended to brightness. Further empirical evaluation will complete the work for brightness and hue as well as include perceived contrast and thus establish the theory as a unifying description for all three domains. Empirical predictions associated with the Torgerson's conjecture and measurement scales of ratings are carried out. The theoretical work extends over the duration of the grant, wherefore not all the associated empirical work could logistically be carried out. However, in all instances, the empirical program is developed in detail.

3 Technical Approach

Our approach has attempted to overcome the much criticized lack of testable, foundational assumptions in psychology, by utilizing the theoretical framework of axiomatic algebra. The method involves the formulation of necessary and sufficient mathematical invariances (axioms) that together are sufficient to conclude a particular representation for the operations described by the axioms. If the axioms are shown to hold, then the representations are concluded on the basis of mathematical logic without any requirement for direct evaluation of the representations themselves.

Formulated are necessary and sufficient behavioral and testable condition from which the detailed representations, measurement scales, are derived. The project began as a detailed axiomatic theory of psychophysics based on the operations of matching and magnitude production. The empirical aspect of the approach requires the primitives of the theory be given precise interpretation in the given psychophysical domain and then experimentally evaluate how well the behavioral properties are, on the basis of data, supported. If satisfactory support is found, the representations are concluded as descriptive theory for the domain. Among the strengths of this approach is that it allows novel predictions to be deduced from the representations in a way analogous to a good deal of classical static physics. An element of the work is thus to examine the theory to derive from it solutions to outstanding questions as well as novel predictions.

Formally, let $x, u \in X$, be signals from a set of physical intensities with 0 being the threshold level. If x is a pure tone presented to the left ear, then let u denote another such tone presented to the right ear and denote their joint presentation as $(x, u) \in X \times X$, where the subjective intensity ordering $\succsim$ is assumed to be a weak order. The stimulus could equally well be lights presented to the two eyes, weight in the two arms, or any other binary presentation.

A signal $\mathbf{z}$ is a matching signal if (z, z) is judged equally intense as (x, u). Matching has the operation notation $x \oplus u$ defined as $x \oplus y := z$.

In ratio magnitude production the experimenter presents a signal x and a number $p > 0$ and asks the respondent to report the signal $\mathbf{x}_p$ such that the "interval from the reference signal ρ to $\mathbf{x}_p$" is perceived as p "times" as intense as the "interval from ρ to x". It is also convenient to think of $\mathbf{x}_p$ as an operator: $\mathbf{x}_p = x \circ_p \rho$. The $\mathbf{z}$ and $\mathbf{x}_p$ are bold faced to emphasize that they are random variable. When $\rho = 0$, this is simply S. S. Stevens' (1975) method of magnitude production.

Luce (2002, 2004, 2008) presented an axiomatization describing the two operation, $\oplus$ and $\circ_p$, separately, as well as properties linking the two operations, which ensured that the same psychophysical function was involved in both operations, establishing a mapping between the structures $\langle X, \succsim, \oplus, \circ_p \rangle$ and $\langle \mathbb{R}^+, \geq, +, \times \rangle$. These properties imply the following numerical representation: A *p–additive* order preserving psychophysical function ψ:

$$\psi(x \oplus y) = \psi(x \oplus 0) + \psi(0 \oplus y) + \delta\psi(x \oplus 0)\psi(0 \oplus y), \quad \delta = -1, 0, 1, \tag{1}$$

And a weighting function W over positive numbers such that

$$W(p) = \frac{\psi(x_p) - \psi(\rho_i)}{\psi(x) - \psi(\rho_i)}, \quad i = \begin{cases} + & \text{if } p \geq 1 \\ - & \text{if } p < 1 \end{cases}, \tag{2}$$

Strong support, some obtained under this grant, for the behavioral axioms has been established for loudness, brightness, and perceived contrast, focusing on the data from individual respondents, (Steingrimsson & Luce,2005a, 2005b, 2006, 2007; Steingrimsson, 2009, 2011,2012a,b,c). These experiments demonstrate that we can successfully apply this theoretical approach to give a formal and unified description of the three domains of loudness, (achromatic) brightness, and perceived contrast. Furthermore, this description is unique in the sense that the scales involved are either ratio or absolute ones.

In the representations (1, 2), the two functions ψ and W figure as unknown functions. In the axiomatic approach, invariance properties are formulated that if hold imply certain functional forms. On basis of such invariances, we have established that ψ has a power function form and W is a Prelec function (Steingrimsson & Luce, 2006, 2007; Steingrimsson, 2012a).

4 Progress Made & Results Obtained

The main results are grouped into seven categories. Because some of this research is a continuation of previous work, most detailed are those areas obtained under this grant. Item 1 and 2 involve direct expansion of the theory; items 2-6 results that are in various ways derived from the theory itself; while the last time, 7, is an alternative axiom underlying the p-additive representation, (1), that simplifies its empirical evaluation. In all cases, the progress on the associated empirical program is reported.

1. Theory of Unary Senses

The ears, eyes, and arms are binary receptors that function as cooperative pairs. Had there had been a grand designer, rather than evolution, we might have considered having 3 arms and hands—2 to hold and the middle one to manipulate. In reality, we need not consider senses beyond the binary ones, but there are numerous unary ones. Examples: taste, electric shock, vibration, force, linear extent, preference for money, etc. A crucial, and novel, realization is that the unary senses present constraints that lead to modeling that differ from the binary one. To separate the two, we refer to the binary cases as 2-D and he unary ones as 1-D.

Central is the observation that for the 2-D cases, the concatenation $x \oplus u$ of signals is over two sensory organs, in which case the important operator property of commutativity, $x \oplus u \sim u \oplus x$, is over, e.g., the two eyes or ears, and that has empirically generally been found to fail. In the unary theory, the interpretation is of simple physical concatenation of x and u, denoted $x \odot u$, which, of course, means that its physical measure is just the sum of the two intensities: $x + u$. Thus, there is no experimental issue about finding $\odot$. This case was modelled in physics by Hölder's (1901) axioms, heavily involving commutativity and associativity such that for all signals x, u, and z,

$$x \odot u \sim u \odot x, \tag{3}$$
$$(x \odot u) \odot z \sim x \odot (u \odot z). \tag{4}$$

This led, via Hölder's theorem, to a mapping $\langle X, \succsim, \odot \rangle$ into $\langle \mathbb{R}^+, \geq, + \rangle$.and a simple additive representation. In contrast, as with the binary theory, we assume magnitude production and a linking axiom between the $\circ_p$ production structure and the $\odot$ structure. These require both addition and multiplication, which means we definitely need to map into $\langle \mathbb{R}^+, \geq, +, \times \rangle$, which according to Hölder's (1901) axioms about $\langle X, \succsim, \odot \rangle$ lead to three possible, very distinct, representations, namely,

$$\varphi(x \odot y) = \varphi(x) + \varphi(y) + \delta\varphi(x)\varphi(y), \qquad \delta = -1, 0, 1. \tag{5}$$

Unlike the binary case, Luce (2012a) shows there are three form for the psychophysical function φ corresponding to the value of δ: For $x \in \mathbb{R}^+$

$$\begin{array}{ll} \varphi_0(x) = \eta x \quad (\eta > 0) & \text{if } \delta = 0 \\ \varphi_1(x) = e^{\lambda x} - 1 \quad (\lambda > 0) & \text{if } \delta = 1 \\ \varphi_{-1}(x) = 1 - e^{-\kappa x} \quad (\kappa > 0) & \text{if } \delta = -1 \end{array}$$

These predictions lead a somewhat complicated predictions of cross-modal mappings between domains, which we have worked out and are reported in Table 1 of Luce (2012a)—discussion of these results in relation to utility theory are also briefly addressed by Luce (2011).

The extant empirical literature (Stevens, 1975 summarizes) for unary domains defends power functions. And indeed, for $\delta = 0$, φ is a special case of a power function, but for. $\delta \neq 0$, these two exponential function clearly are not power functions; what gives? It may simply turn out that the extant data are not as indicative of power-functions as has been argued. . For example, Stevens (1959) reported averaged cross-modal matches between loudness of noise, vibration, and shock each linear fit to these data measured in dB appeared decent, but visual evaluations can be deceiving: Stevens did indeed note them to be far less satisfactory than desired. Our post-hoc fitting with predicted functions for cross-modal matches seem to improve the fits.

By mapping $\langle X, \succsim, \odot \rangle$ just into $\langle \mathbb{R}^+, \geq, + \rangle$ rather than into the full real numbers, Hölder and the rest of the field overlooked these solutions to his axiomatizations (Luce, 2013b). The omission did not matter for physics, but they certainly appear to matter greatly for the behavioral and economic sciences. So it was not until the work carried out under this grant that we fully realized the impact of the incompleteness of Hölder's theorem.

2. Cross-Dimensional and -Modal Matching

A. Steingrimsson and Luce (2007) showed that a certain commutative property proposed by Narens (1996) was equivalent to measurement on a ratio scale, i.e. whether $\mathbf{x}_p$ = "p times x" followed by $\mathbf{x}_{p,q}$ = "q times $\mathbf{x}_p$" is the same as when the order of p and q is reversed to q and p. That is, does $\mathbf{x}_{p,q} = \mathbf{x}_{q,p}$? But, what if, e.g., loudness for stimulus fixed at one frequency was a ratio scale, how did that scale compare to one for another fixed frequency? Luce, Steingrimsson, and Narens (2010) generalized the commutativity property to showing that under any of several cross-dimensional conditions, the property's holding was equivalent to the same ratio scale obtaining across those dimensions. They showed this to hold empirically for loudness and pitch. During this grant Steingrimsson, Luce, and Narens (2012) extended the empirical work to conclude that brightness for different hues also forms a single ratio scale.

B. We realized that axiomatic properties that could relate scales of loudness for different pitch, could equally be applied to cross modal questions such as how scales of loudness relate to scales of brightness and therefore to evaluating the intriguing prospect of establishing a single scale for multiple domains.

C. Luce (2012a) used the properties of the additive binary representation (1) and p-additive unary (5) ones to evolve theoretical predictions for cross modal matches. They are appreciably more complex than Stevens (1975, Ch. 4 for summary) recognized, but appear to have some properties exhibited in his data. The program for the unary theory (Luce, 2012a) as well as the increased complexity of cross-modal matching predictions presents a considerably enlarged empirical program than previously envisions that must be executed to answer the proposition in point B.

3. Torgerson's (1961) conjecture

Torgerson's (1961) conjectured that respondents fail to distinguish subjective differences from subjective ratios. Luce (2012b) showed that when applied to eqisections and fractionation that the conjecture implies that the function W in the production representation (2) is the identity function. That W is the identity function is firmly rejected by existing data (Ellermeier & Faulhammer, 2000; Steingrimsson & Luce, 2005b; Steingrimsson, 2011, 2012c; Zimmer, 2005). Yet, direct evaluations of the conjecture have produced mixed results, some favoring it and others rejecting it. Why might this be?

Most empirical efforts have relied on stimuli which when measured in physically, "ought to" allow respondents to make distinct evaluations of differences and ratios. However, when these stimuli are evaluated in psychological space using plausible instantiations of the summation representation (1), these stimuli appear in many cases to be ill chosen. On the basis of the psychophysical model, we evaluate the conjecture in several different ways. Using auditory stimuli, we decided to try having respondents tell us qualitatively their perceptual order between pairs of signal according to differences and ratios, which according to Torgerson they are unable to do; in brightness, we asked respondents to adjust intensity of illuminated squares to agree either in equality of differences or in ratios. In all cases, the choices of stimuli and prediction of responses are informed by our model. These predictions were formed on the basis of the representations (6, 7), detailed under the next item.

The evidence shows respondents give differential responses consistent with both ratio and differences judgments but they do so based on individual and task dependent strategies that are seemingly unaffected by ratio/difference instructions. These data thus suggest a refinement of the conjecture to include the ability to perform both ratio and difference judgments but that respondents cannot surface these judgments simultaneously. Rather, they shift between the two scenarios in a yet unspecified manner (Steingrimsson & Luce, 2012b). A happy consequence of this seemingly novel characterization of the conjecture is that the data appear to simultaneously give a plausible account for the theoretically derived evidence against the Torgerson's (1961) conjecture based on the W function (Luce, 2012b) as well as of the inconsistency in the empirical record.

4. The scales of ratings scales

The use of ratings on a scale of $\{a, a+1, a+2, ..., n\}$, $a \in \mathbb{Z}$ to evaluate "magnitude" are ubiquitous in psychology. Examples include pain in medical settings, service satisfaction in a restaurant, effect of medication on depression, and so on. Yet, despite wide-spread use, no principled analysis of the underlying scale of these ratings has, to our knowledge, appeared in the literature.

Rating scales, such as $\{1, 2, 3, 4, 5\}$ or $\{0, 1, 2, ..., 10\}$, are dimensionless, and so they are really not compatible with ratio or ordinal scales. Were they ratio scales it should be equally satisfactory to use $\{2.8, 5.6, 8.4, 11.2, 14\}$ instead of $\{1, 2, 3, 4, 5\}$ and were they ordinal, the ratings of $\{1, 2, 3, 4, 5\}$ should yield the same conclusions as their cubes $\{1, 8, 27, 64, 125\}$, neither of which we expect. Whether ratio or ordinal is assumed (often simply implicitly by way of a choice of statistical tool), then the researcher feels it appropriate to use statistical inferences appropriate to that level of scale type, which simply is not really justified just by declaration.

We suggest that when the respondent is asked to scale something using the ratings $\{0, 1, 2, ..., 10\}$, they are using these integers simply to count. That is not a case of scaling in the usual measurement sense. Of course, that leaves it entirely up to the respondent "to decide" what to count. If this view is correct, it means ratings are on absolute scales giving no more trouble using standard statistical measures such as the mean, standard deviation, median than there is with any other counts. For example, a farmer may want to estimate how many oranges will end up in a box. So the farmer counts of the numbers of oranges in several identical boxes and computes the mean and standard deviation, but no one would claim that the mean count should be viewed as a ratio scale of anything. It is a simple fact of counting.

The next question is, in a rating scale procedure, what is it that the respondent is counting? The answer is not obvious. We propose two hypotheses each of which needs to be evaluated experimentally. We wish to explore the very general conjecture that when asked to evaluate intensity on a rating scale, the respondent establishes in his or her mind an ordered partition of subjective intensity, and the count is to report which partition includes the current signal. Suppose a respondent is asked to rate on a scale of $\{1, 2, ..., n\}$ signals from the reference "interval" from x to y, where $x < y$. Then, the ratings will form a distribution from which the subjective category boundaries can be estimated. Based on the psychophysical model as well as the work on the Torgerson's conjecture, two different possible representations present themselves. The first predicts equally spaced intervals, i.e., when the respondent is asked to partition the interval (x, y) into n subjectively "equal" intervals, then the interval of the first $m \leq n$ of these equal ones, called $e_{m,n}$ must satisfy

$$
\begin{aligned}
(n-m)\left(\psi(e_{m,n}) - \psi(x)\right) &= m\left(\psi(y) - \psi(e_{m,n})\right) \\
\Leftrightarrow \psi(e_{m,n}) &= \frac{m}{n}\left[\psi(y) - \psi(x)\right] + \psi(x).
\end{aligned}
\tag{6}
$$

Replacing $\psi(x) = \alpha x^{\beta}$, we have

$$
e_{m,n} = \left[\frac{m}{n}\left(y^{\beta} - x^{\beta}\right) + x^{\beta}\right]^{1/\beta}.
$$

The other is of equal ratios, i.e., when the respondent is asked to report the signal $r_{m,n}$ that divides the interval (x, y) so that $(x, r_{m,n})$ is in the "ratio" $\frac{m}{n}$ to (x, y), then we assume that this means

$$
W\left(\frac{m}{n}\right) = \frac{\psi(r_{m,n}) - \psi(x)}{\psi(y) - \psi(x)}. \tag{7}
$$

Replacing $\psi(x) = \alpha x^{\beta}$ and solving for $r_{m,n}$

$$
r_{m,n} = \left[W\left(\frac{m}{n}\right) y^{\beta} + \left(1 - W\left(\frac{m}{n}\right)\right) x^{\beta}\right]^{1/\beta}.
$$

Data need to be collected to evaluation which of these, if either, best describe the data. Pilot data have been collected.

5. Non-Equal Matches

When respondents match one signal to another signal they do, in fact, not usually produce same physical signal. We call these non-equal matches (NEM). Special cases include the time-order error (TOE) in audition and the space-order error in vision. Steingrimsson and Luce (2012a) collected extensive data on NEM using a certain version of the method of adjustment (MA) and these data accorded well with (2). Psychologists' often used 2IFC method yielded data that are far less coherent than those from MA run in blocks. For example, the MA theory predicts either of two very different patterns of behavior, and our respondents divided equally between the two patterns. That seems to have important implications about the methodology of psychological experiments in general. In particular, one should not average over respondents—there are important individual differences.

6. Regression Effect

The regression effect is the name Stevens gave to the fact that linear equations relating signal intensity to magnitude estimation and to magnitude production do not quite agree. On the assumption that magnitude estimation satisfies the same representation (2) but with x, u the two given signals and $\mathbf{p}$ the respondent determined measure, we derived a simple theoretical account of the regression effect, which in the context of (2) asks whether it does hold for magnitude estimation. This again is both a theoretical question (Luce, 2013a) and an empirical question, which has not yet been investigated.

7. Conjoint Commutativity

The empirical study of the axioms underlying the conjoint additive representation (1) initially focused mostly on the double cancellation axiom. That axiom contains redundant features that make its evaluation a major challenge. The special case of double cancellation, the Thomsen condition, was shown, in the full axiomatic context to be equivalent to the double cancellation property absent the undesirable redundancies. Therefore, in our empirical evaluation of the theory in loudness and brightness, we studied that property. However, although we found substantial support for the it, we observed certain problems in its empirical realization, as had some previous researchers before us (Gigerenzer & Strube, 1983). We concluded the trouble stemmed from unequal number of compound estimate involved in the estimated signals subjected to statistical evaluation (a major culprit is the phenomenon of NEM reported in item 5). We showed that the property of conjoint commutativity, first proposed by Falmagne (1976), who called it the commutativity rule, is equivalent to the Thomsen condition, a result that seems to have been overlooked in the literature. We detailed the issue and subjected this property to empirical evaluation for both loudness and brightness (Luce & Steingrimsson, 2011). Subsequently we evaluated only conjoint commutativity for perceived contrast (Steingrimsson, 2012b). In all cases, the data strongly supported the conjoint commutativity and thereby conjoint additivity in these domains.

5 Significance of Results & Impact on Science

This Grant has enabled the extension of the theory described in Section 3, addition to the already strong evidence for the behavioral invariances underlying the theory described in Section 3, and extension of this evaluation to an additional domain. A number of predictions from the theory have been derived. All predictions that have so far been tested have been sustained. Continuing projects have been outlined.

The extension of the 2-D theory to include the 1-D modalities provides a complete psychophysical theory of global percepts of intensive dimensions. Prior and novel empirical work provides strong support for the 2-D theory across three separate domains and extant data lends plausibility to the 1-D theory, which in turn is associated with a clear empirical program. This results is important in three major ways. First, it is a complete theory of intensive dimensions, which by virtue of being domain independent, and verified or in principle verifiable in all intensive domains, provides a unified formal description for this area of human perception. Second, the theory captures both the general behavior of (normal and healthy) people while explicitly specifying, on the level of parameters, how individual behavior may vary. Consequently, the data are analyzed and evaluated on individual bases rather than, as is a frequent practice in psychology, averaging over individual respondents. The latter is important because unless individual scale properties are identical, which has been soundly rejected empirically, the practice of averaging over respondents limits generalizability of results to a population, and in many cases is not mathematically justified treatment of the data. Third, the theory, by way of its axiomatic approach, has the feature that the behavioral properties (axioms) themselves can be separately evaluated empirically and if found to hold lead to the representations by way of mathematical logic. There is thus no need for fitting data to the model or empirically evaluating the representations themselves. This approach, typical for classical statistic physics, to psychophysics is novel in its scope and, as is the case for physics, possesses considerable power to arrive at derived account for numerous outstanding problems as well as novel predictions. It is possible that we have but scratched the surface of what the results are telling us.

Stimuli can often vary on more than one dimension. We have extended the theory to evaluate scale properties of one while another is varied. This has allowed us to demonstrate that a single ratio scale obtained for, e.g., loudness varying in pitch and brightness varying in hue. More of these evaluations remain to be carried out. A particularly intriguing possibility is that a single ratio scale obtains for two or more domains and we have laid out the empirical program needed to answer this question. This program is simple in the case of 2-D domains, but due to novel predications for the psychophysical functions for the 1-D domains, cross-modal predictions between the 2-D and 1-D domains are unexpectedly complex requiring substantial new data. The empirical program needed for this evaluation outlined in detail. At the completion of this program, we will be able give a complete answer to questions about the scale properties of (in principle) any intensive dimension, both as it varies on other dimensions within a modality as well as how those scales relate across modalities.

The theory established the necessary and sufficient axioms for concluding ratio scale measures using magnitude estimation/production (ME/P). One aim of the grant was to extend methodological axiomatization beyond ME/P, and in particular to ratings. The paradigm of ratings is particularly ubiquitous in psychology, even as the scale properties of the obtained data are poorly established and the results are treated in ways implying unformulated and poorly founded assumptions. The primary goal is to establish the scale type of the data collected by these methods and thereby explicitly establish which statistics are justified. This work is well advanced and when completed promises to vastly improve the tools psychologists use for collecting and analyzing data obtained using ratings.

The issues related to scale properties of data collected by various methods are linked to some long-standing question. For instance, the Torgerson's conjecture (that respondents do not distinguish between judgments of the ratios of sensations and their differences), is a question that turns out to matter substantially in the establishment of measurement scales, yet has so far has resisted theoretical and empirical conclusive resolution. Our derived prediction from the core psychophysical theory turned out to address this question in an unexpected, but successful, way.

Another matter of importance is the well-known effect of temporal and/or spacial stimulus presentation, which we call non-equal matches (NEM). This effect is often thought of as context effects that can be "averaged away". We derived an account of this effect from our psychophysical theory and showed that an, a priori, assumption about how to control for it by averaging is not a sound approach and that the effect can vary by sufficiently from one data collection methods to affect the conclusions drawn from the resulting data.

Evaluating the impact of, for instance, the NEM alerted us to the need to examine both our existing data as well as the underlying behavioral properties of the theory. This analysis led us to seek an alternative for the Thomsen condition, a property central to conjoint additivity, namely the property of conjoint commutativity that from an empirical stand point is superior to the Thomsen condition.

The theory itself is mathematically explicit, formal description of individual behavior. This provides possible applications which are computational in nature. In our work on the Torgerson's conjecture, we used plausible instantiations of the parameters of the theory to choose stimuli that were appropriate in the psychological context and to show that stimuli used in prior research that seemed to be appropriate based on their physical context, were decidedly not so. This is one instance of possible application of the computational aspect of the theory. Any number of other applications involving, e.g., computer simulation of psychophysical properties seem plausible.

If it turns out the idea of a single subjective intensity scale over some intensive continua appears to hold, it suggests the practical possibility of using one easily manipulated measure as a proxy for another much more difficult measure or to manipulate. Can we use that to improve, in practice, the measurement of pain in terms of something easily controlled such as loudness, as has recently been proposed by Bartoshuk (2010)? The major problem here is that for inherently 1-D attributes, the p-additive patterns of item 1, Section 4, need to be thoroughly studied—a very long term project. We do not know for sure which 1-D attributes correspond to $\delta = 1, 0, -1$ and which can be successfully matched.

6 Publications Resulted from Research

Luce, R. D. (2011). Inherent Individual Differences in Utility. *Frontiers in Psychology, 2*, 297.

Luce, R. D. (2012a). Predictions About Bisymmetry and Cross-Modal Matches from Global Theories of Subjective Intensities. *Psychological Review, 119*, 373-387

Luce, R. D. (2012b*). Torgerson's Conjecture and Luce's Magnitude Production Representation Imply an Empirically False Property. *Journal of Mathematical Psychology, 56*, 176–178.

Luce, R. D.(2013a). Analogs In Luce's Global Psychophysical Theory Stevens' Psychophysical Regression Effect. *American Journal of Psychology, 126*, 47–52.

Luce, R. D. (2013b). The Incompleteness Of Hölder's Theorem During Most of The 20^{th} Century. To appear in a Festschrift on the occasion of Suppe's 90th birthday.

Luce, R. D., & Steingrimsson, R. (2011*). Theory and Tests of the Conjoint Commutativity Axiom for Additive Conjoint Measurement. *Journal of Mathematical Psychology, 55*, 379–385.

Steingrimsson, R., Luce, R. D, & Narens. L. (2012). Brightness of Different Hues is a Single Psychophysical Ratio Scale of Intensity. *American Journal of Psychology, 125*, 321–333.

Steingrimsson, R., & Luce, R. D (2012a*) Predictions From a Model of Global Psychophysics About Differences Between Perceptual and Physical Matches. *Attention, Perception, & Psychophysics, 74*,1668–80.

Steingrimsson, R., & Luce, R. D. (2012b). When Might It Be Sensible To Report A Rating Scale Of Subjective Intensity? In preparation.

Steingrimsson, R. (2012a*) Empirical Evaluation of a Model of Global Psychophysical Judgments for Brightness and Perceived Contrast: Forms for the psychophysical and weighting functions. In advanced preparation.

Steingrimsson, R. (2012b). Empirical Evaluation of a Model of Global Psychophysical Judgments for Perceived Contrast: I. Behavioral Properties of Summations and Productions. Submitted.

Steingrimsson, R. (2012c). Empirical Evaluation of a Model of Global Psychophysical Judgments for Perceived Contrast: II. Behavioral Properties Linking Summations and Productions. In advanced preparation.

7 References Cited

Bartoshuk, L (2010). Profile: Linda Bartoshuk. A taste for controversy. *Science, 328*, 1471-13.

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

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

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

Luce, R. D. (2002). A psychophysical theory of intensity proportions, joint presentations, and matches. *Psychological Review, 109*, 520–532

Luce, R. D. (2004). Symmetric and asymmetric matching of joint presentations. *Psychological Review, 111*, 446–454.

Luce, R. D., Krantz,D. H., Suppes, P. & Tversky, A. (1990). *Foundations of Measurement, Vol. III.* New York: Academic Press.

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

Luce, R. D. (2008). Correction to Luce (2004). *Psychological Review, 115*, 601.

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. (2011). Evaluating a model of global psychophysical judgments for brightness: II. Behavioral Properties Linking Summations and Productions. *Attention, Perception & Psychophysics, 73*, 872–885.

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

Torgerson, W. S. (1961). Distances and ratios in psychological scaling, *Acta Psychologica* 19, 201–205.

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