TECHNISCHE UNIVERSITÄT DARMSTADT

Evaluating the Ratio Scalability of Perceived Duration

An Axiomatic Study

Inaugural-Dissertation
Submitted to Technische Universität Darmstadt
Fachbereich Humanwissenschaften
in Partial Fulfillment of the
Requirements for the Degree of
Doctor Rerum Naturalium

by

M. Sc. Psych. Jana Birkenbusch
from Darmstadt

First Reviewer: Prof. Wolfgang Ellermeier, Ph.D.
Second Reviewer: Prof. Dr. Jürgen Heller
Approved: 2015/07/24
Examination: 2015/10/12

D17: Darmstadt 2015

for my family

Acknowledgment

I would like to express my gratitude to my supervisor Prof. Wolfgang Ellermeier. He patiently guided me throughout the period of my doctoral thesis and gave me the opportunity to pursue my work independently, but always provided professional and personal support whenever needed. Thank you for all the motivating conversations always pointing out a perspective to carry on, for the plenty of elaborate comments on my manuscripts, and for your bright laughter on the corridor!

I am very grateful to Prof. Jürgen Heller for kindly acting as an expert referee for this thesis.

I am very thankful to my colleagues and friends Dr. Florian Kattner, Josef Schlittenlacher, and Michael Gutjahr for accompanying me longer than three years and for creating a very pleasant atmosphere in our working group and beyond. They always supported my research as regards content, on technical issues and by giving comments on the earlier draft of this thesis. Thank you so much for being my "elder brothers"!

I am particularly grateful to Christian Wolff for the greatest "partner colloquium" I can imagine. Thank you for making me look at things from a perfectly different perspective!

I am very thankful to my student assistants Frederik Loy, Lukas Bezold, and Sebastian Warda who supported me during the data collection and with plenty of interesting ideas and questions.

Most of all, I am grateful to Matthias, the most amiable, bright and caring man I ever met. Thank you for your patient listening, for understanding me when I don't even understand myself, and for giving me the feeling of having arrived where I want to stay. Whatever happens: The time we are spending together will never pass!

The results of this dissertation production experiment: I perceived the last 94,608,000 seconds on a sensory continuum, but definitely not on a ratio scale. And even though all those marvelous moments passed off and, from now on, have to be remembered: The marks they left remain where I can always find them - in my mind and in my heart.

Abstract

The relationship between the physical intensity t of a stimulus and its perceived magnitude $\varphi(t)$ can be described by Stevens' power law $\varphi(t) = \alpha t^{\beta}$ (Stevens, 1956). The exponent of the power function β, crucial for the shape of the psychophysical function, depends on the sensory modality studied and can be estimated via direct scaling experiments. As recent developments in axiomatic measurement theory have shown, the application of direct scaling is based on fundamental assumptions concerning the participants' scaling behavior: The observers' perception of the investigated modality needs to be ordered on a sensory continuum and has to be valid on a ratio scale. Furthermore, the numbers as presented in the experimental instructions have to be processed as exact mathematical values. Narens (1996) made these implicit assumptions empirically testable by expressing them in the behavioral axioms of monotonicity, commutativity and multiplicativity. However, rigorous axiomatic testing showed that most participants fail to veridically process the numerical instructions used in production or estimation tasks. Steingrimsson and Luce (2007) have thoroughly analyzed the kind of "numerical distortion" that appears to be operating and claimed that the relationship between perceived and mathematical numbers can be described by a power function. To make this assumption empirically testable, they formulated the axiom of k-multiplicativity.

The present thesis aimed to empirically evaluate this axiomatic framework to the perception of short durations. This was accomplished by combining axiomatic testing strategies derived from different theoretical approaches (Augustin, 2008; Narens, 1996; Steingrimsson & Luce, 2007) in a single reinvestigation thereby affording a much more precise determination of the concept of ratio scalability than in the most earlier empirical studies. Furthermore, the application domain was human time perception, which had not been subjected

to this kind of axiomatic approach before.

The aim of Manuscript A ($N = 25$) was to find out whether the basic assumptions for the application of direct scaling methods are valid for the perception of short durations. Furthermore it was tested whether the estimated power law parameters are invariant under changes of the reference stimulus and thus psychologically relevant. In accordance with previous findings for other sensory continua, monotonicity held for the duration adjustments of most participants. Significant violations of the commutativity axiom were found in 12.5% of all pertinent tests, whereas multiplicativity was violated in 32% of such tests. The axioms of weak multiplicativity and invertibility were violated in over 50% of the tests, indicating a problem with psychological relevance.

Manuscript B examined whether a relationship between mathematical and perceived numbers can be described by a power function with a constant exponent and whether there is a difference between the processing of integers and fractions. To that effect, the validity of k-multiplicativity was evaluated for $N = 35$ participants. The axiomatic tests showed a power function with a constant exponent to appropriately describe the relationship between mathematical and perceived numbers. However, different values of k were found for integers and fractions indicating that they are processed differently.

Manuscript C investigated whether the functional relationship between standard duration and power law parameters can be determined. Furthermore, it tested whether the standard dependency of the power law parameters is an artifact of the ratio production procedure or whether this finding is stable even if other measures of sensitivity are used. The power law parameters were estimated for six different standard durations t $(0.1, 0.2, 0.3, 0.4, 0.5$ and 0.6 s) and compared to the corresponding Weber fractions. The results of two experiments with the same $N = 10$ participants show a positive power relationship between the duration of the standard and the estimated exponent of Stevens' power law, which can be described by the function $\beta = 0.13t^{0.3}$. A negative power relationship of the form $W = 0.84t^{-0.3}$ was found between the Weber fractions and the duration of the standard.

In conclusion, the present doctoral thesis shows that if using ratio production of temporal intervals, the measurement is based on a sensory continuum and on a ratio scale. Therefore, the application of direct scaling methods in order to determine the power law parameters for perceived duration is legitimate. It was further found that a large proportion of the participants does not process the numerals that are presented in the experimental instructions at face value, i.e., an inherent numerical distortion impedes an unequivocal interpretation of the scale values. However, this numerical distortion does not reflect an entirely arbitrary or intractable interpretation of numbers, but a well-characterized mathematical relationship – a power function with a constant exponent. Because it was shown that fractions and integers are processed differently, they should not be intermixed within one ratio production experiment.

Furthermore, the present thesis showed that modeling perceived time as a function of physical time, regardless of whether a power function or a linear relationship holds, is difficult: Even if both kinds of models seem to describe the relationship quite well, the estimated parameters depend on the magnitude of the reference stimulus used in the experiment and thus can hardly be interpreted in a psychologically relevant way.

However, the influence of the standard on the size of the exponent seems to be systematic: Increasing standard durations go along with increasing exponents. Weber fractions measured under identical conditions were found to decrease with increasing standard durations and thus, combining both findings, it can be assumed that differential sensitivity for duration perception increases between 100 and 400 ms and remains at a constant level between 400 and 600 ms. A bias due to the ratio production procedure is thus ruled out.

Zusammenfassung (Abstract in German)

Der Zusammenhang zwischen der physikalischen Intensität t eines Reizes und dessen wahrgenommener Größe $\varphi(t)$ kann mit Hilfe von Stevens' (1956) Potenzgesetz, $\varphi(t) = \alpha t^{\beta}$, beschrieben werden. Der Exponent der Potenzfunktion β, der für die Form der psychophysischen Funktion entscheidend ist, hängt von der untersuchten sensorischen Modalität ab und wird mittels direkter Skalierung geschätzt. Wie jüngste Entwicklungen in der axiomatischen Messtheorie zeigen, basiert die Anwendbarkeit der direkten Skalierung jedoch auf einigen Grundannahmen über das Skalierungsverhalten der Versuchsperson: Die Wahrnehmung von Reizen der untersuchten Modalität muss auf einem sensorischen Kontinuum sowie auf einer Verhältnisskala beruhen. Außerdem müssen die in den Instruktionen verwendeten Zahlen von der Versuchsperson veridikal, d.h. wie tatsächliche mathematischen Zahlen verarbeitet werden. Um diese impliziten Grundannahmen empirisch testbar zu machen, entwickelte Narens (1996) die behavioralen Axiome Monotonie, Kommutativität und Multiplikativität. Eine strenge Testung dieser Axiome ergab, dass die meisten Versuchspersonen die dargebotenen Zahlen nicht veridikal verarbeiten. Steingrimsson und Luce (2007) untersuchten diese "numerische Verzerrung" und vermuteten stattdessen, dass das Verhältnis zwischen mathematischen und wahrgenommenen Zahlen als Potenzfunktion dargestellt werden kann.

Das Ziel dieser Doktorarbeit bestand darin, diese Axiomatik für die Wahrnehmung kurzer Zeitdauern empirisch zu überprüfen. Dies wurde durch die Zusammenführung axiomatischer, aus unterschiedlichen theoretischen Ansätzen (Augustin, 2008; Narens, 1996; Steingrimsson & Luce, 2007) abgeleiteten Testverfahren erreicht. In einer gemeinsamen Testung evaluiert, bietet dieses

Verfahren eine wesentlich genauere Bestimmung des Konzepts der Verhältnisskalierbarkeit als die meisten vorangegangenen Untersuchungen. Außerdem stand die menschliche Zeitwahrnehmung im Fokus der Untersuchung, die dieser axiomatischen Prüfung zuvor noch nicht unterzogen wurde.

Ziel von Manuskript A ($N = 25$) war es herauszufinden, ob die impliziten Grundannahmen der direkten Skalierung für wahrgenommene Dauer gelten. Weiterhin wurde untersucht, ob die geschätzten Parameter von Stevens' Potenzgesetz unter Änderung des Standards invariant bleiben und somit psychologisch relevant sind. Im Einklang mit Ergebnissen zu anderen Sinnesmodalitäten zeigte sich, dass Monotonie für die meisten Versuchspersonen gültig ist. Kommutativität wurde in 12.5% der Tests verletzt, während Multiplikativität in 32% der Tests ungültig war. Schwache Multiplikativität und Invertibilität wurden in über 50% der Tests verletzt, was auf ein Problem der psychologischen Relevanz hinweist.

Manuskript B untersuchte, ob der Zusammenhang zwischen mathematischen und wahrgenommenen Zahlen von einer Potenzfunktion mit einem konstanten Exponenten darstellbar ist und ob ein Unterschied in der Verarbeitung von Brüchen und ganzen Zahlen besteht. Deshalb wurde das Axiom der k-Multiplikativität für $N = 25$ Versuchspersonen getestet. Die Prüfung des Axioms ergab, dass der Zusammenhang zwischen mathematischen und wahrgenommenen Zahlen sehr gut von einer Potenzfunktion mit einem konstanten Exponenten beschrieben werden kann. Die unterschiedliche Verarbeitung von ganzen Zahlen und Brüchen zeigte sich u.a. dadurch, dass verschiedene Werte für k gefunden wurden.

In Manuskript C wird die Bestimmung des Funktionszusammenhangs zwischen der Dauer des Standardreizes und der Größe der Potenzgesetz-Parameter beschrieben. Weiterhin wurde überprüft, ob die Standard-Abhängigkeit der Parameter auf die Methode der Verhältnisherstellung zurückzuführen ist oder ob tatsächlich eine Änderung der differentiellen Sensitivität vorliegt. Deshalb wurden die für sechs verschiedene Standarddauern t $(0.1, 0.2, 0.3, 0.4, 0.5$ und $0.6)$ geschätzten Potenzgesetz-Parameter mit den entsprechenden Weber-Brüchen verglichen. Die Ergebnisse der beiden Experimente mit $N = 10$ Versuchspersonen zeigten eine positive Potenzfunktion zwischen der Dauer des Standards und dem Exponenten der Form $\beta = 0.13t^{0.3}$. Zwischen den Weber-Brüchen und der Dauer des Standards wurde eine negative Potenzfunktion der Form $W = 0.84t^{-0.3}$ gefunden.

Zusammenfassend zeigte diese Doktorarbeit, dass die Wahrnehmung kurzer Zeitdauern auf einem sensorischen Kontinuum und einer Verhältnisskala beruht. Deshalb kann die Verwendung von direkter Skalierung zur Schätzung der Potenzgesetz-Parameter für Zeit als gerechtfertigt betrachtet werden. Weiterhin wurde gezeigt, dass ein Großteil der Versuchspersonen die im Experiment verwendeten Zahlen nicht wie mathematische Zahlen verarbeitet und daher eine inhärente numerische Verzerrung die eindeutige Interpretation der Skalenwerte beeinträchtigt. Trotzdem stellt diese numerische Verzerrung kein völlig willkürliches und undefinierbares Zahlenverständnis dar, sondern einen mathematisch gut beschreibbaren Zusammenhang – eine Potenzfunktion mit einem konstanten Exponenten. Da gezeigt wurde, dass ganze Zahlen und Brüche unterschiedlich verarbeitet werden, sollten diese innerhalb eines Skalierungsexperiments nicht kombiniert werden.

Weiterhin zeigte diese Arbeit, dass die Modellierung von psychophysischen Funktionen, egal ob linear oder exponentiell, schwierig ist: Obwohl beide Modelle den Zusammenhang zwischen physikalischer und wahrgenommener Zeit gut abbilden können, hängen die entscheidenden Parameter von der Dauer des im Experiment verwendeten Standards ab und können deshalb nur bedingt interpretiert werden.

Der Zusammenhang zwischen der Standarddauer und der Größe des Exponenten scheint jedoch systematisch zu sein: Mit steigender Standarddauer zeigten sich steigende Exponenten und, unter identischen Bedingungen ermittelte, sinkende Weber-Brüche. Insgesamt kann also angenommen werden, dass die Sensitivität für wahrgenommene Dauer zwischen 100 und 400 ms steigt und sich zwischen 400 und 600 ms auf konstantem Level einpendelt. Eine von der Methode der Verhältnisherstellung verursachte Verzerrung konnte ausgeschlossen werden.

Contents

List of Figures

List of Tables

Chapter 1

Theoretical Background

1.1 The Psychophysical Law

This first section illuminates the psychophysical law as the central issue of the present doctoral thesis. It discusses Stevens' power law as a method to relate the physical intensity and the perceived magnitude of a stimulus, and describes direct scaling methods necessary to estimate the parameters of the psychophysical function as well as their difficulties and advantages.

A psychophysical law describes the relationship between the physical intensity of a stimulus and its perceived, i.e., its psychological magnitude. The determination of the form of this psychophysical function is one of the fundamental questions in psychophysics. Ptolemy (about 150 A.D.) proposed to measure the size of stars by their apparent brightness and thereby provided one of the first attempts of psychophysical scaling. More than 1500 years later, the mathematician Bernoulli (1738) was the first who attempted to describe the relationship between the actual – physical – amount of money to its perceived – psychological – value and proposed that the perceived value of money increases at a decreasing rate as the actual amount of money grows. He thereby formulated a logarithmic function for the relationship between stimulus magnitude, i.e., the amount of money, and sensation, i.e., its perceived value. The same proposition was expressed by Fechner (1860) in his work "Elemente der Psychophysik". Fechner was the first psychologist who assumed that increasing a stimulus' intensity by a constant ratio leads its perceived magnitude to increase by a constant amount, i.e., doubling the stimulus intensity should always result in a sensation magnitude by the same increment. Fechner's law had an foundational influence on psychophysics and dominated research in many fields – until Stevens (1953) and others (Fletcher & Munson, 1933; Richardson & J. S. Ross, 1930) raised serious doubts about the validity of Fechner's law. Stevens, who used ratio scaling methods (see section 1.1.2) to determine the psychophysical function for brightness and loudness, could not find parameters corresponding to a logarithmic function, but found both to be proportional to the cube root of the stimulus' physical intensities.

1.1.1 Stevens' Power Law

In the "new psychophysics", it is assumed that, according to Stevens (1957), the relationship between the physical intensity of a stimulus t and its perceived magnitude $\varphi(t)$ can be described by a power function of the form:

$$\varphi(t) = \alpha t^{\beta}, t > 0 \tag{1.1}$$

The parameter α is an arbitrary factor depending on the scale units used, whereas the parameter β is thought to depend on the sensory modality and determines the exact shape of the power function. If the value of β is > 1, the perceived intensity grows faster than the intensity of the physical stimulus. If β is < 1, what would be proposed by Fechner's law, the function is negatively accelerated, and if $\beta = 1$, there is a directly proportional relationship between physical and perceived stimulus intensity, i.e., the relationship can be described by simple linear function resulting in a straight line (Gescheider, 1997).

1.1.2 Direct Scaling Methods

Measuring perception on a ratio scale has always been a target criterion in psychology (Plateau, 1872). Merkel (1888), who investigated the validity of Weber's law for loudness was the first who reasoned which magnitude corresponds to a stimulus that is perceived to be twice as intense as an original stimulus. To that effect, he presented a stimulus of a certain loudness and asked the participants to produce an corresponding stimulus that is perceived to be twice as loud as the first. Some years later, Fullerton and Cattell (1892) used a comparable magnitude production method requiring the participants to adjust comparison stimuli according to a certain ratio to a standard stimulus. The two procedures have in common that ratios between stimulus intensities are to be adjusted and thus, the produced stimulus intensities intended to represent measurements valid on ratio scale level.

However, these experiments were carried out because the indirect scaling of loudness showed Weber's law not to hold. Therefore, the application of this "new method" aimed to find another valid psychophysical law describing the relationship between physical intensity and perceived loudness. During the following years, different modalities were investigated using different methodical procedures satisfying the requirements of ratio scaling: magnitude estimation, ratio estimation, magnitude production and ratio production. In the following, these methods are described in detail.

Magnitude Estimation

The method of magnitude estimation introduced by Richardson and Ross (1930) is one of the most commonly applied methods in ratio scaling. In a magnitude estimation procedure, the participant is required to directly estimate a stimulus' intensity by denominating it with a numerical value. Stevens (1958) described two different procedures of magnitude estimation: According to the first procedure, a standard stimulus is presented to the participant, which is assigned to a certain numerical value. In the following trials, the participant is asked to describe the intensity of other (comparison) stimuli by assigning appropriate numerical values relative to the standard stimulus. For example, if the standard is assigned a value of 10, a stimulus that is perceived to be three times as intense as the standard is assigned a value of 30, and a stimulus that is perceived to be half as intense as the standard is called 5. In the second procedure, sometimes called "free" or "absolute" magnitude estimation, no standard stimulus is presented and the participant is simply asked to give direct numerical estimations of stimulus intensity. The psychophysical function is estimated by plotting the averaged magnitude estimates as a function of the physical stimulus intensities.

Magnitude Production

The method of magnitude production is an "inverted" version of magnitude estimation: The participant is given a numerical value and is required to adjust the stimulus' intensity corresponding to this value. For example, the participant is given the number 10 and is then asked to produce a stimulus that is perceived to be as intense as 10. If a number 20 is presented, the produced stimulus intensity is assumed to feel twice as intense as the previously adjusted stimulus 10. The magnitude production procedure requires the stimuli to be continuously adjustable or, at least, to be adjusted in very small steps. Magnitude production can be used to check the validity of magnitude estimation and vice versa. A reciprocal validation can be helpful especially to detect a regression to the mean (S. S. Stevens & Guirao, 1962), i.e., the participants' tendency to avoid extremely low or high judgments even though they may be appropriate to their perception. The "method of numerical magnitude balance" (Hellman & Zwislocki, 1968) provides a technique that combines magnitude production and estimation and estimates particularly unbiased psychophysical functions via geometric means. Altogether, strong agreement was found between magnitude production and magnitude estimation, yielding a high validity of both methods and the corresponding scales (Marks & Gescheider, 2002).

Ratio Estimation

The method of ratio estimation requires the participant to estimate the numerical ratio of two stimulus magnitudes. Thus, the method is very similar to magnitude estimation, but usually requires specifying ratios of sensation magnitudes (Ekman, 1958). For example, if two stimuli are presented to the participant and the first stimulus is perceived to be twice as intense as the second stimulus, then the participant might describe them by the ratio of 2 : 1, 20 : 10 or 200 : 100. Ratio estimation can be used to validate the results of a ratio production experiment and vice versa.

Ratio Production

The method of ratio production, sometimes refered to as fractionation – especially when proportions < 1 are to be adjusted – was invented by Churcher

(1935) and developed by Stevens (1936). The participant's task is to adjust a variable comparison stimulus according to a certain subjective ratio to a fixed standard stimulus. Many different psychophysical scales, such as loudness, brightness, weight, taste, duration, pain and vibration were generated by means of ratio production (S. S. Stevens, 1975; S. S. Stevens & Galanter, 1957). As in magnitude production, the method requires the stimuli to be adjustable on a sensory continuum. However, if the stepsize is small enough, an adjustment in discrete steps is also possible, since the participant is often instructed to gradually approximate the intensity of the comparison stimulus until it is clearly perceived to correspond to the given ratio to the standard. Many ratio production experiments employ ratios of $1 : 2$ and check the validity of the participants' adjustments by instructing them to "undo" a $\times 2$ adjustment by means of a $\times \frac{1}{2}$ adjustment. If this validity check fails, a bias due to the numerical values used or the standard magnitude may be assumed (Augustin, 2008).

Difficulties of Direct Scaling

One problem central to this thesis is that psychophysical scaling is, however, prone to context effects. Mellers and Birnbaum (1982), Garner (1954) as well as Ward et al. (1996) showed that a change in the standard stimulus range has an influence on the shape of the psychophysical function. Furthermore, the number examples given in the experimental instruction (Robinson, 1976), as well as the number values assigned to the standard stimuli (Beck & Shaw, 1965) or even the entire experimental context might have an influence on the size of the estimated power law parameters. They might also vary under changes of the physical measurement scale f (Narens & Mausfeld, 1992) and the size of the standard (Augustin, 2008) used in the scaling experiment. Therefore, the psychological relevance of the parameters has been called into question (Lockhead, 1992).

In contrast to this point of view, other investigators have argued that finding the "true" exponent is still possible (M. Teghtsoonian & R. Teghtsoonian, 2003; R. Teghtsoonian, 2012). Engen and Tulunay (1957) further showed that practice and a standardized experimental setting serve to minimize context effects.

Furthermore, Anderson (1970) questions whether the participants' judgments in a direct scaling experiment are proportional to sensation magnitudes, whereas Birnbaum (1982) argues that direct scaling has no advantages over category ratings, because both are in fact no more than ordinal scales of subjective value.

Advantages of Direct Scaling

However, Stevens provides a very easy and straightforward procedure to investigate perceptional and judgmental attributes by assuming that participants are able to directly describe the perceived magnitude of a stimulus (S. S. Stevens, 1956). Especially in contrast to the Fechnerian approach (Fechner, 1860) using the indirect route via discriminability (e.g., Dzhafarov & Colonius, 1999), direct scaling provides considerable advantages: It usually needs a smaller number of trials to determine the participant's sensitivity to a certain sensory modality and thus provides an economic benefit. Furthermore, it uses ratios of stimulus intensities rather than just noticeable differences (JNDs) as the perceptional basis for the estimated psychophysical function. The output of a perceptual process, i.e., the perceived magnitude, is not measured in the units of the input, i.e., the underlying physical continuum – as is in indirect scaling. Instead, it is measured in sensation units and thereby provides a more detailed picture of the investigated sensory system (Gescheider, 1997). Additionally, in contrast to other findings (Pradham & Hoffman, 1963; J. C. Stevens & Guirao, 1964), individual psychophysical functions can be determined and thus, it can be assumed that general power functions are not an artifact of averaging.

Furthermore, direct scaling methods were repeatedly applied to several sensory modalities (Jones, 1974; S. S. Stevens & Galanter, 1957). Therefore, they provide a high comparability of findings within a certain modality as well as across different modalities. Their wide application in spite of the identified (and partially corrected) weaknesses indicates their outstanding and noteworthy position in psychophysics.

1.2 Axiomatic Measurement Theory

A particular scaling procedure and thus direct scaling as well, lacks scientific and psychological relevance until it is validated by means of measurement theory.

However, there is only a small number of efforts to develop an empirically testable framework based on axiomatic measurement theory (Ellermeier & Faulhammer, 2000; Luce, 1959; Narens, 1996) that can be used to evaluate the applicability of direct scaling. Since some of these approaches cannot be applied to all types of direct scaling procedures, they are not even generalizable. Nonetheless, this thesis combines several axioms – each constituting a necessary condition for a certain aspect of scietific relevance – into an axiomatic framework that can be used to evaluate the direct ratio scaling.

This section introduces axiomatic measurement theory and explains the particular tests employed in the studies reported in this thesis. The first section explains axioms formulated by Narens (1996) to empirically evaluate assumptions fundamental to direct scaling. The second section provides an extended axiomatic approach by Steingrimsson and Luce (2007) testing the relationship between mathematical and perceived number, whereas the third paragraph shows an axiomatic extension by Augustin (2008) to test the psychological relevance of the parameters estimated for the psychophysical function.

Mathematically formulated properties that have to be valid for a certain empirical reference system are called *axioms*. Axiomatic testing can be used to derive conclusions in form of mathematical or logical propositions for the underlying basic set (Orth, 1974). In representational measurement theory, one can distinguish cognitive and behavioral axioms (Narens, 1996): The cognitive axioms describe the relationship between the participants' unobservable sensation of a stimulus' intensity and its numerical representation. In contrast to the behavioral axioms, they cannot be empirically tested. The behavioral axioms characterize the participants' behavior in a scaling experiment and relate their numerical representation to the number words used to describe the stimulus' intensity. The behavioral axioms can be used to evaluate implicit assumptions fundamental to direct scaling and will be discussed in the following.

There are two basic assumptions fundamental to the application of Stevens's direct scaling methods (Narens, 1996): It is assumed that the participants are able to estimate or to produce the perceived intensities on a sensory continuum and on a ratio scale. Furthermore, it is assumed that the participants interpret the numerals – presented in the experimental instruction and used to describe the sensation magnitudes – like rational mathematical numbers. Stevens himself never explicitly tested these assumptions although they are basic to the approach of direct scaling. Therefore, Narens (1996) formulated mathematical axioms to provide a possibility to empirically test these assumptions.

Notation

In the current thesis, a notation according to Luce (2002) will be used to describe the instructions and stimuli applied in a ratio production experiment. Usually, the participant is required to adjust the magnitude of a comparison stimulus w, x, y or z to a ratio of $\mathbf{p}$, $\mathbf{q}$ or $\mathbf{r}$. The latter describe the perceived magnitude of the standard stimulus t. The notation $(x, \mathbf{p}, t)$ represents a participant's adjustment x, which is perceived to be $\mathbf{p}$ times as intense as the standard t, with the boldface letter referring to the number word used in the magnitude production instructions.

1.2.1 Narens' Axioms

The axiom of *ordering* (Narens, 1996) or *monotonicity* (Augustin, 2008) can be tested to prove the assumption of stimulus intensity being perceived on a sensory continuum. According to Narens, it is a necessary condition for the subsequently tested axioms of commutativity and multiplicativity as well as for any scaling procedures at all, because even the categories of an ordinal scale can be arranged in an ascending or descending and therefore monotonic order. The axiom of monotonicity is formulated as:

$$\text{If } (x, \mathbf{p}, t) \in E \text{ and } (y, \mathbf{q}, t) \in E, \text{ then } p > q \Leftrightarrow x \succ y. \tag{1.2}$$

This axiom implies that if x has been adjusted to appear $\mathbf{p}$ times as intense as the standard t and another adjustment y is $\mathbf{q}$ times as intense as a standard t, and $\mathbf{p}$ is greater than $\mathbf{q}$, then the produced magnitude of $t \times \mathbf{p} = x$, must be greater than the magnitude of $\times\mathbf{q}$, y. The subsequently tested axiom is called commutativity and evaluates whether the participants' perception of the investigated modality is based on a ratio scale[1]. It is formulated as:

[1]Because for each positive integer p and each φ being an element of a ratio scale $\mathcal{S}$, $\varphi(\mathbf{p})$ is a function that is multiplication by some positive real c, with not necessarily $c = p$.

$$\text{If } (x, \mathbf{p}, t) \in E, (z, \mathbf{q}, x) \in E, (y, \mathbf{q}, t) \in E, \qquad (1.3)$$
$$\text{and } (w, \mathbf{p}, y) \in E, \text{ then } z = w.$$

The axiom of commutativity holds, if the order of two successive ratio productions $\times\mathbf{p} \times \mathbf{q}$ does not affect the finally adjusted stimulus magnitude, i.e., the production of $\times\mathbf{p} \times \mathbf{q}$ results in the same outcome magnitude as $\times\mathbf{q} \times \mathbf{p}$. Testing the axiom of multiplicativity shows whether the participants are able to interpret the scale values as scientific numbers, i.e., whether they have a veridical understanding of numbers[2]. The axiom of multiplicativity is formulated as:

$$\text{If } (x, \mathbf{p}, t) \in E, (z, \mathbf{q}, x) \text{ and } r = qp, \text{ then } (z, \mathbf{r}, t) \in E. \qquad (1.4)$$

Multiplicativity holds, if the outcome magnitude of a successive production sequence $\times\mathbf{p} \times \mathbf{q}$ is equal to the outcome of a single production of $\times\mathbf{r}$ with r being the mathematical product of p and q. During the past 20 years, the axiomatic approach to direct scaling pioneered by Narens (1996) has been extended, e.g., by Luce and colleagues (Luce, 2002, 2008; Luce, Steingrimsson, & Narens, 2010). One of their recent interpretations concerning the axiom of multiplicativity argues that a veridical interpretation of numbers as presented in the instructions and thus the validity of multiplicativity is not mandatory for direct ratio scaling. Luce argues that if the axiom of commutativity is valid and thus ratio scalability for the investigated modality can be assumed, it may be implied that the participants interpret the numbers as some ratio, although not as the exact ratio stated in the instructions.

1.2.2 Steingrimsson's and Luce's Axioms

However, the axiom of multiplicativity was found to be violated for many sensory modalities (Augustin & Maier, 2008; Ellermeier & Faulhammer, 2000; Kattner & Ellermeier, 2014; Steingrimsson & Luce, 2007; Zimmer, 2005) and thus, the assumption of participants having a veridical interpretation of numbers had to

[2]Because for each positive integer p and each φ being an element of a ratio scale S, $\varphi(\mathbf{p})$ is a function that is multiplication by the integer p.

be rejected. Therefore, Steingrimsson and Luce (2007) formulated a weaker axiom to test whether the transformation function between perceived and mathematical numbers follows a power relationship with a constant exponent multiplied by a constant k. This axiom, which is called *k-multiplicativity*, is formulated as:

$$\text{If } (x, \mathbf{p}, t) \in E, (z, \mathbf{q}, x) \text{ and } r = kqp, \text{ then } (z, \mathbf{r}, t) \in E. \tag{1.5}$$

That means, k-multiplicativity holds, if the outcome magnitude of a successive production sequence $\times\mathbf{p} \times \mathbf{q}$ multiplied by a constant factor k is equal to the outcome of a single production of $\times\mathbf{r}$ and if k is invariant over several pairs of $\mathbf{p}$ and $\mathbf{q}$. Because empirical observations have shown fractions to be processed in a different way than are integers, one may have to distinguish experimental instructions using fractions ($p < 1$) and integers ($p \geq 1 \wedge p \in \mathbb{N}$).

1.2.3 Augustin's Axioms

For direct scaling, a further crucial question is whether the estimated exponent of the power function is invariant under certain transformations and especially under changes of the standard stimulus t constituting the basis for the participants' estimates or adjustments. In order to examine the dependency of the exponent β on the standard t, Augustin (2008) proposed two further empirically testable axioms: Weak multiplicativity and invertibility.

Weak multiplicativity is formulated as:

$$\begin{gathered}\text{For } t, y, z \in X \text{ and a real number } p > 0, (y, \mathbf{p}, t) \in E, (z, \mathbf{1/p}, y) \in E \\ \Rightarrow (z, \mathbf{1}, t) \in E.\end{gathered} \tag{1.6}$$

The axiom of weak multiplicativity is valid, if the stimulus intensity resulting from successive adjustments $\times\mathbf{p} \times \frac{1}{\mathbf{p}}$ is equal to the stimulus intensity resulting from the basic adjustment with $\mathbf{p} = 1$. Augustin's axiom of weak multiplicativity appears to be very similar to Narens' axiom of multiplicativity with a crucial difference: Multiplicativity has to hold for all cases in which $\mathbf{p} > 0$ and $\mathbf{q} > 0$, whereas weak multiplicativity is a special case of multiplicativity with $\mathbf{q} = \frac{1}{\mathbf{p}}$. That means even if the axiom of multiplicativity is violated in general, this special case, i.e., weak multiplicativity, might hold.

The axiom of *invertibility* is formulated as:

$$\text{For } t, y \in X \text{ and } \mathbf{p} > 0, (y, \mathbf{p}, t) \in E \Leftrightarrow (t, \frac{1}{\mathbf{p}}, y) \in E. \tag{1.7}$$

In other words, invertibility holds, if the outcome intensity of a stimulus resulting from successive adjustments $\times\mathbf{p} \times \frac{1}{\mathbf{p}}$ is equal to the stimulus intensity of the standard t or, put more simply, if it is possible to "undo" a $\times\mathbf{p}$ adjustment by asking the participant to produce its reciprocal $\times\frac{1}{\mathbf{p}}$. So weak multiplicativity and invertibility are very similar, but differ in whether the successive adjustment resulting from $\times\mathbf{p} \times \frac{1}{\mathbf{p}}$ is equal to the adjustment of $\times\mathbf{1}$ in the first case and the actual standard in the second case. As Augustin (2008) stated, both axioms are necessary and sufficient conditions for the exponent of Stevens' power law to be invariant under changes of the standard t.

1.2.4 Axiomatic Testing

Testing the previously presented axioms requires data produced in a direct scaling experiment (Narens, 1996). Typically, ratio production experiments are used for this purpose, instructing the participants to adjust the stimulus intensity of a comparison according to a certain ratio $\mathbf{p}$ or $\mathbf{q}$ to a standard. To evaluate the different axioms, certain combinations of $\mathbf{p}$ and $\mathbf{q}$ are required. Evaluating monotonicity requires several ordered $\mathbf{p}$, e.g., $\mathbf{p} = 2, 3, 4, 5$ in so-called *basic trials*, i.e., trials with the same fixed standard t. To test commutativity, so-called *successive trials*, in which the individual adjustments produced in the basic trial are used as standards, are needed. For example, one might choose the combinations of $(p, q) = 2, 3$ and $(q, p) = 3, 2$ or $(p, q) = 2, 4$ and $(q, p) = 4, 2$. The evaluation of multiplicativity requires basic- ($\times\mathbf{r}$) as well as successive trials ($\times\mathbf{p} \times \mathbf{q}$), e.g., $(p, q) = 2, 3$ ($(q, p) = 3, 2$, respectively) and $\mathbf{r} = 6$ or $(p, q) = 2, 4$ ($(q, p) = 4, 2$, respectively) and $\mathbf{r} = 8$. Testing k-multiplicativity calls for several pairs of $\times\mathbf{p} \times \mathbf{q}$ and $\times\mathbf{r}$ to check whether k is constant over these pairs. For all of these $\mathbf{p}$, $\mathbf{q}$, and $\mathbf{r}$, fractions as well as integers may be employed (Luce, 2002). To test weak multiplicativity and invertibility, it is necessary to mix fractions and integers as ratio production factors, because $\mathbf{p} = \frac{1}{\mathbf{p}}$ is required. For example, successive trials with $\mathbf{p} = 3$ and $\mathbf{q} = \frac{1}{3}$ might be used. In addition, basic trials with $\mathbf{p} = 1$ are needed.

Altogether for meaningful statistical testing, all types of adjustments should be made, at least, ten times. Because of the lengthly adaptive adjustment procedure and the required number of repetitions, experiments to collect data for axiomatic tests usually need considerable time. Therefore, it is helpful to divide the experiment into different test sessions to assure the participants' attention and alertness. For the same reason, it is also difficult to test all the axioms mentioned in one single experiment, because too many different types of trials are required.

Even though it is possible to test axioms by collapsing data across the entire sample and comparing overall mean adjustments, major virtue of the measurement-theory approach is that behavioral axioms can be evaluated separately for each participant and thus can be helpful to draw conclusions about the scaling behavior of each individual in the sample (Ellermeier & Faulhammer, 2000; Steingrimsson & Luce, 2007). For this reason, ratio scaling experiments that are conducted in order to evaluate behavioral axioms usually use small sample sizes with $N = 5$ up to $N = 15$.

1.3 Duration Perception

Since there is a very wide range of theories and fascinating findings about the perception of duration and the processing of temporal information, this section runs the risk of getting off track. Therefore, it discusses only issues that are directly relevant for the research questions of this thesis. The first section examines basic findings on duration perception implying some decisions on the experimental setting of the experiments conducted. The second section presents essential timing theories and models that are necessary for understanding timing mechanisms. The third section concentrates on the psychophysics of duration perception, since this is the central aspect of this thesis. Especially findings on the relationship between physical and perceived duration, i.e., the psychophysical function, and work on measuring temporal discrimination, i.e., JNDs are presented. For a more detailed review on timing behavior, see Allan (1979), Fraisse (1963, 1978), or, more up to date, Grondin (2008, 2010).

1.3.1 Basic Findings

What is time? The notion of time can distinguish between two different concepts which result from our individual experience of change (Fraisse, 1963). The concept of succession corresponds to the fact that two or even more events can be perceived as different and thus organized sequentially. It is based on the experience of continuous changes and on the experience of the present to become tha past. The concept of duration corresponds to the interval which lies between two successive events.

Prospective vs. Retrospective Timing

Experiments on duration perception differ in a methodically with respect to the moment at which participants are instructed to give a duration estimate (S. W. Brown & Stubbs, 1992; A. D. Eisler, H. Eisler, & Montgomery, 2007; Stubbs, 1988): Prospective timing is measured when participants are instructed at the beginning of the experiment to estimate the duration of the intervals that will be presented. In retrospective timing tasks, participants are asked to estimate the duration of a past event after its occurence. In another kind of retrospective setting, participants' timing performance might be indirectly tested and they are informed about that purpose at the end of the experiment.

Prospective timing is often used in tasks employing short duration intervals (milliseconds and seconds), whereas retrospective timing is preferred in studies using long intervals (minutes and hours) (Bisson, Tobin, & Grondin, 2008). Whereas prospective timing produces more accurate estimations, because the attentional focus lies on the perception of duration itself, retrospective timing focusses on the task (S. W. Brown, 1985). Thus, memory effects are included in the duration estimates making them less accurate (Block & Zakay, 1997). In this doctoral thesis, prospective timings are employed throughout, because axiomatic tests require accurate timing precluding memory effects. Furthermore, duration production tasks are difficult to present in a retrospective task.

Critical Durations

Since time is infinite, it is necessary to identify meaningful duration ranges that can be investigated in a sensible way. Even if humans seem to have timing mechanisms operating for very short (neural activity) as well as very long durations (circadian rhythms), different internal structures can be assumed to be involved in timing (Wackermann, 2007). Fraisse (1963) distinguished between *duration perception* and *duration estimation*: Duration perception refers to the concept of *presence*, i.e., the temporal extend of a stimulus that can be perceived at one given moment in time without rehearsal during or after the stimulus presentation. According to Fraisse, presence ranges up to 5 s, whereas Pöppel (2004) assumes the "sense of nowness" to end at an upper limit of 3 s for neurocognitive reasons. Grondin (2010) argues time perception to occur between 100 ms and a few seconds. Duration perception in this range has been thoroughly investigated, because it is important for behavior such as speech and music perception or motor coordination. Furthermore, durations in this range are investigated in detail since they are close to the "indifference interval" (about 700 ms) meaning they are neither over- nor underestimated (H. Eisler, A. D. Eisler, & Hellström, 2008). Several studies (Drake & Botte, 1993; Fraisse, 1967; Friberg & Sundberg, 1995) reported the highest discrimination performance for intervals between 300 and 800 ms. In contrast to duration perception, duration estimation is less sensory and needs cognitive and memory resources (Hellström & Rammsayer, 2004). For duration estimation, participants often use segmentation strategies in order to divide the interval into equal shorter segments of time, e.g., by counting (Grondin, Meilleur-Wells, & Lachance, 1999). To avoid these complications, the present doctoral thesis employed short intervals in the temporal range of duration perception. Due to the ratio production task, care was taken that the total duration of standard, inter-stimulus intervals and comparison did not exceed a limit of 5 s to preclude over- or underestimation and memory effects.

Filled vs. Empty Intervals

Another difference in temporal discriminability occurs because the beginning and the end of a given duration may be indicated in different ways. If a continuous signal is presented between the onset and offset of an interval, this interval is called *filled*. *Empty*, in contrast, means that an onset and an offset marker are presented at the beginning and at the end of the interval to be judged. The discrimination performance of short intervals is often stated to be better in tasks using filled intervals than in experiments employing empty intervals (Abel, 1972a, 1972b; Rammsayer & Lima, 1991). However, other studies find that empty intervals are easier to discriminate than filled ones (Grondin, 1993; Henry, 1948; Small & Campbell, 1962). For duration (re-)production, Thomas and Brown (1974) as well as Long and Beaton (1980) reported filled intervals to result in longer estimates than empty ones. Furthermore, it is assumed that timing performance depends on the type of the markers, on the range of the investigated durations, on individual differences and on the methodical procedure employed in the experiment. Rammsayer (2010a) assumed temporal processing of filled intervals to be functionally different from processing of empty intervals. The experiments reported in this doctoral thesis all use filled intervals, because participants reported comparing two continuous stimuli to be less confusing than keeping track of four markers constituting three intervals one of which is the inter-stimulus interval.

Auditory vs. Visual Stimuli

Comparing the discriminability of durations from different sensory modalities revealed a greater temporal acuity for auditory than for visual stimuli (Penney & Tourret, 2005; van Wassenhove, 2009). Investigating filled intervals of identical duration showed auditory stimuli to be perceived to be longer than visually presented stimuli (Behar & Bevan, 1961; Goldstone & Goldfarb, 1964; Goldstone & Lhamon, 1974; S. S. Stevens & Greenbaum, 1966). It was further found that auditory stimuli are perceived to last longer than visual stimuli, even in the case of simultaneous presentation. This finding could be replicated for interrupted stimuli (Walker & Scott, 1981). Furthermore, it was found that both visual and auditory temporal resolution power correlate with psychometric intelligence (Haldemann, Stauffer, Troche, & Rammsayer, 2012) and further depends on whether the intervals are filled or empty. The effect was demonstrated in several experiments using a number of durations and various timing tasks. Nevertheless, other investigations could not find an effect of modality (Bobko, Thompson, & Schiffman, 1977; D. R. Brown & Hitchcock, 1965). Since they aregenerally assumed to be perceived and adjusted more accuralety, the current thesis exclusively employed auditory signals as stimuli.

Time Order Error

In the case of two or more stimuli being presented in succession, a time-order-error (TOE) may occur, i.e., the order of the stimulus presentation has an influence on the participants' judgment or decision behavior (e.g., Allan, 1977; Hellström, 1985). A duration discrimination task usually uses a shorter duration (S_0) and a longer duration (S_1) to be told apart. They can be presented in the order S_1S_0, i.e., the longer duration is presented before the shorter duration, or in the order S_0S_1, i.e., the shorter duration is followed by the longer duration. The participant is required to indicate the order of presentation: R_{10} means that S_1 is presented first followed by S_0, whereas R_{01} means that S_0 is followed by S_1. The signed difference between the two conditional probabilities for the two types of correct answers, $P(R_{10}|S_1S_0) - P(R_{01}|S_0S_1)$ is called TOE (Allan, 1979). The TOE even occurs when the successively presented stimuli are rated in a sensory dimension other than duration, i.e., brightness, in fact that is where the TOE was discovered. However, TOE in duration discrimination is much more crucial than in brightness discrimination: In contrast to brightness discrimination offering the possibility of presenting both stimuli together, duration discrimination cannot avoid the TOE because duration intervals cannot be presented simultaneously. The TOE depends on the duration of the intervals that are being judged: For short durations, the TOE is positive, i.e., it is more likely to correctly identify a long duration to be followed by a short duration than vice versa, whereas for long duration, it is negative. The duration range in which no TOE occurs is called "indifference interval" (Hellström, 1985). Furthermore, the TOE depends on the stimulus range presented in the experiment (Jamieson & Petrusic, 1975). Even though this thesis did not use discrimination tasks except for Experiment C2, the TOE might, nevertheless, have an influence on the timing behavior. Ratio production procedures also employ two stimuli, i.e., standard and comparison that are presented in succession and have to be judged or adjusted, respectively. Therefore, Experiment A2 employed both orders of stimulus presentation (standard followed by comparison and comparison followed by standard) to check whether adjusted durations significantly differ for the two orders. Because no TOE was found, one might assume that the presented durations were in the neighborhood of the indifference interval.

1.3.2 Timing Theories and Models

There is a number of different quantitative models for the perception of brief temporal intervals attempting to explain the mechanism of encoding and decoding of durations, judgment and discrimination. Most of these models assume one common mechanism that underlies the perception of different types of intervals, e.g., visual and auditory, which is supported by empirical evidence. For example, transfer effects were found between reproduction tasks applying series of both stimulus types (Warm, Stutz, & Vassolo, 1975). Additionally, inter- and intra-model correlations of comparable size were observed between category ratings of the two modalities (Loeb, Behar, & Warm, 1966). However, there is no theory that can explain all empirically observed phenomena, because most of them are adapted to either Stevens' law or Weber's law, to certain experimental methods or to further task demands such as non-temporal processing of, e.g., affective stimulus material or distracting cognitive tasks. The most established models will be introduced in the following.

Creelman Model

One of the first models primarily focused on the psychophysical law was stated by Creelman (1962) assuming that the judgment of duration is based on a number of pulses that are accumulated during a certain interval. The pulses are emitted by independent generators with a fixed firing rate λ. During an interval t, the expected number of emitted pulses and its variance is λt. For short intervals t, the distribution of λt is Poisson, whereas for long intervals, it is approximately normal. However, λ was found not to be constant and thus, the Creelman model was extended, e.g., by adding more parameters (Divenyi & Danner, 1977; Getty, 1975; Kinchla, 1972).

Internal Clock Model

A comparable model was proposed by Treisman (1963), which is referred to as the internal clock model. It assumes a pacemaker that produces regular sequences of pulses at a constant rate. This pulse rate can be affected by the person's emotional state, e.g., the arousal. The pulses are counted by a counter unit and stored in a store unit, if necessary. A comparator unit compares the counted or stored pulses and judges, e.g., which of two intervals is longer. Mistakes in duration judgments can be caused by the counter skipping pulses, by forgetting or false remembering in the store or by comparison errors made by the comparator.

Scalar Expectancy Theory

In this context, the theory of scalar timing has to be mentioned (Gibbon, 1977; Wearden & Lejeune, 2008). According to scalar expectancy theory (SET), the timing behavior is required to exhibit two properties: *Mean accuracy* requires the means of the adjusted durations to vary linearly with increasing standards, whereas the *scalar property of variance* necessitates timing sensitivity to remain a proportion of the increasing standard (Wearden & Lejeune, 2008). The general model behind this theory is related to the internal clock model (Treisman, 1963) and assumes a timing mechanism containing a pulse-emitting pacemaker, a counter, an accumulator and, in some models, a memory component. If the scalar property holds for duration perception, the *coefficient of variation* (CV = SD/M) stays constant with varying standard duration, because the number of pulses accumulated during an interval is constant over intervals.

Onset-Offset Model

The onset-offset-model by Allan, Kristofferson and Wiens (1971) states that the variability in perceived durations – all generated by a given stimulus t – is caused by perceived variations in the on- and offsets of t. The perceptual on- and offset latencies are uniformly and independently distributed over a range of q ms and assumed to be independent of stimulus duration. Thus, the perceived durations are distributed in a range of $2q$ ms with an expected value proportional to the stimulus duration t and a variance of $q^2/6$ due to the distribution.

Real-Time Criterion Model

Another model is the so-called real-time criterion model by Kristofferson (1977), who argues that duration discrimination can be interpreted as a temporal order discrimination. The onset of an interval t triggers an internally timed interval which is called the criterion. If the internal criterion ends before the offset of t, the stimulus is categorized as short. Otherwise, if the stimulus t ends before the internal criterion ends, the interval is classified as long. According to this model, errors in timing behavior are entirely due to the variability in the criterion, which depends on the duration values and the range used in the experimental task.

Other Models and Conclusion

Several other models have also been proposed, e.g., by Eisler (1975), who assumes parallel processing of intervals and a comparative judgment, i.e., the categorization of the interval presented in a trial n always depends on the duration presented on trial $n-1$. Thomas and Brown (1974) also propose a parallel temporal processing: They assume each stimulus duration to be encoded as a vector and to be decoded as the inverse vector. Thomas and Weaver (1975) integrate the role of non-temporal information and state that if attention is divided between the temporal encoding and the encoding of non-temporal information, the perceived duration is composed of the timing process and the duration of the information encoding process depending on the complexity of the information.

As mentioned before, none of these models is appropriate to explain all empirically observed phenomena associated with timing behavior. Nonetheless, each of them provides an explanation for a certain aspect of temporal processing and thus noteworthy to mention. Although this doctoral thesis does not concentrate on the development and evaluation of timing models, the empirical findings of the present experiments can be related to some of the introduced timing models. The participants' duration adjustments, e.g., can be analyzed with regard to scalar expectancy theory (Gibbon, 1977; Wearden & Lejeune, 2008) and thus, related to the internal clock model (Treisman, 1963).

1.3.3 Psychophysics of Duration Perception

The first psychophysical investigation of duration perception used, as many of the following studies, the method of limits (Holway & Pratt, 1936) to determine a difference lime (DL). Today, typically indirect and often adaptive psychophysical methods are applied to estimate difference thresholds and just noticeable differences (JNDs). In these procedures, on a given trial, participants are to decide, e.g., which one of two successively presented intervals is perceived to last longer. According to their responses, a psychometric function or a specific point estimate on it can be determined by plotting the response probability of "First interval is longer!" as a function of the interval duration. In addition, direct psychophysical methods such as magnitude production or magnitude estimation can be used to estimate the parameters of Stevens' power law and thus the form of the psychophysical function for duration.

JNDs und Weber Fractions: Hypotheses and Empirical Findings

The investigation of difference thresholds and just noticeable differences (JNDs) led to the examination of Weber's law for duration perception (Estel, 1885; Mehner, 1885) stating the ratio of the JND to the standard to be constant across different durations ($\frac{\Delta I}{I} = c$). However, a number of doubts concerning the validity of Weber's law arises due to different findings on timing behavior. "Breaks" in the psychophysical function (H. Eisler, 1975; Richards, 1964) were found for increasing standard durations, which might be explained by the timing process to switch to new measuring ranges or processing stages. Furthermore, different neural mechanisms for the perception of duration of intervals up to 500 ms and beyond are assumed (Rammsayer, 1994). The separation of two stimuli by an inter-stimulus interval, i.e., the "gap" between the presentation of a standard and a comparison interval, can interfere with the perceptual process and thus influence discrimination performance. Other methodical effects that might affect the validity of Weber's law are the time order error (TOE) and differential stimulus weighting, i.e., the order of standard and comparison stimulus: For a negative TOE meaning an overestimation of the shorter interval being followed by the longer interval as well as for a standard-first-comparison-second order (Stott, 1935), smaller JNDs were found.

Nevertheless, different empirical studies can be consulted to answer the question whether Weber's law holds for duration perception: Estel (1885) found constant Weber fractions in a stimulus range of 1.5 to 5 s, whereas Henry (1948) investigated shorter durations and found Weber fractions to be constant between 0.4 and 4.0 s with a minimum at 1 s. Grondin, Ouellet and Roussell

(2001) found Weber's law to hold between 0.6 and 0.9 s and increasing Weber fractions between 0.9 and 1.2 s.

In general, one might argue that the simple form of Weber's law does not hold for duration perception. However, generalizations of Weber's Law, i.e., modified forms of Weber's theory are able to explain increasing or decreasing Weber ratios for certain duration ranges. For example, the generalized form postulated by Getty (1975) explains very high Weber fractions for very short standards (< 200 ms) by assuming a constant sensory noise interfering with the process of duration perception and predicts higher Weber fractions for short standard durations: The noise as a duration-independent source of timing variability especially influences short standard durations whereas longer durations remain unaffected (Getty, 1975; Rammsayer, 2010b). Other generalized forms were reported by Guilford (1932), Creelman (1962) and Killeen and Weiß (1987). The determination of JNDs and Weber fractions is subject of Experiment 2 in Manuscript C.

The Psychophysical Function for Subjective Duration

Measuring subjective duration requires the determination of the psychophysical function of internal duration, i.e., it has to be examined, how psychological duration increases with physical duration. It is noteworthy that in contrast to most other sensory modes like audition and vision, time perception is not based on the input of specialized "duration receptors", such as dedicated neural wiring or specific brain regions (Grondin, 2010). There is just an equivocal concept of a "time sense" describing the perception of duration and thus, one might argue that time perception is not comparable to other sensations. However, regular psychophysical scaling methods can be used to estimate a psychophysical function for internal time. As well as for other sensory continua, it was found that the psychophysical function of perceived duration is not linear, even though if it comes close.

A finding constantly observed in many studies is an overestimation of short intervals (< 600 ms) and an underestimation of long intervals (> 600 ms). This finding can be accounted by Vierordt's law (Vierordt, 1868) and is usually described by the power function parameters.

Typical $\beta-$exponents reported in ratio production and magnitude and verbal estimations vary between 0.8 to 0.9 (H. Eisler, 1975, 1976), whereas other experiments revealed exponents > 1 (Björkman & Holmkvist, 1960; Frankenhaeuser, 1960).

Several sources that influence the size of the power function exponent β were reported, e.g., the intensity (H. Eisler & A. D. Eisler, 1992) and modality of the stimuli presented (visual or auditory), the participants' age (H. Eisler, 1976), gender (H. Eisler & A. D. Eisler, 1992), geographic origin (A. D. Eisler, 1995), neurological impairment (A. D. Eisler, H. Eisler, & Mori, 2001) as well as the experimental method (Bobko et al., 1977; Painton, Cullinan, & Mencke, 1977). Especially the experimental method seems to be a crucial source of variance, even though an attempt was made to minimize methodological biases. Eisler (1975) reported a low correlation ($r = .14$) between exponents estimated from reproduction and ratio production data. Bobko, Thompson, and Schiffman (1977), Allan (1979), as well as Fraisse (1984) take this finding as evidence for a linear relationship between physical and internal time. They argue that power function exponents clearly differing from 1 only result from ratio production experiments being an inappropriate method to determine the psychophysical function. In some data analyses (H. Eisler, 1975; Michon, 1967), a breakpoint was found in log-log plotted coordinates of the psychophysical function that supposedly divided the function into two or even more segments. However, there is no satisfying theoretical reason for the existence such a breakpoint and therefore, these approaches were not further investigated.

However, there are several findings that do *not* support the existence of a power function relationship between physical and perceived duration. Kaner and Allan (unpublished data; Allan, 1979) compared linear and power function models for each participant and found the linear function to fit better in 23 of 32 cases, even though the power function model seemed to better fit the averaged data with an exponent of $\beta = .77$. Kristofferson (1980) even claimed that there was no transformation of physical time to perceived time and thus argued that internal time is identical to real time.

In conclusion, time perception is not assumed to be veridical. Even if interpreting exponents "close to" 1 as a veridical transformation would be easy, this doctoral thesis assumes a power relationship between internal time and real time, because empirical evidence supports this more complex modeling.

Nonetheless, the estimation of the psychophysical function for duration perception is not as straightforward as is for other sensory modalities, because the exponent might be influenced by a number of context effects. For that very reason, it is necessary to investigate the psychophysical function for duration perception in more detail and to validate assumptions fundamental to ratio scaling of perceived duration.

1.4 Research Questions

This doctoral thesis consists of three manuscripts – one of them published, one accepted for publication and one to be revised and resubmitted – each reporting two or three experiments. In this section, the research questions of each manuscript resulting from the theoretical background and the previous empirical findings are briefly derived. The most important issues of each manuscript are explained as leading to five central research questions that are addressed in this thesis. Furthermore, an overview on the three manuscripts is provided in Table 1.1, whereas their relations are depicted in Figure 1.1.

Previous studies revealed inconsistent findings regarding the relationship between physical and perceived duration. These ambiguous findings often result from different scaling procedures, but also from different standard durations employed in the experiments. This doctoral thesis particularly investigated the applicability of direct scaling methods, i.e., ratio scaling of duration perception. Therefore, the fundamental assumptions of direct scaling, which were implicitly expressed by Stevens (1957), are validated. These assumptions that have to be tested separately for each sensory modality, require the stimulus intensities to be perceived on a sensory continuum and at ratio scale level. This thesis applies rigorous testing of methods developed in representational measurement theory to the scalability of perceived durations. Furthermore, it has to be assumed that the experiment's participants interpret the numbers presented in the instructions as mathematical values. Narens' axioms (Narens, 1996) express these assumptions as mathematical conditions that can be tested empirically using data collected in a ratio production experiment. These axioms are monotonicity, commutativity, and multiplicativity (see section 1.2) and they

Table 1.1

Overview on the manuscripts (MS), the short titles, the journal of publication, and the size of the sample (N).

MS	Short Title	Journal	N
A	Octuplicate this interval!	*Attention, Perception & Psychophysics* 2015: *77*(5), 1767-1780	25
B	Investigating numerical distortion	*Journal of Mathematical Psychology* (accepted for publication)	35
C	Quantifying subjective duration	*Attention, Perception & Psychophysics* (revise & resubmit)	10

are evaluated in this doctoral thesis for the perception of short durations. If the axioms are found to be valid, the power function parameters for duration perception can be estimated to determine the shape of the psychophysical function. Because the estimated power law parameters are often employed to compare the sensitivity of different modalities, they have to be invariant under changes of the standard stimulus used in the experiment. That is why it has to be tested whether the parameters estimated on the basis of a ratio production procedure are independent of the employed standard durations. Otherwise, the parameters are not psychologically relevant (Augustin, 2008; Narens & Mausfeld, 1992). Both questions were addressed in two separate experiments which are reported in the Manuscript A.

As hitherto investigated, the axioms of monotonicity and commutativity hold for most sensory modalities, whereas the axiom of multiplicativity was usually found to be violated. As reported in Manuscript A, a violation of multiplicativity was found for most participants for the perception of duration, as well. Thus, it can be assumed that participants have difficulties to interpret the numbers presented in the experimental instructions as mathematical values, i.e., their number representation is based on some ratio scale, but, however, not veridical. Based on this finding, another functional relationship between mathematical and perceived numbers has to be assumed (Steingrimsson & Luce, 2007), i.e., a power function with a constant exponent multiplied by a constant factor k. This assumption is weaker than the axiom of multiplicativity and can be empirically evaluated by validating the axiom of k-multiplicativity.

Ratio production experiments usually employ both fractions and integers to instruct participants. The axiom of k-multiplicativity was tested separately for $\mathbf{p} < 1$ and $\mathbf{p} \geq 1$ (for notation, see section 1.2), because empirical findings suggest fractions and integers to be processed differently. According to the latter assumption, testing k-multiplicativity in a mixed condition should result in axiom violations. Thus, to confirm the divergent processing of fractions and integers, k-multiplicativity was evaluated in an additional intermixed condition with both $\mathbf{p} < 1$ and $\mathbf{p} \geq 1$. The findings of three separate data analyses are reported in Manuscript B.

Presumably because the power law parameters were supposed and sometimes found to be invariant under changes of the standard stimulus, the relationship between the power function exponent and the size of the standard stimulus was not examined for duration perception so far. However, as reported in Manuscript A, the standard duration does have an influence on the size of the exponent and thus has to be interpreted carefully. Beyond the axiomatic approach showing a standard dependency to exist, it appears worthwile to determine the exact relationship between the power law exponent as a function of the standard. To preclude that the standard dependency arises due to specific artifacts due to the method of ratio production, differential sensitivity for perceived durations was determined by yet another measure, i.e., by just noticeable differences (JNDs), in order to check whether a comparable relationship exists between standards and Weber fractions. Both of these experiments were conducted testing the same participants. The results are reported in Manuscript C.

In conclusion, five central research questions are answered in the current doctoral thesis:

1. Are the basic assumptions for the application of direct scaling methods – sensory continuum, ratio scalability and veridical number processing – valid for the perception of duration? (Manuscript A, Experiment 1)

2. Are the estimated power law parameters invariant under changes of the reference stimulus and thus psychologically relevant? (Manuscript A, Experiment 2)

3. Can the relationship between mathematical and perceived numbers be

described by a power function with a constant exponent and is there a difference between the processing of integers and fractions? (Manuscript B)

4. Can the functional relationship between standard duration and power law parameters be determined? (Manuscript C, Experiment 1)

5. Does the standard dependency of the power law parameters occur due to the ratio production procedure or can this finding be transferred to other measures of sensitivity (JNDs)? (Manuscript C, Experiment 2)

Ambiguous findings on the psychophysical function for the perception of short duration led to research question 1, i.e., whether the fundamental assumptions for direct scaling of duration are valid at all and thus, whether ratio scaling is an appropriate method to estimate the parameters of the psychophysical function. Since these assumptions were found to predominantly hold, research question 2 arises concerning the psychological relevance of the estimated power law parameters: It has to be evaluated whether the exponent of the power law is invariant under changes of the standard. Research question 3 results from the observation that participants' number perception does not seem to be veridical and thus, another transformation function between mathematical and perceived number has to be determined separately for fractions and integers. Since the power law parameters for duration were found to vary with the standard, it is important to know the exact relationship between the exponent as a function of the standard, as formulated in research question 4. Furthermore, for interpreting power law parameters as a measure of sensitivity, it is a sensible question 5 whether other, unequivocal measures of sensitivity, such as JNDs, vary with the standard as well, or whether the standard dependency arises due to the idiosyncrasies of the method of ratio production.

Research questions 1 and 2 are addressed in the two experiments reported in Manuscript A, whereas question 3 is discussed on the basis of three experiments in Manuscript B. Research questions 4 and 5 are examined in the two studies reported in Manuscript C (see Table 1.1 and Figure 1.1).

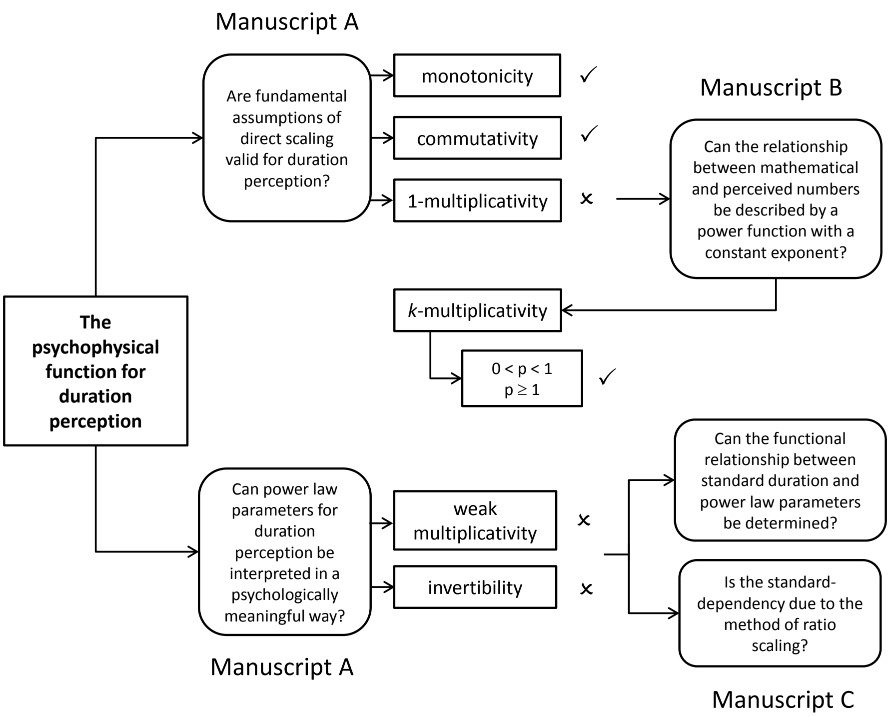

Figure 1.1: Relationship between the research questions, the evaluated axioms and the manuscripts. Outcomes of axiom testing are indicated as ✓ (found to be valid) or ✗ (found to be violated).

Chapter 2

Overview of the thesis

2.1 Manuscript A: Octuplicate This Interval! Axiomatic Examination of the Ratio Properties of Duration Perception

2.1.1 Purpose of the Study

Estimating the parameters of a psychophysical function such as Stevens' power law (1957) is usually based on data collected in direct scaling experiments. Their applicability depends on the validity of the basic assumption that participants are capable of processing ratios of magnitudes. Testing Narens' (1996) axioms of monotonicity, commutativity, and multiplicativity poses an empirical method to evaluate whether the assumptions fundamental to direct scaling hold for a certain sensory modality. Since previous studies investigating loudness (Ellermeier & Faulhammer, 2000; Zimmer, 2005), brightness (Steingrimsson, Luce, & Narens, 2012), area (Augustin & Maier, 2008) and pitch (Kattner & Ellermeier, 2014) yielded comparable violation rates for monotonicity (0% violations), commutativity (11 - 22% violations) and multiplicativity (61 - 94% violations), the purpose of Manuscript A, Experiment A1 was to determine the validity of Narens' axioms for the perception of duration. A further basic assumption necessary for an unambiguous and therefore psychologically relevant interpretability of the power law parameters (Augustin, 2008; Narens & Mausfeld, 1992) is the invariance of the exponent under changes of the standard.

Thus, two axioms formulated by Augustin (2008), weak multiplicativity and invertibility, were tested in Experiment A2 to examine whether the size of the standard duration used in the ratio production experiment influences the size of the estimated power function exponent.

2.1.2 Method

In Manuscript A, two experiments are described both employing a ratio production procedure. According to the experimental instructions, the participants were required to adjust the duration of a comparison interval in a certain ratio to a standard interval. The intervals were marked by continuous A4 sine tones presented with a sound pressure level of 65 dB SPL. The standard intervals used in Experiment A1 were of 100 and 400 ms duration, whereas in Experiment A2, a standard duration of 600 ms was chosen. The initial comparison intervals varied between one and ten times the standard in Experiment A1 and between one third and three times the standard in Experiment A2. These durations were chosen in relation to the particular ratio production factors: In Experiment A1, the basic trials employed integer ratios of $\mathbf{p} = 1, 2, 3, 4, 6$, and 8. The successive trials, i.e. trials that used a comparison interval of a previous $\times 2$, $\times 3$, or $\times 4$ as a standard, ratios of $\mathbf{q} = 2, 3$, and 4 were used, resulting in the pairs $(p, q) = (2, 2), (2, 3), (2, 4), (3, 2)$, and $(4, 2)$. In Experiment A2, basic $\mathbf{p}$-ratios of $\frac{1}{3}, \frac{1}{2}, 1, 2$, and 3 were employed and combined with $\mathbf{q} = \frac{1}{3}, \frac{1}{2}, 2$, and 3 thus constituting the pairs $(p, q) = (\frac{1}{3}, 3), (\frac{1}{2}, 2), (2, \frac{1}{2})$, and $(3, \frac{1}{3})$. During one trial, the participants stepwisely increased or decreased the duration of the comparison interval by pressing a corresponding cursor key. After each keystroke, the fixed standard and the varied comparison interval were replayed, until the participant was satisfied with his or her adjustment and registered the final durations. Experiment A1 employed $N = 10$ participants, $N = 15$ participants took part in Experiment A2.

2.1.3 Results and Discussion

The axiom of monotonicity is satisfied, if the mean individual duration adjustments resulting from two adjacent ratio production factors significantly differ, i.e., if the adjusted durations monotonically increase with increasing ratio production factor. Analyses across the entire sample as well as individual graphical analyses (Augustin & Maier, 2008) revealed the axiom of monotonicity to hold for the entire sample of $N = 10$ participants, i.e., each participant was capable of producing monotonically increasing durations with increasing ratio production factor. The axiom of commutativity is taken to be satisfied, if the successive adjustment $\times\mathbf{p} \times \mathbf{q}$ is statistically indistinguishable[1] from a inverted $\times\mathbf{q} \times \mathbf{p}$ adjustment, i.e., if the order of the ratio production factors does not affect the adjusted outcome duration. Descriptive analyses as well as individual inferential testing revealed the axiom of commutativity to be violated in 12% of the tests. The axiom of multiplicativity holds, if the duration resulting from the successive $\times\mathbf{p} \times \mathbf{q}$ (or $\times\mathbf{q} \times \mathbf{p}$) adjustments is statistically not different from basic $\times\mathbf{r}$ adjustments with $r = pq$. Descriptive analyses as well as individual inferential testing revealed multiplicativity not to be valid in 32% of the tests. In Experiment A2, the axiom of monotonicity was found to be violated by 4 of 15 participants. The axiom of weak multiplicativity is taken to be satisfied, if the adjusted duration resulting from two successive adjustments $\times\mathbf{p} \times \frac{1}{\mathbf{p}}$ is indistinguishable from the basic adjustment of $\times 1$, whereas for the axiom of invertibility, this outcome has to be indifferent from the standard duration. Both axioms were found to be violated by almost all participants (55% and 57% violations). These results of Experiment A1 imply that most participants' perception of short durations, comparable to other sensory modalities, is based on a sensory continuum and on a ratio scale. However, one third of the participants have difficulties in interpreting the ratio production factors as presented in the experimental instruction as mathematical numbers, i.e., a "numerical distortion" impedes an unequivocal interpretation of the scale values. Furthermore, the axiomatic analyses in Experiment A2 revealed that the size of the power law exponent for duration perception depends on the size of the standard duration. This finding implies that even though a power function model fits the relationship between physical and perceived duration quite well, a psychologically relevant interpretation of the power law parameters is impaired.

[1]This is the case, if the difference between the two successive adjustments is smaller than ca. $t * p * 0.1$ ms.

2.2 Manuscript B: Axiomatic Evaluation of *k*-Multiplicativity in Ratio Scaling: Investigating Numerical Distortion

2.2.1 Purpose of the Study

The most crucial axiom of Narens' (1996) framework is the axiom of multiplicativity: Its validity provides evidence for participants taking the numbers as presented in the experimental instruction at face value, i.e., for a veridical transformation between mathematical and perceived number. Most previous studies examining the validity of multiplicativity for different modalities, as well as Experiment A1 came to the conclusion that a majority of participants does not have a veridical understanding of numbers. One recent interpretation concerning multiplicativity argued that a veridical interpretation of numbers is not mandatory for direct scaling, because the validity of commutativity implies ratio scalability and thus the participants to interpret the numbers at some ratio, though not the exact ratio stated in the instructions (Luce et al., 2010). However, another axiomatic approach (Steingrimsson & Luce, 2007) assumed a weighing function between mathematical and perceived numbers of $W(\mathbf{p}) \neq p$ instead of a veridical one ($W(\mathbf{p}) = p$). According to this assumption, the relationship between mathematical and perceived numbers can be described by a power function with a constant exponent multiplied by a constant factor k. To empirically evaluate whether this assumption holds, the axiom of k-multiplicativity can be tested. It is comparable to the axiom of multiplicativity (or 1-multiplicativity, $r = 1pq$), but weaker, because the more general form $r = pqk$ has to be satisfied. Because empirical evidence is given for the assumption that participants process and represent fractions and integers in a different way (Bonato, Fabbri, Umiltà, & Zorzi, 2007; Ganor-Stern, 2012; Steingrimsson & Luce, 2007), it is necessary to examine k-multiplicativity for fraction and integer ratio production factors in separate experiments. Intermixing them in a further experiment should result in a violation of k-multiplicativity and thus providing evidence for different processing. Therefore, the aim of Manuscript B was to check the validity of k-multiplicativity for duration perception. For integer ratio production factors ($p \geq 1$), the results of Experiment A1 were reanalyzed. Experiment B1 was conducted using fractions as ratio production factors ($p < 1$) to separately check k-multiplicativity. For both experiments, the k-multiplicativity was assumed to hold. The data of Experiment A2 using both fractions and integers as ratio production factors were reanalyzed to test the axiom of k-multiplicativity in a mixed condition. In this case, an axiom violation was assumed to occur due to the participants' different processing of fractions and integers.

2.2.2 Method

The method used for Experiment B1 was similar to the procedure employed in Experiment A1 and A2: A ratio production procedure was conducted using a standard duration of 1600 ms and an initial comparison interval which was randomly chosen between one eighth and one time the standard. The ratio production factors of the basic trials were $\mathbf{p} = \frac{1}{8}, \frac{1}{6}, \frac{1}{4}, \frac{1}{3}$, and $\frac{1}{2}$. In the successive trials, they were combined with the ratios of $\mathbf{q} = \frac{1}{2}, \frac{1}{3}$, and $\frac{1}{4}$, resulting in the pairs $(p, q) = (\frac{1}{2}, \frac{1}{2}), (\frac{1}{2}, \frac{1}{3}), (\frac{1}{2}, \frac{1}{4}), (\frac{1}{3}, \frac{1}{2})$, and $(\frac{1}{4}, \frac{1}{2})$. The participants' task was to adjust the duration of the comparison tone to a certain ratio of a standard. Stepwise adjustments could be made by pressing the appropriate cursor keys. Experiment B1 employed $N = 10$ participants. The analysis of k-multiplicativity is arranged in three paragraphs: In the first paragraph, the data resulting from Experiment A1 using integers as ratio production factors are analyzed. The second paragraph shows the results of the new experiment B1 only using fractions. In the third paragraph, the data from Experiment A2, in which both fractions and integers were used as ratio production factors, are analyzed.

2.2.3 Results and Discussion

The axiom of k-multiplicativity holds if the adjusted duration of a successive trial $\times\mathbf{p} \times \mathbf{q}$ (or $\times\mathbf{q} \times \mathbf{p}$, respectively) multiplied by a constant factor k is statistically indistinguishable from a single adjustment $\times r$ (with $r = pq$) and if k is indistinguishable across several associated $\times\mathbf{p} \times \mathbf{q}$ and $\times\mathbf{r}$ adjustments. The axiomatic analyses revealed k-multiplicativity to hold for all $n = 20$ participants in Experiment A1 and B1 and thus, separately for both integers and fractions. However, different values of k were found for fractions ($k > 1$ and thus, $W(\mathbf{p}) > p$) and integers ($k < 1$ and thus $W(\mathbf{p}) < p$) indicating a differing way of processing for $p < 1$ and $p \geq 1$. The validity of k-multiplicativity confirms the assumption that number representation in participants is not veridical but follows a power relationship with a constant exponent. Intermixing fractions and integers in Experiment A2 revealed a violation of k-multiplicativity for 13 of $n = 15$ participants, i.e., the assumption of fractions and integers to be processed in different ways – also indicated by different values of k extracted from Experiment A1 and B1 – could be confirmed. This finding explains a bias observed in most ratio production experiments: When integers are used as ratio production factors, the adjusted magnitudes of successive trials often exceed the adjustment of single trials (Augustin & Maier, 2008; Birkenbusch, Ellermeier, & Kattner, 2015; Ellermeier & Faulhammer, 2000) whereas when using fractions, by contrast, the outcomes of successive trials fall below the outcomes of single trials (Steingrimsson & Luce, 2007; Zimmer, 2005). Since, however, knowing the size of k and the approximate form of the weighting function $W(\mathbf{p})$ does not reveal the exact functional relationship between perceived and mathematical numbers, correcting the power law exponent by somehow "undoing" the numerical distortion – as proposed by Schneider et al. (1974) – still seems impossible at this point.

2.3 Manuscript C: Quantifying Subjective Duration: Both Power Function Exponents and Weber Fractions Vary With the Standard

2.3.1 Purpose of the Study

In Experiment A1, the estimation of the parameters α and β of the psychophysical function revealed significantly differing exponents for the two different standard durations t. This finding was confirmed by the axiomatic evaluation of weak multiplicativity and invertibility in Experiment A2: It was found that the size of the exponent β is not invariant under changes of the standard duration and thus, a psychologically relevant interpretation of the shape of the psychophysical function for duration perception is impaired. However, the results of Manuscript A led to the assumption that the size of the exponent increases with increasing standard, but this assumption was neither experimentally investigated nor statistically tested. Even though previous research (H. Eisler, 1976; Kane & Lown, 1986) attempted to clarify the dependency of the exponent on the standard duration, there is neither a clear proposition nor an exact functional form describing the relationship between standard and exponent. Therefore, the purpose of Experiment C1 of Manuscript C was to determine the exact relationship of the exponent as a function of the standard duration. Since the parameters of the psychophysical function are used to compare the sensitivity of different sensory modalities, it seems plausible to investigate whether the exponent-standard dependency is evidence for a change in sensitivity of the perception of short durations or whether it is just a result of the ratio production procedure. That is why in Experiment C2, an adaptive procedure measuring discrimination "thresholds" (Kaernbach, 1991) was used to determine the Weber fractions as another measure of sensitivity at the same standard durations as used in Experiment C1 and for the same participants.

2.3.2 Method

The method of Experiment C1 is comparable to the procedures of Experiment A1. The ratio production procedure employed six different standard durations ranging from 100 to 600 ms and initial comparison durations ranging between one and six times the corresponding standard. As in the previous experiments, sine tones at a frequency of 440 Hz (A4 standard pitch) were used. The ratio production factors were $\mathbf{p} = 2, 3$, and 6. There were no successive trials as in Experiment A1. Again, participants were asked to adjust the duration of the comparison interval to the ratios $\mathbf{p}$ to the standard. Adjustments were made by keypress. In Experiment C2, the participants had to complete an adaptive weighted up-down procedure to determine the 75%-correct threshold of the same standard stimuli as used in Experiment C1. In each trial, a standard and a comparison tone were presented in random order and the participants had to indicate, which of the tones was perceived to be longer. Due to the weighted up-down procedure, the response on the current trial determined the difficulty of the next trial: If the response was correct, the distance between the standard and the comparison tone in the next trial was decreased, whereas it was increased, if the answer was incorrect. Each of the six standard durations was combined with ascending and descending comparison tones, i.e., in half of the trials, the standard was shorter than the comparison, whereas it was longer in the other half of the trials. The same $N = 10$ participants took part in both experiments.

2.3.3 Results and Discussion

The results of Experiment C1 revealed increasing exponents β with increasing standard durations t ranging from 100 to 600 ms. A positive power relationship of the form $\beta = 0.13t^{0.3}$ was found between standard and exponent. Between 100 and 400 ms, the slope of the function is steeper, whereas it decreases between 400 and 600 ms. Individual analyses of the relationship between standard duration and estimated exponent revealed a comparable relationship for seven of ten participants. The present finding confirms previous axiomatic tests of weak multiplicativity and invertibility (Manuscript A) stating that the size of the exponent varies under changes of the standard. Furthermore, one might assume that differential sensitivity of duration perception increases with increasing standard, since a higher sensitivity is associated with greater exponents β (R. Teghtsoonian, 2012; Ward, Armstrong, & Golestani, 1996). Experiment C2 revealed a negative power relationship between the standard duration and the Weber fractions of the form $W = 0.84t^{-0.3}$, i.e., an increase in the sensitivity for short durations can be assumed. Especially between 100 and 400 ms the function has a steep slope, whereas it decreases between 400 and 600 ms. This observation is in line with former findings (Getty, 1975; Grondin, 2001, 2010) also reporting high Weber fractions below 400 ms and constant but lower values beyond. According to this result, an effect due to the method of ratio production can be precluded. Furthermore, the results contribute to the debate on the relationship between power function exponents and Weber fractions. A negative correlation of $r = -.94$ ($p = .004$) was found between exponent and Weber fraction supporting previous findings that both measures are negatively correlated (R. Teghtsoonian, 1971).

Chapter 3

General Discussion and Conclusions

The present doctoral thesis investigated central aspects relevant for the psychophysical scaling of duration perception and thus, closely examined conditions necessary for the application of direct methods to determine the relationship between physical and internal duration. For this purpose, five experiments – four ratio production and one adaptive procedures – were carried out. Altogether, 55 participants took part in the experiments and completed 16,425 trials.

Note that the axiomatic conditions investigated stem from different theoretical approaches (Augustin, 2008; Narens, 1996; Steingrimsson & Luce, 2007) and have never been evaluated in a single empirical study, thereby affording more stringent conclusions about the validity of the psychophysical law than earlier work. Furthermore, this thesis constitutes the first application of the axiomatic measurement approach to studying duration perception.

In this section, a brief résumé of the results is provided followed by a paragraph on their theoretical significance and their practical implications. Furthermore, limitations and future perspectives of the work reported in this thesis are summarized.

3.1 Summary of Results

Manuscript A aimed to find out whether the basic assumptions for the application of direct scaling methods – sensory continuum, ratio scalability, and veridical number processing – are valid for the perception of short duration (Research Question 1; see section 1.4). Therefore, Narens' (1996) axioms of monotonicity, commutativity, and multiplicativity were tested on the basis of the data set resulting from Experiment A1. The results showed perceived duration to occur on a sensory continuum, because monotonicity was found to hold for all $N = 10$ participants. The axiom of commutativity was valid in 88% of all tests (12% violations) and thus, it can be assumed that the majority of participants processes duration on a ratio scale. Therefore, the two fundamental conditions for the application of direct scaling are valid, indicating that direct scaling and especially ratio production are appropriate methods to determine the shape of the psychophysical function, i.e., the parameters of Stevens' power law. However, the evaluation of multiplicativity revealed axiom violations in 32% of all tests, indicating that many participants do not process the numerical instructions as presented in the experiment as scientific numbers. This finding implies that there is no veridical transformation between perceived and mathematical numbers and thus, the scale values as produced by the participants in a ratio production experiment cannot be interpreted at face value. Manuscript B investigated the issue of the transformation from mathematical to internal number representation in more detail.

Manuscript A further examined whether the estimated power law parameters are invariant under changes of the reference stimulus and thus psychologically relevant (Research Question 2). Therefore, Augustin's (2008) axioms of weak multiplicativity and invertibility were evaluated on the basis of Experiment A2. The analyses found both crucial axioms to be violated by more than 50% of the $N = 15$ participants and thus, the assumption of invariant power law parameters has to be rejected. These findings were supported by an observation resulting from Experiment A1 yielding significantly different power function parameters for the two different standard durations employed. Manuscript C investigated the issue of standard-dependent parameters in more detail.

The question central to Manuscript B was whether the relationship between mathematical and perceived numbers can be described by a power function with a constant exponent and whether there is a difference between the processing of integers and fractions (Research Question 3). Therefore, Steingrimsson's and Luce's (2007) axiom of k-multiplicativity, which constitutes a more general assumption about the transformation function than 1-multiplicativity, was tested for three different data sets. The results revealed the axiom of k-multiplicativity to hold for fractions ($0 < p < 1$, Experiment B1, $N = 10$) as well as for integers ($p \geq 1$ Experiment A1, $N = 10$) indicating that the relationship between perceived and scientific magnitude of both fractions and integers can be described by a power function with a constant exponent multiplied by a constant factor k. However, the results further showed these functions not to be identical, i.e., fractions and integers are processed differently, because different values of k were found for fractions and integers. Additionally, k-multiplicativity turned out to be violated in the intermixed condition (Experiment A2, $N = 15$) also indicating that participants have different representations of numbers $0 < p < 1$ and $p \geq 1$.

Manuscript C aimed to find out whether the functional relationship between standard duration and power law parameters can be determined (Research Question 4). Therefore, ratio production data for six different standard durations (100 – 600 ms) produced by $N = 10$ participants were analyzed (Experiment C1). The results yielded increasing β exponents between 0.55 and 0.91 for increasing standard durations. A power function of the form $\beta = 0.13t^{0.3}$ fit the data with an adjusted R^2 of 99%. This finding shows that the exponent as a measure of sensitivity increases with increasing standard between 100 and 400 ms, whereas it stabilizes for standards longer than 400 ms.

It was further examined whether the standard dependency of the power law parameters occurs due to the ratio production procedure of whether this finding can be substantiated by other measures of sensitivity (Research Question 5). Therefore, Experiment C2 employed an adaptive weighted up-down procedure with the same sample of $N = 10$ participants as took part in Experiment C1 to determine the JNDs and Weber fractions for the same standards. The analyses revealed decreasing Weber fractions with increasing standard that can be described by a power function of the form $W = -0.84t^{-0.3}$. Since small Weber fractions indicate a higher differential sensitivity, the findings are in line with the results from Experiment C1: The sensitivity for duration perception increases below 400 ms and remains stable above 400 ms.

3.2 Theoretical Conclusions and Practical Implications

The present thesis contributes to the empirical evaluation of the psychophysical method of direct scaling and its application by means of representational measurement theory. Furthermore, it provides theoretical conclusions helpful for the understanding of context effects in direct scaling and on aspects relevant for the issue of duration perception. These theoretical conclusions and their practical implications will be discussed in the following section.

3.2.1 Application of Direct Scaling

One central aim of this thesis was to confirm assumptions fundamental to direct scaling concerning the scalability of perceived duration. For the application of direct scaling methods in order to determine the parameters of the psychophysical function, it is necessary that the investigated sensory modality is perceived on a sensory continuum and on a ratio scale (Narens, 1996; S. S. Stevens, 1975). The axiomatic tests conducted in this thesis revealed both assumptions to hold for internal duration and thus, the application of direct scaling and ratio scaling in particular is justified. Because direct scaling is a commonly employed method to determine the form of the psychophysical function of duration, this thesis reveals important practical implications by legitimating this procedure for the past and future ratio scaling experiments.

Ratio scaling experiments typically employ both integers, and – more rarely – fractions as ratio production factors (Luce, 2002), which might be interpreted as a critical issue, since numbers < 1 and > 1 are assumed to be processed differently (Bonato et al., 2007; Ganor-Stern, 2012; Steingrimsson & Luce, 2007). However, the use of fractions and integers revealed comparable violation rates for monotonicity and commutativity indicating that ratio production can be applied in both settings[1].

[1]Furthermore, the violation rates for all axioms were shown to remain stable across different

Nonetheless, investigating the processing of integers and fractions in order to examine the relationship between mathematical and perceived numbers showed that not all participants have a veridical understanding of numbers: The evaluation of the axiom of multiplicativity revealed violations for $0 < p < 1$ as well as for $p \geq 1$. Although Narens (1996) stated a veridical transformation function and thus, the validity of the axiom of multiplicativity to be a mandatory condition for applying direct scaling, Luce (2008, 2010) assumed this condition not to be necessarily required. He argued that the validity of commutativity shows the numbers to be processed as some ratio, though not as the exact ratio stated in the experimental instructions. Therefore, even though multiplicativity is violated in about 30% of all tests, direct scaling of duration can be performed, but interpreting the scale values remains difficult.

Because number processing was found not to be veridical, another approach by Steingrimsson and Luce (2007) was taken in order to check whether the relationship between mathematical and perceived numbers can be described by a power function with a constant exponent multiplied by a constant factor k. Axiomatic analyses of k-multiplicativity showed this assumption to be valid for both fractions and integers. For $0 < p < 1$, $k > 1$ was found, whereas $p \geq 1$ revealed $k < 1$ indicating fractions to be over- and integers to be underestimated relative to their "true" numerical value. Overestimation increases with a decreasing value of the fraction, whereas underestimation increases with increasing value of the integer. This finding replicates previous findings in the context of axiomatic measurement (Steingrimsson & Luce, 2007), but also in the context of decision making and utility theory investigating wider ranges of numbers (e.g., Fishburn, 1988) and is therefore very remarkable.

After determining k, conclusions on the form of the weighting function relating mathematical and perceived numbers can be drawn. For fractions, a weighing function of the form $W(\mathbf{p}) > p$ can be assumed, whereas for integers, $W(\mathbf{p}) < p$ is derived. However, because different values of k were found for

experiments, since they were also tested, but not reported in Manuscript C. Experiments conducted in an Experimental Practicum course taught by the author of this thesis in the winter term 2014/15 revealed similar violation rates for visually marked durations that were either temporally or graphically structured. Therefore, it can be assumed that duration perception occurs on a ratio scale independent from modality or stimulus content, even though this finding has to be more systematically investigated.

fractions and integers, a general weighting function of the form $W(\mathbf{p}) = kp^{\omega}$ cannot be determined. This finding practically implies that ratio scaling experiments should not use fractions and integers as combined ratio production factors, because they are represented differently.

Furthermore, the divergent weighting functions for fractions and integers explain a bias frequently observed in ratio scaling: For integers employed as ratio production factors, successive adjustments, e.g., $\times 2 \times 3$ usually exceed the basic adjustments, e.g., $\times 6$ (e.g., Ellermeier & Faulhammer, 2000). In fractionation (e.g., Zimmer, 2005), by contrast, the successive adjustments ($\times \frac{1}{2} \times \frac{1}{3}$) commonly fall short of the basic adjustments ($\times \frac{1}{6}$).

Additionally, the individual differences concerning axiom violations observed in the present studies lead to the conclusion that the fundamental assumptions of direct scaling should be validated for each participant, before his or her data are included in the general analyses or in an overall estimation of psychophysical parameters. Especially data produced by participants who violate the axiom of commutativity and thus have difficulties to process duration or other sensory modalities on ratio scale level might strain a proper interpretation of power law parameters.

The axiomatic tests of commutativity (about 9% violations for fractions and integers) and multiplicativity (about 32% violations) for perceived duration can further be compared to findings for other sensory continua: For the perception of area, Augustin and Maier (2008) reported violation rates of 12% for commutativity and 61% for multiplicativity. Ellermeier and Faulhammer (2000) found commutativity to be violated in 11% of all cases and violations of multiplicativity in 94% of the tests using integer ratios, whereas Zimmer (2005) reported violation rates of 14% and 89% using fractionation, with both studies examining loudness. For the perception of pitch, Kattner and Ellermeier (2014) found the axioms to be violated in 22% and 33% of all tests, respectively. It seems as if the violation of commutativity is robust across different sensory modes ($9 - 22\%$) indicating that the perception of most sensory modes occurs on a ratio scale. For the axiom of multiplicativity, by contrast, obviously a wider range of violations was found ($32 - 94\%$). Since multiplicativity tests whether the participants interpret the numbers as presented in the instructions (i.e., like mathematical numbers), this finding is somewhat counterintuitive: It would rather be assumed that commutativity varies due to the divergent representation of different modalities whereas multiplicativity reflecting participants' number representation should remain robust. Therefore, it might be further investigated whether the wide range of observed violations occurs due to variable experimental instructions, differences in the samples, or other unconsidered influences.

3.2.2 Context Effects

Furthermore, the role of context effects constituting a persistant nuisance in direct scaling was reassessed. Since it had repeatedly been questioned whether the estimated parameters are invariant under certain context effects (Gescheider, 1997; Poulton, 1989; M. Teghtsoonian & R. Teghtsoonian, 2003), it is plausible to investigate context effects and their systematic behavior in more detail. Although the present thesis investigated the role of the standard size, one might interpret the current findings as a reassessment of context effects, because they seem to have a systematic but not merely a distorting influence in direct scaling.

A practical finding concerning the comparison of power function exponents across different modalities identified in this thesis results from their dependency on the size of the standard duration: Particularly in a range between 100 and 400 ms, the exponent of the power function increases, i.e., the differential sensitivity of duration perception changed in this short standard range. The individual analyses showed the same pattern, i.e., increasing exponents with increasing standards for seven of ten participants. Thus, it can be assumed that some participants' duration perception is influenced by the standard size, whereas for other participants it is not.

Therefore, the results imply, for determining the power function exponent that a standard longer than 500 ms should be employed, or, otherwise, the size of the reference stimulus should be referred to when power law parameters are reported. This might be inevitable, since comparisons between sensory modalities using empirical power functions exponents[2] cannot be conducted

[2]For example, the exponent β can be used to compare the gradients of different sensory modalities to draw conclusions about their sensitivity (H. E. Ross, 1997; Ward et al., 1996), i.e., if the exponent of a sensory modality A is larger than the exponent of a modality B,

without an element of arbitrariness, because there is no plausible way to find a comparable standard size in the other modality.

Further research, however, should investigate whether the influence of the standard size on the exponent interacts with other context effects (Garner, 1954; Mellers & Birnbaum, 1982), e.g., the range of the standard stimuli (Ward et al., 1996) or the numbers assigned to them (Beck & Shaw, 1965), which have been examined in previous studies.

3.2.3 Duration Perception

As regards duration perception, the present findings contribute to the debate on whether the psychophysical relationship between physical and internal duration is a linear (Allan, 1979; Fraisse, 1984) or a power function (H. Eisler, 1975, 1976). The results revealed a power function to fit the relationship between physical and internal duration very well ($R^2 = 0.92$). However, comparing the power function to a simple linear model ($R^2 = 0.85$) reveals both models to fit rather well. Furthermore, the overall estimated power function exponent is close to one ($\beta = 0.99$ across three different standards, Manuscript A). This finding indicates that even though the size of the estimated individual exponents seems to rule out a linear transformation, a simple linear function might fit the group data better than a power function. Nonetheless, especially with regards to different standard durations, the power function model seems to be a more appropriate than the linear model, because it is able to describe the varying sensitivity across different standards (exponents ranging from 0.64 to 0.91; see section 1.3).

Furthermore, the present findings can be related to scalar timing properties (Gibbon, 1977; Wearden & Lejeune, 2008) and therefore, to the pacemaker-counter model of internal timing (Creelman, 1962; Rammsayer & Ulrich, 2001; Treisman, 1963; see section 1.3). The scalar expectancy theory (SET) assumes the two fundamental properties of *mean accuracy* and *scalar property of variance*.

---

it can be assumed that the observer is more sensitive to modality A than to modality B. In relation to indirect scaling methods, it was found that the ratio of two power function exponents β_A and β_B of two different sensory modalities A and B corresponds to the inverse ratio of the Weber fractions of these modalities W_A and W_B (R. Teghtsoonian, 2012; Ward, 1995).

That implies that the mean representation of time for a series of temporal judgments has to be proportional to real time, which is shown in psychophysics when an exponent of $\beta = 1$ is found and thus, the relationship between physical and perceived duration is linear. In the present thesis, an overall exponent of $\beta = 0.99$ was found indicating mean accuracy to hold in general. Especially for standards longer than 400 ms, mean accuracy holds, but, when considering short standard durations, the property would be rather invalid. The second property, the variability of timing, holds if the variability of temporal judgments increases linearly with the mean representation of duration, i.e., if the proportion between variability and the mean remains constant, the scalar property of variance is valid. For Experiments A1, A2 and C1, this property seems to hold, because the *coefficient of variation* (CV = SD/M) stays approximately constant for each standard duration. For C2, one might assume scalar property not to be valid, because the Weber fraction decreases up to 300 ms and remains constant beyond 300 ms. These results contradict the findings of Wearden and Lejeune (2008), who found scalar properties to hold for most discrimination tasks, but not for "classical" timing tasks such as interval production, temporal reduction and verbal estimation. Grondin (2010) found Weber fractions for standards of 200 and 1000 ms not to conform to the scalar property, but found lower Weber fractions for the shorter standard employed. Altogether, the present results fit the scalar expectancy theory quite well. This finding might be interpreted as an indirect confirmation of the internal clock models assuming a pacemaker emitting pulses being accumulated in a counter unit. Since it was not the main objective of this thesis, this issue is not discussed in more detail.

3.2.4 Other Methodical Implications

The results of the present thesis might be relevant for the debate about the relationship between Stevens' power law and Weber fractions. It was qualitatively argued that there should not be a correlation at all between the Weber fraction as a measure of resolving power and the power function exponent as a measure of sensation magnitudes (Laming, 1997; S. S. Stevens, 1961). By contrast, R. Teghtsoonian (1971) found a negative Spearman rank correlation coefficient of $r = -.44$ between Weber fractions and power law exponents when aggregating data across several sensory modalities, suggesting that lower difference thresholds go along with higher power function exponents. The current analyses yielded a correlation coefficient of $r = -.94$ between the power function exponents and Weber fractions estimated for six different standard durations. Therefore, the present results clearly support the finding of a negative correlation, even though within perceived duration and not across different modalities.

3.3 Limitations and Future Perspectives

The findings of the current thesis are limited to a certain scope of applications and might be extended by future investigations. This section discusses issues concerning the methodical limitations and perspectives followed by issues concerning the selection of stimuli.

3.3.1 Methodical Issues

An issue of generalization that might be further investigated is the evaluation of the axioms in other direct scaling procedures. The present thesis exclusively used the method of ratio production, but the method of ratio estimation could be evaluated by axiomatic means as well. Additionally, the axiomatic framework could be extended to magnitude production or estimation without a designated standard or even to other multimodal methods. Since ratio production procedures as applied in this thesis are often criticized for being prone to context effects, it could be helpful to find out whether the present findings can be generalized to a wider range of methods.

Another question remaining unanswered results from Manuscript B: Even though the form of the weighting function relating perceived and mathematical numbers was specified more precisely, the exact functional relationship is still undetermined. It was found that the weighting function is not veridical ($W(\mathbf{p}) \neq p$) but follows a power function of the form $W(\mathbf{p}) = kp^{\omega}$ with a constant exponent ω multiplied by a constant factor k. However, since ω cannot be determined yet, this might be a goal for further research as might be "correcting" the size of the power function exponent β by the size of the numerical distortion, as suggested by Schneider et al. (1974).

Furthermore, a methodical challenge results from the inconsistencies found in treating fractions and integers that concerns the axioms of weak multiplicativity and invertibility: Both axioms necessitate an experimental setting in which both fractions and integers are used as ratio production factors, because $p = \frac{1}{q}$ is required. Since the present thesis found fractions and integers to be processed differently and thus, should not be intermixed in a ratio production experiment, one might argue that the axioms of weak multiplicativity and invertibility are bound not to be valid due to the numerical distortion. Therefore, an alternative axiomatic framework might be developed avoiding the joint application of fractions and integers as ratio production factors to test the standard dependency of the exponent.

In addition, the axiomatic analyses offer the possibility of explicit individual testing. In the present thesis, individual inspections were carried out and integrated to draw more general conclusions on fundamental aspects of direct scaling. However, it might be interesting to find out whether the validity of the axioms is specific to certain modalities or subjects and thus, further studies should concentrate on a small number of selected participants but extend the analyses to a greater number of sensory modalities. Another potential question might arise when regarding the method: Are certain participants particularly "good" at ratio production, i.e., are the interindividual differences produced by the experimental method itself?

3.3.2 Stimulus Issues

In each of the experiment reported in this thesis, sine tones of A4 standard pitch, i.e., filled auditory intervals were used as stimuli. Since the passage of time can be perceived via other sensory modalities such as the visual or the haptic system, it has to be investigated whether the present results are generalizable for other sensory modes that can be involved in perceiving duration. Especially the visual system has to be investigated, since temporal sensitivity for visual stimuli was found to be lower than for auditory stimuli[3] (Penney & Tourret, 2005; van Wassenhove, 2009).

Furthermore, the processing of non-temporal information particularly by

[3]This was done in the same Experimental Practicum (winter term 2014/15): It was found that axiom violation rates were similar to those determined for auditory stimuli.

attention mechanisms can affect the perceived duration of an interval (S. W. Brown & Boltz, 2002; Buhusi & Meck, 2009). For example, it can be assumed that the emotional arousal (Angrilli, Cherubini, Pavese, & Manfredini, 1997; Droit-Volet & Gil, 2009) caused by a stimulus, its temporal structure or the complexity of its content (Thomas & Weaver, 1975) all have an influence on the estimated stimulus duration. Therefore, it might be investigated whether the assumptions fundamental to direct scaling are valid for stimuli that are more complex than sine tones. These were neither temporally structured nor spectrally complex or changing and thus might be "easier" to adjust than intervals filled with distracting content[4].

The aim of the present thesis was to investigate the perception of very short durations. Therefore, another question resulting from this thesis is what happens to duration perception and especially to the validity of the axiomatic framework when using standard durations longer than 1000 ms. Since the procedure requires ratio production, this is somewhat difficult, because the total duration of stimulus presentation, i.e., the standard, the inter-stimulus interval and the comparison multiplied by a certain factor should not exceed an upper limit of about 5 to 6 s. Otherwise, undesired memory effects might become an issue. Developing an experimental setting employing longer standards and still investigating duration perception – and not estimation – could be an interesting aim for further studies.

A further research question subsequently arises from the findings on context effects: Besides the standard dependency of the power function exponent, other stimulus characteristics, e.g., their range, intensity, and structure, should be investigated, because they might have a systematic effect on the size of the exponent as well. The instructions, especially the number values assigned to the stimuli, could be systematically varied to test their influence on the size of the power function parameters. Since there is no pertinent axiomatic framework for an empirical evaluation, axioms comparable to weak multiplicativity and

[4]For the visual stimuli, this was piloted in the same Experimental Practicum (winter term 2014/15): For visually structured intervals, more violations of commutativity and multiplicativity were found with increasing complexity, indicating the visual structure to impair participants' ability to process duration on a ratio scale. Temporally structured stimuli produced the fewest axiom violations at a low frequency of visual events, showing moderate temporal structuring to facilitate ratio perception.

invertibility (Augustin, 2008) might have to be developed. Otherwise, an experiment similar to Experiment C1 might be conducted.

Chapter 4

Manuscript A: Octuplicate This Interval! Axiomatic Examination of the Ratio Properties of Duration Perception

Abstract

The relationship between the physical intensity of a stimulus and its perceived magnitude can be described by Stevens' power law (Stevens, 1956), i.e., a power function with an exponent depending on the sensory modality studied. Direct scaling methods used to determine the power function exponent are based on the assumption that subjects are capable of processing ratios of magnitudes. The present experiments investigate whether this assumption holds for duration perception by empirically testing Narens' (1996) fundamental axioms of monotonicity, commutativity and multiplicativity. To determine, whether the exponent can be interpreted in a meaningful way, i.e. whether it is invariant under changes of the reference stimulus, two further axioms, invertibility and weak multiplicativity (Augustin, 2008) are evaluated. $N = 25$ participants were required to adjust the duration of a comparison tone to specific ratios of different standard durations in two experiments. In accordance with

previous findings for other sensory continua, monotonicity held for the duration adjustments of most participants. Significant violations of the commutativity axiom were found in 12.5% of all pertinent tests, whereas multiplicativity was violated in 32% of such tests. The axioms of weak multiplicativity and invertibility, however, were violated in over 50% of the tests. These results indicate that even though a ratio scale for perceived duration exists, the numbers as used by the participants cannot always be taken at face value and that even though power functions fit the data quite well, the exponent depends on the size of the standard and therefore cannot always be interpreted in a meaningful way.

4.1 Introduction

Specifying the relationship between physical time and perceived duration has been explored in many facets in psychophysics. Particularly when duration perception is compared with other sensory modalities, Stevens' power law is invoked. Employing it implies two related, and fundamental questions: First, whether perceived duration satisfies the condition of ratio scalability and second, whether the power law parameters obtained in duration scaling experiments remain unaffected by certain characteristics of the task. This study examines these questions by testing the validity of a number of pertinent axioms from representational measurement theory.

The relationship between the physical intensity of a stimulus and its perceived magnitude can be described by Stevens' power law (1956, 1975), which is formulated as:

$$\varphi(t) = \alpha t^{\beta}, t > 0. \tag{4.1}$$

That is, the perceived magnitude of a stimulus t is described by a power function αt^{β}. Whereas the parameter α is a proportionality factor depending on the units used, the exponent β depends on the sensory modality. If the value of β is > 1, the perceived magnitude of the stimulus grows faster than the intensity of the physical stimulus. If β is < 1, the increments in perceived stimulus magnitude become smaller with increasing physical stimulus intensity. In the case of $\beta = 1$, there is a directly proportional relationship between physical and perceived stimuli, i.e., the relationship can be described by a simple linear function.

Physical time and its perceived duration were also found to be related by a power function (Allan, 1979; S. S. Stevens & Galanter, 1957). The power function was fitted in several experiments applying different scaling methods (H. Eisler, 1975), among them ratings and magnitude estimation with and without a standard (Bobko et al., 1977). These approaches yielded exponents ranging from 0.44 to 1.87 (Kornbrot, Msetfi, & Grimwood, 2013), with an average exponent of 0.90 most suitably describing the relationship between physical and perceived duration (H. Eisler, 1976).

Established methods to determine the exponent of Stevens' psychophysical function are scaling procedures, in which participants are asked to produce correspondences between the perceived intensity of stimuli and numerical values consistent with the instruction. Stevens (1956) described two direct scaling methods, which are called magnitude estimation and magnitude production.

Though Stevens, in his later writings (e.g., S. S. Stevens, 1975) expressed a preference for using these methods without any constraints such as fixed standards or pre-assigned numerical values, their earliest applications were implemented in a similar manner as the classical methods to measure sensory thresholds, that is they used a fixed stimulus, the *standard*, and a variable stimulus called the *comparison*. These versions of magnitude estimation and production have later been termed "ratio estimation" and "ratio production", respectively (Gescheider, 1997).

There are two implicit assumptions fundamental to these direct scaling procedures: It is assumed that the participants are able to estimate or to produce perceived intensities on ratio scale level and, furthermore, that the numerals the participants use to describe their sensations may be treated like rational numbers in mathematics and therefore can be taken at face value.

Narens (1996) may be credited with making these assumptions explicit - never actually tested by Stevens or his followers - and formulated mathematical axioms providing a possibility to validate them. He distinguishes between behavioral and cognitive axioms: The untestable cognitive axioms describe the relationship between the participant's unobservable sensation of a stimulus' intensity and its numerical representation. The behavioral axioms characterize the participant's behavior in a scaling experiment and relate their numerical representation to the number words used to describe the stimulus' intensity. In contrast to the cognitive axioms, the behavioral axioms are empirically testable.

The behavioral axioms crucial for the assumption that participants are capable of estimating or producing ratios of stimulus intensities are monotonicity, commutativity and multiplicativity. Their validity can be evaluated by analyzing data collected in magnitude or ratio production experiments (Luce, 2002). In the latter, when applied to the psychophysics of duration, the participant is instructed to adjust the duration of a comparison stimulus (such as *w, x, y, z* in the following), of the ratio of $\mathbf{p}$, $\mathbf{q}$ or $\mathbf{r}$ of the perceived duration of the standard stimulus t: The notation $(x, \mathbf{p}, t)$ represents a participant's adjustment x, that is perceived to last $\mathbf{p}$ times as long as the standard interval t, with the bold face letter referring to the number word used in the magnitude production instructions.

First of all, besides a number of technical axioms concerning the continuity of the physical stimulus values, the axiom of monotonicity (Augustin, 2008; Axiom 3.1 in Narens, 1996), also known as *ordering*, has to be tested. It is formulated as:

$$\text{If } (x, \mathbf{p}, t) \in E \text{ and } (y, \mathbf{q}, t) \in E, \text{ then } p > q \Leftrightarrow x \succ y. \tag{4.2}$$

That means, if x has been adjusted to appear $\mathbf{p}$ times as long ($\times\mathbf{p}$, in the following) as the standard t and another adjustment y is $\mathbf{q}$ times as long ($\times\mathbf{q}$, in the following) as the standard, and $\mathbf{p}$ is greater than $\mathbf{q}$, then the adjusted duration x must be longer than the duration y. According to Narens' (1996) theory, if the axiom of monotonicity holds it can be assumed that the perception of stimuli of the investigated modality occurs on a sensory continuum. It is a necessary condition not only for the subsequently elaborated axioms of commutativity and multiplicativity, but also fundamental to any scaling at all, because even the categories of an ordinal scale can be arranged in an ascending or descending and therefore monotonic order.

Furthermore, the axiom of commutativity can be evaluated, which is formulated as:

$$\text{If } (x, \mathbf{p}, t) \in E, (z, \mathbf{q}, x) \in E, (y, \mathbf{q}, t) \in E, \tag{4.3}$$
$$\text{and } (w, \mathbf{p}, y) \in E, \text{ then } z = w.$$

In other words, commutativity holds, if the stimulus duration resulting from a successive production sequence $\times\mathbf{p} \times \mathbf{q}$ is equal to the stimulus duration resulting from successive adjustments with interchanged ratio production factors $\times\mathbf{q} \times \mathbf{p}$. For example, doubling the duration of a standard tone and then then tripling the outcome should result in the same final duration as tripling the standard duration first and then doubling the result. Narens showed that, if the axiom of commutativity holds, it can be assumed that the participant perceives stimulus magnitudes of the investigated modality on ratio scale level. But even if a ratio scale of perception might exist, there is no evidence that the scale values used by the participants can be interpreted as scientific numbers. To show the latter, the axiom of multiplicativity has to be evaluated, which is formulated as:

$$\text{If } (x, \mathbf{p}, t) \in E, (z, \mathbf{q}, x) \text{ and } r = qp, \text{ then } (z, \mathbf{r}, t) \in E. \tag{4.4}$$

In other words, the multiplicativity property holds, if the stimulus duration resulting from the successive adjustments $\times\mathbf{p} \times \mathbf{q}$ is equal to the stimulus duration resulting from a single adjustment $\times\mathbf{r}$ with r being the mathematical product of p and q. For example, doubling the duration of a standard tone and then tripling this adjustment should result in the same final duration as making the standard six times as long in a single adjustment. If the axiom of multiplicativity holds, the numerals as used by the participants to describe the perceived stimulus magnitudes can be taken at face value.

During the last decade, the axiomatic approach to magnitude scaling pioneered by Narens (1996) has been extended by Luce and colleagues (Luce, 2002, 2008; Luce et al., 2010). One recent interpretation concerning the axiom of multiplicativity argues, that a veridical interpretation of numbers and thus the validity of multiplicativity is not mandatory for direct ratio scaling: If the axiom of commutativity is satisfied, thus implying ratio scalability for the modality studied, it may be said that the participants interpret the numbers as some ratio, though not the exact ratio stated in the instructions.

The axiomatic framework has been applied to a number of psychophysical dimensions such as loudness (Ellermeier & Faulhammer, 2000; Steingrimsson & Luce, 2005a, 2005b; Zimmer, 2005), area (Augustin & Maier, 2008), brightness (Steingrimsson, 2011; Steingrimsson, Luce, & Narens, 2012), and, most recently, pitch (Kattner & Ellermeier, 2014). Duration perception, however, has not been studied in this axiomatic manner.

Therefore, the aim of the first experiment was to investigate, whether the fundamental axioms of Narens' theory hold for duration perception, i.e., whether participants are capable of processing durations on a ratio scale. This was tested in a ratio production experiment, in which participants were required to adjust the duration of a comparison tone to specific positive integer ratios of two different standard durations ($t_1 = 100$ ms, $t_2 = 400$ ms).

The experiment employed a method that is typical for axiomatic testing requiring the participant to adjust the duration of the comparison interval in an iterative fashion until it subjectively matches with the desired ratio. In contrast to one-shot-estimations (e.g., "Turn the sound off as soon as it is p times as long."), which seem to be less cumbersome, this procedure does not introduce a bias due to motor latency. Furthermore, the initial duration of the comparison was randomly chosen to fall above and below the estimated "target duration" for the purpose of counterbalancing trials in which the participants had to shorten or lengthen the comparison tone.

In the second experiment, two further axioms, weak multiplicativity and invertibility (Augustin, 2008), were tested to provide evidence for the psychological meaningfulness of scaling perceived duration, i.e., whether the size of the power law exponent for duration perception remains unaffected by the size of the standard used in ratio production. Again, participants had to adjust the duration of auditory intervals to a certain ratio with respect to a standard tone ($t_3 = 600$ ms). This time, fractions as well as integers were used as ratio production factors.

4.2 Experiment 1

4.2.1 Method

Participants

$N = 10$ participants took part in the experiment. The sample consisted of 4 female and 6 male participants with a median age of 24 years ranging from 21 to 56 years. They did not have any prior knowledge of the hypotheses being tested. The experiment was conducted individually in a double-walled sound-attenuated listening chamber (IAC).

Stimuli and Apparatus

All stimuli were sine waves of the same frequency of 440 Hz (A4 standard pitch) converted with a sampling rate of 44.1 kHz, and with 16-bit resolution. Their duration varied, as a result of the protocol, and contained 10 ms cosine-shaped rise and decay ramps to avoid unwanted switching transients. The standards were of fixed durations of 100 and 400 ms or of individual duration generated according to the adjustments of the participants. The comparison stimuli varied accordingly; their initial length was randomly chosen between 1 and 10 times the duration of the corresponding standard. The tones were preset to a comfortable sound pressure level of 65 dB SPL. After passing through a headphone amplifier (Behringer HA 800 Powerplay PRO 8), the tones were presented diotically via headphones (Beyerdynamics DT 990 PRO). The experiment was programmed in MATLAB using the PsychToolbox-package by Brainard (1997) and Pelli (1997).

Procedure

In the first time-production experiment, the participants had to complete 264 trials altogether. They were divided into four identical test sessions taking place on different days. Each session was composed of three blocks of 22 trials resulting in a total of 66 trials, respectively. After the completion of a block, the participants were allowed to take a short break. The recording of the data started after the participants had become familiar with the task during three practice trials at the beginning of each session.

Each trial consisted of two duration intervals marked by continuous tones, which were presented successively. The first tone, or standard, was of fixed duration, either 100 or 400 ms, while the second tone, or comparison, was of variable starting duration and could be adjusted by the participants. The tones were separated by a fixed silent inter-stimulus interval of 500 ms. During the presentation of both tones, a yellow numeral $\mathbf{p}$ ($\mathbf{p} = 1, 2, 3, 4, 6, 8$) was displayed in the upper part of the screen, which was the instruction for the participant to adjust the duration of the second tone so that it was perceived to be $\mathbf{p}$-times as long as the first tone. The adjustments could be made by pressing either the left cursor key for decreasing or the right cursor key for increasing the duration of the comparison tone. The steps for incrementing/decrementing duration were $\frac{1}{20}$ of the standard interval, that is 5 ms for the standard of 100 ms and 20 ms for the standard of 400 ms. To increase step size, participants could press the shift key together with the cursor key resulting in steps being ten times as long as the original steps, that is 50 ms or 200 ms, respectively.

The participants were asked to adjust the duration of the comparison tone step by step, i.e., after each key press response, the current standard and the altered comparison were replayed and the instruction was presented again. The participants were instructed to adjust the comparison until they were satisfied with the result and to eventually press the enter key to register the final value. The next trial started after an inter-trial interval of 2000 ms. There was no time restriction to performing the task.

In each of the blocks, the standards of $t_1 = 100$ and $t_2 = 400$ ms were combined with the ratio production factors $\mathbf{p} = 1, 2, 3, 4, 6$ and 8. These trials are called *basic trials* and their outcomes are primarily used to test monotonicity. The testing of commutativity and multiplicativity is based on the outcomes of so-called *successive trials*, in which the individual adjustments produced by the participants in the basic trials were used as standards. They were combined with the ratio production factors $\mathbf{q} = 2, 3,$ and 4. Each type of adjustment was made 12 times, $i = 12$. In the following, the basic adjustments are indicated by $(x_i, \mathbf{p}, t)$. As an example, $(x_3, \mathbf{2}, 100)$ is the third ($i = 3$) adjustment of a trial with a ratio production factor $\mathbf{p} = 2$ and a standard stimulus $t = 100$ ms.

In the successive trials, for each participant, the individual basic adjustments of each $(x_i, \mathbf{p}, 100)$ and $(x_i, \mathbf{p}, 400)$ were used as standard stimuli. More precisely, the new standards $(x_i, \mathbf{2}, 100)$ and $(x_i, \mathbf{2}, 400)$, derived from a basic doubling trial, had to be made $\mathbf{q} = 2, 3$, and 4 times as long. Likewise, the standards $(x_i, \mathbf{3}, 100)$, $(x_i, \mathbf{4}, 100)$, $(x_i, \mathbf{3}, 400)$ and $(x_i, \mathbf{4}, 400)$ were subsequently doubled $(\mathbf{q} = 2)$. The procedure might become more obvious by inspecting Figure 4.2: The arrows starting from the x-axis depict the basic adjustments, whereas the arrows starting from the arrowheads depict the successive adjustments.

On the whole, there were 22 different types of adjustments: Each of the two standard stimuli was paired with each of the six ratio production factors $\mathbf{p} = 1, 2, 3, 4, 6$, and 8, resulting in 12 types of basic $\times\mathbf{p}$ adjustments. In addition, each standard was combined with each of the five pairs $(p, q) = (2, 2), (2, 3), (2, 4), (3, 2)$ and $(4, 2)$, resulting in 10 different types of successive $\times\mathbf{p} \times \mathbf{q}$ adjustments. Each type of adjustment was made twelve times, resulting in 264 trials per participant.

4.2.2 Results and Discussion

Overall Results

Overall mean adjustments for $(N = 10)$ participants are depicted in Figure 4.2, in the upper panel for the shorter standard duration of $t_1 = 100$ ms and in the lower panel for the longer standard of $t_2 = 400$ ms.

The mean number of adjustments made in one trial was $M = 13.3$. In 66% of all trials, participants made fine-step adjustments of duration. In further analyses, after a brief descriptive overview, the data sets of each participant are treated separately.

Monotonicity

The axiom of monotonicity was tested to confirm that duration perception of short intervals (100 to 4000 ms) occurs on a sensory continuum, i.e., that unequal temporal intervals are perceived as such and can be discriminated, respectively. From a descriptive point of view, the axiom of monotonicity seems to hold, because, as Figure 4.2 shows, the mean outcome durations increase for increasing ratio production factors.

For the inferential statistics, two one-factor, repeated-measures analyses of variance (ANOVAs) tested the effect of the ratio production factor on the mean individual duration adjustments produced in basic trials only, separately for the two standards. For the standard $t_1 = 100$ ms, the ANOVA yielded significant differences among the different ratio production factors, $F(5, 45) = 306.9, p < .001, \eta^2 = .97$. A post hoc Tukey HSD test was conducted to check whether the mean adjustments of a pair of two adjacent ratio production factors are similar ($\sim$). The results showed that all pairs of ratio production factors $((x, \mathbf{1}, 100) \sim (x, \mathbf{2}, 100), (x, \mathbf{2}, 100) \sim (x, \mathbf{3}, 100), (x, \mathbf{3}, 100) \sim (x, \mathbf{4}, 100), (x, \mathbf{4}, 100) \sim (x, \mathbf{6}, 100)$ and $(x, \mathbf{6}, 100) \sim (x, \mathbf{8}, 100))$ differ significantly at $p < .001$. For the standard $t_2 = 400$ ms, a comparable ANOVA also yielded significant variations among the ratio production factors, $F(5, 45) = 140.7, p < .001, \eta^2 = .94$. Post hoc Tukey HSD comparisons revealed significant differences for all pairs of ratio production factors, $p < .001$ for $(x, \mathbf{4}, 400) \sim (x, \mathbf{6}, 400)$ and $(x, \mathbf{6}, 400) \sim (x, \mathbf{8}, 400)$, $p < .01$ for $(x, \mathbf{1}, 400) \sim (x, \mathbf{2}, 400)$ and $(x, \mathbf{3}, 400) \sim (x, \mathbf{4}, 400)$, and $p < .05$ for $(x, \mathbf{2}, 400) \sim (x, \mathbf{3}, 400)$. Further analyses of variance containing the factors block and session revealed no main effects for them, thus any practice effects can be ruled out.

Furthermore, a graphical analysis based on cumulative sums of the adjustments made, as proposed by Augustin and Maier (2008), was conducted for each participant. The axiom of monotonicity requires, that, for a fixed standard stimulus t, a ratio production factor $\mathbf{p}$ and a fixed number of repetitions i, the inequality $S(x_i, \mathbf{p}, t) < S(x_i, \mathbf{q}, t)$ holds, with $p < q$ and S representing the sum of duration adjustments x made up to the i-th trial. That is, the axiom of monotonicity holds, if for each standard t and each number i of repetitions (adjustments), the cumulative sums can be ordered by the ratio production factors used: $S(x_i, \mathbf{1}, t) < S(x_i, \mathbf{2}, t) < S(x_i, \mathbf{3}, t) < S(x_i, \mathbf{4}, t) < S(x_i, \mathbf{6}, t) < S(x_i, \mathbf{8}, t)$. Thus, for each participant and both standards t_1 and t_2, the $n = 12$ outcome durations of each type of $\times\mathbf{p}$ adjustments were summed up successively across trials. The cumulative sums, $S(x, \mathbf{1}, t), S(x, \mathbf{2}, t), S(x, \mathbf{3}, t), S(x, \mathbf{4}, t), S(x, \mathbf{6}, t)$ and $S(x, \mathbf{8}, t)$, of participant mg12, who is representative for the sample, are depicted in Figure 4.1, the left panel shows the shorter and the middle panel shows the longer standard duration. Although the outcome durations of all trials $n = 1$ to 12 were summed up successively, only the cumulative sums in the range of trials $n = 7$ to 12 are plotted, in order to avoid inspecting the effects resulting from random influences for a small number of observations.

Both graphs show that the curves for different ratio production factors never cross, e.g., that for the standard duration t_1, each cumulated outcome duration for $\mathbf{p} = 2$ is shorter than the corresponding cumulated outcome duration for $\mathbf{p} = 3$, meaning that at no point in the sequence of trials is monotonicity violated, thereby providing a more rigorous test than a comparison of overall condition means would.

Commutativity

The axiom of commutativity provides evidence for the assumption that duration perception is based on a ratio scale. For testing commutativity, adjustments produced in successive trials are analyzed: Commutativity is taken to be satisfied, if a successive $\times\mathbf{p} \times \mathbf{q}$ adjustment is statistically indistinguishable from a successive $\times\mathbf{q} \times \mathbf{p}$ adjustment, i.e., if both types of raw adjustments emanate from the same distribution. For a descriptive analysis, Figure 4.2 shows that most of the corresponding pairs of successive adjustments $\times\mathbf{p} \times \mathbf{q}$ and $\times\mathbf{q} \times \mathbf{p}$ which are connected by dashed lines coincide, indicating that the axiom holds for the overall means.

For individual inferential testing, nonparametric Mann-Whitney U-tests (two-tailed, $\alpha = .1$) for both pairs $(p, q) = (2, 3)$ and $(2, 4)$ and both standards were conducted resulting in four tests per participant and a total of 40 tests for the entire sample.

A standard significance level of $\alpha = .1$ was used, because the aim of the analysis was to accept a statistical null hypothesis, thus making it harder to assume that an axiom holds for a particular comparison. A correction for multiple comparisons was not applied for the same reason.

For the entire sample, five violations in the 40 tests of the axiom of commutativity were observed (compare Table 4.1). Four of the five violations were produced by two participants (ml06, mn21), both for the standard of 100 ms. For seven of ten participants, the axiom of commutativity held in all cases.

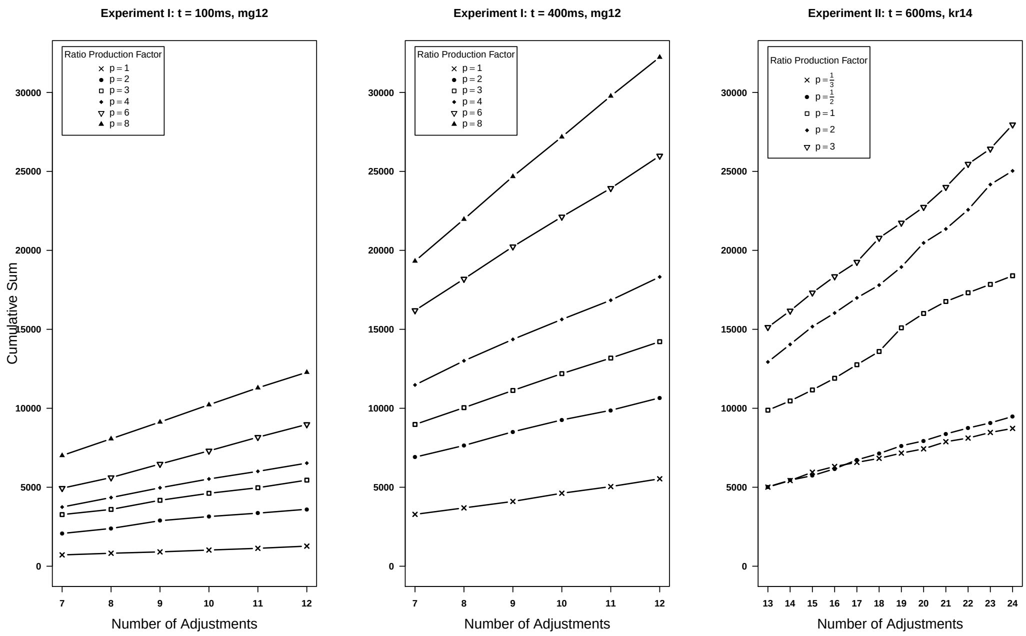

Figure 4.1: Cumulative sums of the ratio productions made in Experiments 1 and 2. Each curve depicts the cumulative sums for a particular ratio production factor **p** as a function of the trial number (7 to 12, or 13 to 24, respectively, to minimize the effect of random influences for "small" number of repetitions). The left graph refers to the shorter standard duration ($t_1 = 100$ ms), the middle one refers to the longer standard duration($t_2 = 400$ ms), both produced by participant mg12 and showing no violations of monotonicity, representative for the outcome of Experiment 1. The right graph refers to Experiment 2 and a standard of $t_3 = 600$ ms, showing magnitude productions by participant kr14 and violations of monotonicity for basic trials with $\mathbf{p} = \frac{1}{3}$ and $\mathbf{p} = \frac{1}{2}$.

Table 4.1

Experiment 1: Empirical evaluation of the commutative property for both standard stimuli with $t_1 = 100$ ms and $t_2 = 400$ ms for each ($N = 10$) participant. The table entries are $z(U)$-values of the computed Mann-Whitney U-Tests (two-tailed, $\alpha = 0.1, z_{(crit)} = 1.68$). Violations of commutativity are printed in boldface.

	$100_{p,q} = 100_{q,p}$		$400_{p,q} = 400_{q,p}$	
	(p,q)			
Part.	(2,3)	(2,4)	(2,3)	(2,4)
as11	-0.64	1.27	0.14	1.21
jb13	-0.06	1.39	0.69	0.29
mg12	0.29	1.39	1.62	0.81
mh15	-1.04	0.32	1.44	0.64
ml06	**2.02****	**2.71****	0.92	0.01
ml16	0.75	0.01	-0.23	-0.06
mn21	**-2.14***	**-1.85***	-0.06	0.55
mw28	-0.98	-1.27	0.40	0.75
tb01	0.90	0.92	1.33	1.50
we28	0.52	0.46	**-2.37***	-1.27

Levels of significance: $.1^{*}$, $.01^{**}$, $.001^{\dagger}$

Multiplicativity

The axiom of multiplicativity was tested to check whether the numerals as used by the participants can be taken at face value, i.e., whether there is a veridical transformation between perceived and mathematical numbers. For testing multiplicativity, the adjusted durations resulting from successive trials are compared with durations adjusted in basic trials: The axiom holds, if the duration resulting from the successive $\times\mathbf{p}\times\mathbf{q}$ ($\times\mathbf{q}\times\mathbf{p}$, respectively) adjustment is statistically indistinguishable from the basic $\times\mathbf{r}$ adjustment, with $r = pq$. In a descriptive manner, Figure 4.2 also shows that most of the pairs of successive adjustments $\times\mathbf{p} \times \mathbf{q}$ and $\times\mathbf{q} \times \mathbf{p}$ are commensurate with the corresponding adjustments of $\times\mathbf{r}$ (with which they are connected by dashed lines), thus indicating multiplicativity to hold for the entire sample.

The individual inferential statistics tested multiplicativity by conducting Mann-Whitney U-tests (two-tailed, $\alpha = .1$) for the three pairs $(p, q) = (2, 2)$, $(2, 3)$ and $(2, 4)$ and both standards, which results in six tests for each participant and a total of 60 tests for the entire sample. Altogether, 19 violations of 60 comparisons for the axiom of multiplicativity were observed (compare Table 4.2). For only two participants did the axiom of multiplicativity hold in all cases, whereas the other participants showed violations in one to five of six tests.

Model Fitting Procedure

Furthermore, linear regressions were computed for all participants and both standards to estimate the parameters α and β for the power law ($\varphi(t) = \alpha t^{\beta}$) as well as the parameters a and b for a simple linear function ($\varphi(t) = a + bt$). It was assumed that the individually adjusted durations of $(x, \mathbf{p}, 100)$ and $(x, \mathbf{p}, 400)$ are perceived to be $\mathbf{p}$ times as long as the standards, respectively. Thus, for the linear model, a linear regression of the ratio production factors $\mathbf{p}$ constituting the dependent variable on the individual adjustments constituting the independent variable was computed. For the power function, a linear regression was computed as well, with the logarithmically transformed ratio production factor $\mathbf{p}$ as the dependent variable and the logarithmically transformed individual adjustments serving as the independent variable.

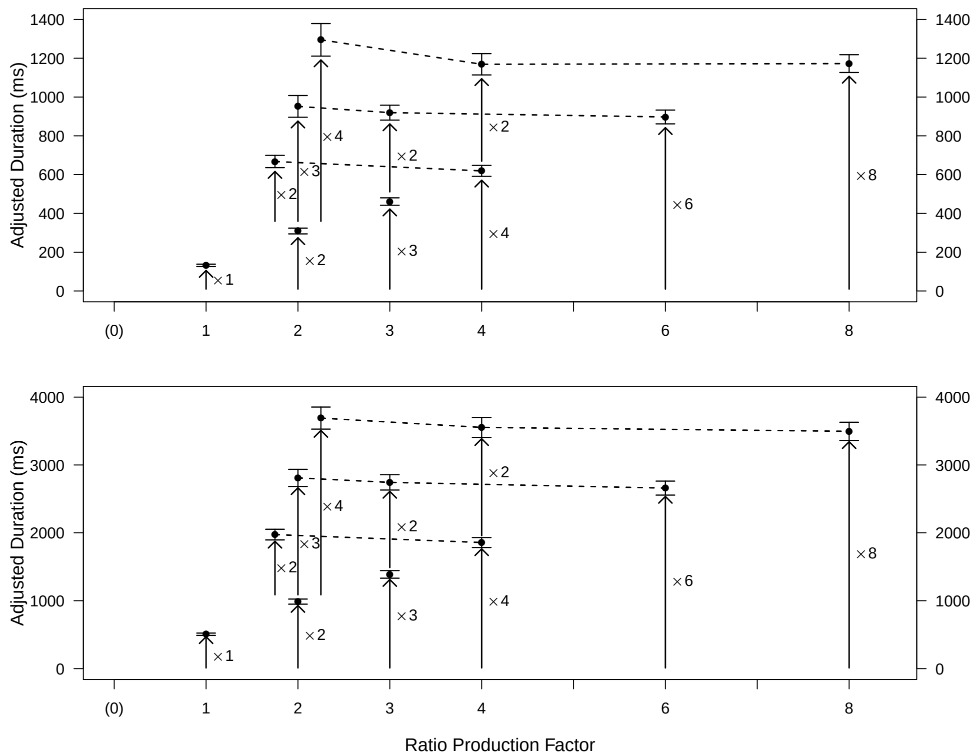

Figure 4.2: Ratio productions made by $(N = 10)$ participants in Experiment 1: Arithmetic means and standard deviations of basic and successive trials for $t_1 = 100$ ms top and $t_2 = 400$ ms bottom. Adjustments connected by dashed lines should coincide, if commutativity and multiplicativity hold.

Table 4.2

Experiment 1: Empirical evaluation of the multiplicative property for both standard stimuli with $t_1 = 100$ ms and $t_2 = 400$ ms for each ($N = 10$) participant. The table entries are $z(U)$-values of the computed Mann-Whitney U-Tests (two-tailed, $\alpha = 0.1, z_{(crit)} = 1.68$). Violations of multiplicativity are printed in boldface.

	$100_{p,q} = 100_r$			$400_{p,q} = 400_r$		
	(p,q)					
Part.	(2,2)	(2,3)	(2,4)	(2,2)	(2,3)	(2,4)
as11	0.40	-0.10	1.34	**3.41**†	**2.10***	**2.15***
jb13	**2.02***	1.17	**-1.91***	0.06	-0.39	-0.94
mg12	1.47	**1.88***	-0.52	0.64	0.13	1.41
mh15	**2.83****	**2.65****	0.97	**2.71****	**3.12****	**3.39**†
ml06	**3.23**†	**-2.89****	**1.75***	-0.55	-0.13	1.38
ml16	0.17	-0.97	0.37	-0.17	0.84	0.54
mn21	-0.09	0.64	-1.02	1.04	-0.40	-0.44
mw28	-1.13	-0.87	0.44	1.56	**2.58****	-0.07
tb01	**-2.02***	**-2.05***	-1.17	0.17	1.07	0.50
we28	-1.67	**-3.39****	-1.44	-0.23	**-1.78***	-0.71

Levels of significance: $.1^{*}, .01^{**}, .001^{\dagger}$

The estimated parameters and squared correlation coefficients R^2 for both linear model and power function and for both standards are shown in Table 4.3. The comparison between linear and power-function model shows, that for the short standard, the power-function model results in a slightly better fit ($t(13.15) = -1.885, p = .082$) explaining 4.7% more of the variance. For the longer standard, the linear model seems to fit the data as well as the power-function model ($t(15.96) = -0.735, p = .47$), the latter explaining only 2.3% more of the variance. Furthermore, the power function exponents estimated for the two standards significantly differ in size, $t(11.58) = -3.67, p = .003$. The exponent β of the power function yielded an average of $\beta(t_1) = 0.87$ ($\beta < 1$ in all cases) for the shorter standard and $\beta(t_2) = 1.02$ ($\beta > 1$ in 6 of 10 cases) for the longer standard duration. Both the linear and the power function indicate a reasonable fit to the data with R^2 ranging from 0.71 to 0.98 for the raw-data adjustments.

Summary

The analyses showed that the axiom of monotonicity was not violated, i.e., the participants were able to produce monotonically ordered adjustments according to the different ratio production factors. The axiom of commutativity was violated in 12.5% of all tests, while multiplicativity was violated in 32% of all tests. The estimated power function exponents for the two standards clearly differ in value, that is, the estimation of the parameters of the power law seems to depend on the duration of the standard, and, for the longer standard, seems to be close to 1 resulting in a simple linear function.

4.3 Experiment 2

The previous experiment investigated the axioms of monotonicity, commutativity and multiplicativity for the perception of duration to test the validity of assumptions basic to Stevens' direct scaling methods. Since the axiom of commutativity was found to be valid in 87.5% of all cases, it can be assumed that participants' processing of short duration in a ratio production experiment is based on a ratio scale. However, it might be difficult to describe the relationship between the mathematical numbers provided in the experimental

Table 4.3

Experiment 1: Estimated parameters and squared correlation coefficients for linear model and power function for both standard stimuli with $t_1 = 100$ *ms and* $t_2 = 400$ *ms and each (*$N = 10$*) participant.*

		Linear Model			Power Function		
Part.	t	a	b	R^2	$ln(\alpha)$	β	R^2
as11	t_1	0.76	5.23	0.71	0.78	0.90	0.84
	t_2	-0.34	2.41	0.91	0.29	1.15	0.92
jb13	t_1	0.48	7.32	0.88	0.90	0.95	0.91
	t_2	-0.11	2.49	0.97	0.36	1.07	0.97
mg12	t_1	0.02	7.53	0.93	0.84	0.89	0.93
	t_2	-0.32	2.91	0.92	0.40	1.12	0.94
mh15	t_1	0.41	7.20	0.91	0.88	0.91	0.95
	t_2	-0.58	2.72	0.95	0.32	1.19	0.96
ml06	t_1	0.67	5.46	0.84	0.78	0.80	0.90
	t_2	0.29	1.41	0.98	0.24	0.87	0.98
ml16	t_1	0.91	4.09	0.76	0.71	0.85	0.85
	t_2	0.52	1.77	0.77	0.34	0.87	0.82
mn21	t_1	0.44	6.30	0.90	0.82	0.85	0.94
	t_2	-0.18	2.62	0.96	0.37	1.10	0.97
mw28	t_1	0.25	6.55	0.89	0.81	0.88	0.90
	t_2	0.47	2.13	0.80	0.36	1.05	0.85
tb01	t_1	0.46	5.51	0.89	0.77	0.88	0.91
	t_2	0.26	1.94	0.95	0.35	0.89	0.97
we28	t_1	0.57	4.86	0.89	0.73	0.79	0.93
	t_2	0.56	1.95	0.90	0.38	0.91	0.94

instruction and the numbers as interpreted by the participants, because the axiom of multiplicativity held in only 68% of the tests, i.e., roughly a third of the participants do not appear to process the numbers at their face value. Comparisons of the estimated exponents of the power functions describing the relationship between physical and perceived duration yielded significantly different exponents for the two standard durations employed.

The observation that the two different standard durations used in Experiment 1 result in diverging exponents has traditionally been classified as a context effect. In the domain of psychophysical scaling, several types of context effects have been described: Besides the stimulus range used in the experiment (Garner, 1954; Ward et al., 1996), the numerical examples given in the experimental instruction (Robinson, 1976) and the number values assigned to the standard stimuli (Beck & Shaw, 1965), or even the entire experimental context might have an influence on the size of the exponent. Therefore, the psychological meaningfulness of the exponent has been called into question (Lockhead, 1992). In contrast to this point of view, other investigators have argued that finding the "true' exponent is still possible (M. Teghtsoonian & R. Teghtsoonian, 2003; R. Teghtsoonian, 2012).

However, in the axiomatic-measurement literature, this problem has been framed as a more fundamental issue of *meaningfulness* (Luce, 1978; Narens, 1981; S. S. Stevens, 1946). For each power function describing the relationship between the physical intensity of a stimulus and its perceived magnitude, one might ask whether the parameters of this function are psychologically meaningful, i.e., invariant under certain transformations. Note that the exponent of the power function depends on the sensory continuum, the participant's individual perception – which does not exert a very strong influence (M. Teghtsoonian & R. Teghtsoonian, 1983) – and potential contextual influences as mentioned above. Furthermore, it might also vary under changes of the physical measurement scale f (Narens & Mausfeld, 1992) and the size of the standard (Augustin, 2008) used in an experiment. If, for example, the measurement scale f is transformed to another scale g measuring the same physical intensity as f and if these scales are neither log-interval nor ratio scales, then it must be assumed that the choice of the scale has an influence on the exponent of the power function. Thus, the obtained exponent is not meaningful.

But even if the exponent of the power function is invariant under changes of the physical stimulus scale applied in the experiment, it has to be investigated, whether the exponent is invariant under changes of the standard stimulus t being the basis for the estimates or adjustments made by the participants. Augustin (2008) suggests a mathematical method to examine the dependency on the standard by postulating two further axioms that can be evaluated empirically that is weak multiplicativity and invertibility. The axiom of *weak multiplicativity* is formulated as:

$$\text{For } t, y, z \in X \text{ and a real number } p > 0, (y, \mathbf{p}, t) \in E, (z, \mathbf{1/p}, y) \in E \quad (4.5)$$
$$\Rightarrow (z, \mathbf{1}, t) \in E.$$

That means, weak multiplicativity holds, if the stimulus intensity resulting from successive adjustments $\times\mathbf{p} \times \frac{\mathbf{1}}{\mathbf{p}}$ is equal to the stimulus intensity resulting from the basic adjustment with $\mathbf{p} = 1$. For example, doubling the duration of the standard and then halving this adjustment should result in the same final duration as matching the duration of the comparison interval to that of the standard. Weak multiplicativity is very similar to Narens' axiom of multiplicativity. But while multiplicativity has to hold for all cases $\mathbf{p} > 0$ and $\mathbf{q} > 0$, weak multiplicativity is a special case of multiplicativity with $\mathbf{q} = \frac{\mathbf{1}}{\mathbf{p}}$, i.e., even if the axiom of multiplicativity is violated, the axiom of weak multiplicativity might hold.

The axiom of *invertibility* is formulated as:

$$\text{For } t, y \in X \text{ and } \mathbf{p} > 0, (y, \mathbf{p}, t) \in E \Leftrightarrow (t, \mathbf{1/p}, y) \in E. \quad (4.6)$$

In other words, invertibility holds, if the intensity of a stimulus resulting from successive adjustments $\times\mathbf{p} \times \frac{\mathbf{1}}{\mathbf{p}}$ is equal to the stimulus intensity of the standard t or, put simply, if it is possible to undo a $\times\mathbf{p}$ adjustment by requiring to produce its reciprocal $\times\frac{\mathbf{1}}{\mathbf{p}}$. So weak multiplicativity and invertibility differ in whether the successive adjustment resulting from $\times\mathbf{p} \times \frac{\mathbf{1}}{\mathbf{p}}$ is equal to the adjustment of $\times\mathbf{1}$ in the first case and the actual duration of the standard in the second case. As Augustin (2008) stated, both axioms are necessary and sufficient conditions for the exponent of Stevens' power law to be invariant under changes of the standard t.

However, previous magnitude production experiments using ratio production factors $p < 1 < q$ assume fractions and integers to be processed differently: A study by Luce, Steingrimsson and Narens (2010) showed the axiom of commutativity to be violated for the $N = 2$ participants tested when fractions and integer ratios were mixed. Steingrimsson and Luce (2007) found comparable discrepancies for the axiom of multiplicativity for $N = 3$ participants in an experiment on loudness production. Augustin (2008) explicitly tested the two crucial axioms of weak multiplicativity and invertibility and found them to be violated for all $N = 10$ participants who performed ratio productions of the area of visually presented circles.

For the perception of duration, numerous experiments to determine the exponent of Stevens' power law were conducted using standard durations ranging from 50 ms to 300 s (H. Eisler, 1976). Although the exponents derived from these experiments vary between $\beta = 0.23$ and 1.36, it has not been sufficiently investigated whether these differences may be caused by the use of different standards. A study by Kane and Lown (1986) used standard durations of 30 and 180 s and did not find the length of standard duration to affect the size of the power law exponent. Eisler's (1976) review of 111 studies on duration perception, however, reported lower exponents obtained from experiments using standard durations shorter than 500 ms, but they did not specify this observation in more detail. Because, in contrast, even the exponents derived from Experiment 1, using standards of $t_1 = 100$ and $t_2 = 400$ ms, significantly differ in size, $\beta(t_1) = 0.87, \beta(t_2) = 1.02$, it is plausible to investigate the meaningfulness of the power law exponent for the perception of duration by means of Augustin's (2008) additional axioms.

4.3.1 Method

Participants

$N = 15$ participants were tested in the experiment. The sample consisted of 14 female participants and one male with a median age of 23 years, ranging from 23 to 45 years. They were all students of psychology, but did not have any prior knowledge of the current hypotheses. Again, testing was conducted individually in a double-walled sound-attenuated listening chamber (IAC).

Stimuli and Apparatus

Stimuli were generated using the same apparatus and signal parameters as in Experiment 1. The fixed standard, however, had a duration of 600 ms, while comparison stimuli varied in duration; their initial length was randomly chosen between 200 and 1800 ms.

Procedure

In Experiment 2, the participants had to complete 216 trials altogether. The trials were divided into two identical test sessions taking place on two different days. Each session was composed of 12 blocks of nine trials, each. After three practice trials, data were recorded. After having completed three blocks, the participants could take a short break.

As in Experiment 1, participants had to adjust the comparison interval, separated from the standard[1] by an inter-stimulus interval of 500 ms, according to a certain ratio production factor $\mathbf{p}$ ($\mathbf{p} = \frac{1}{3}, \frac{1}{2}, 1, 2, 3$) presented on the screen. To increase or decrease the duration of the comparison interval, participants had to press the appropriate cursor key, either in small (20 ms) or in large steps (200 ms). Again, both tones were replayed after each keystroke, with the comparison tone having changed in duration.

In each of the 24 blocks, the standard of $t_3 = 600$ ms was combined with the ratio production factors $\frac{1}{3}, \frac{1}{2}, 1, 2$, and 3 resulting in five types of $\times\mathbf{p}$ adjustments and 120 basic trials altogether. In the successive trials, the individual basic adjustments $(x_i, \mathbf{p}, 600)$ were used as standard stimuli, i.e., the new standard $(x_i, \frac{1}{3}, 600)$ had to be adjusted using the ratio production factor $\mathbf{q} = 3$, the standard $(x_i, \frac{1}{2}, 600)$ was combined with the ratio production factor $\mathbf{q} = 2$, the standard $(x_i, 2, 600)$ was combined with the ratio production factor $\mathbf{q} = \frac{1}{2}$ and the standard $(x_i, 3, 600)$ had to be adjusted with the ratio production factor $\mathbf{q} = \frac{1}{3}$. The four types of $\times\mathbf{p}\times\mathbf{q}$ adjustments resulted in 96 successive trials altogether.

[1] The order of standard and comparison was counterbalanced in this experiment, i.e., in 12 of 24 repetitions of each trial type, the standard was presented *after* the comparison tone. However, the analyses did not reveal any effects of the order of standard and comparison.

4.3.2 Results and Discussion

Overall Results

The overall means based on all ($N = 15$) participants are depicted in Figure 4.3. The mean number of adjustments made in one trial was $M = 7.4$. In 35% of the adjustments, participants were using large steps to reach their final decision. In further analyses, after a brief descriptive overview, the data sets of each participant are treated separately.

Monotonicity

An ANOVA on the duration adjustments yielded significant differences between the different ratio production factors, $F(4, 56) = 200.6, p < .001, \eta^2 = .93$. Post hoc Tukey HSD comparisons revealed significant differences ($p < .001$) for all but one pair of ratio production factors, i.e., $(x, \frac{1}{3}, 600) \sim (x, \frac{1}{2}, 600)$, $p = .55$.

Furthermore, a graphical analysis based on cumulative sums was applied. As suspected from the results of the Tukey test, 4 of 15 participants exhibited violations of monotonicity in their adjustments of $\times\frac{1}{3}$ and $\times\frac{1}{2}$. An example is shown in the right panel of Figure 4.1 for participant kr14, whose lines for $\times\frac{1}{3}$ and $\times\frac{1}{2}$ are at the same level or even cross. These four participants were excluded from further analyses.

Weak Multiplicativity

The axiom of weak multiplicativity is satisfied, when the outcome of the $\times\mathbf{p} \times \frac{\mathbf{1}}{\mathbf{p}}$ adjustment is statistically indistinguishable from the duration of the $\times\mathbf{1}$ adjustment. This axiom was tested by performing nonparametric Mann-Whitney U-tests (two-tailed, $\alpha = .1$) comparing the outcome of the four combinations $(p, q) = (\frac{1}{3}, 3)$, $(\frac{1}{2}, 2)$, $(2, \frac{1}{2})$, and $(3, \frac{1}{3})$ with that of the $\times 1$ adjustment. That was done individually for each of ($N = 11$) participants, resulting in a total of 44 tests. For the entire sample, 24 violations in 44 tests of the axiom of weak multiplicativity were observed. 18 of 22 violations were produced in trials with $(p, q) = (\frac{1}{3}, 3)$ and $(p, q) = (\frac{1}{2}, 2)$, while in trials with $(p, q) = (2, \frac{1}{2})$ and $(p, q) = (3, \frac{1}{3})$, only six violations of 22 tests were found; compare Table 4.4, left column.

Table 4.4

Experiment 2: Empirical evaluation of weak multiplicativity and invertibility for the standard of $t_3 = 600$ ms duration for each ($N = 11$) participant. The table entries are $z(U)$-values of the computed Mann-Whitney U-Tests (two-tailed, $\alpha = 0.1, z_{(crit)} = 1.68$). Violations of weak multiplicativity and invertibility are printed in boldface.

	$600_{p,q} = 600_{(1)}$				$600_{p,q} = 600$			
	(p,q)							
Part.	$(\frac{1}{3}, 3)$	$(\frac{1}{2}, 2)$	$(2, \frac{1}{2})$	$(3, \frac{1}{3})$	$(\frac{1}{3}, 3)$	$(\frac{1}{2}, 2)$	$(2, \frac{1}{2})$	$(3, \frac{1}{3})$
ah06	**3.90**†	**2.06***	-0.98	-1.00	**4.95**†	**1.98***	-0.99	-1.48
ar14	**4.75**†	**3.28****	-0.77	-1.63	**5.44**†	**4.45**†	0.99	0.01
ar18	**4.61**†	**2.24***	-0.15	**-2.03***	**4.95**†	**2.47***	0.99	**-2.47***
cb22	**2.86****	1.05	-1.24	**-2.21***	**2.97****	1.48	-1.48	**-3.46**†
cg26	**1.93***	0.33	-0.03	**-2.06***	**2.47***	0.49	1.48	-0.99
dy02	**4.76**†	**2.97****	-0.65	0.02	**5.94**†	**3.96**†	0.49	0.99
ek23	**4.93**†	**4.31**†	0.88	1.18	**5.94**†	**5.44**†	1.48	**1.98***
hs29	**4.33**†	**5.53**†	-0.86	-0.94	**4.95**†	**5.94**†	-0.49	-0.49
ji08	**4.68**†	**4.50**†	**-1.68***	**-2.12***	**5.44**†	**5.44**†	-0.99	-1.48
kw22	**5.44**†	**3.81**†	-0.96	0.06	**5.94**†	**5.44**†	1.48	**2.47***
lm15	0.25	0.48	-0.95	**-3.14****	**3.96**†	**4.45**†	**4.45**†	0.99

Levels of significance: $.1^{*}$, $.01^{**}$, $.001^{\dagger}$

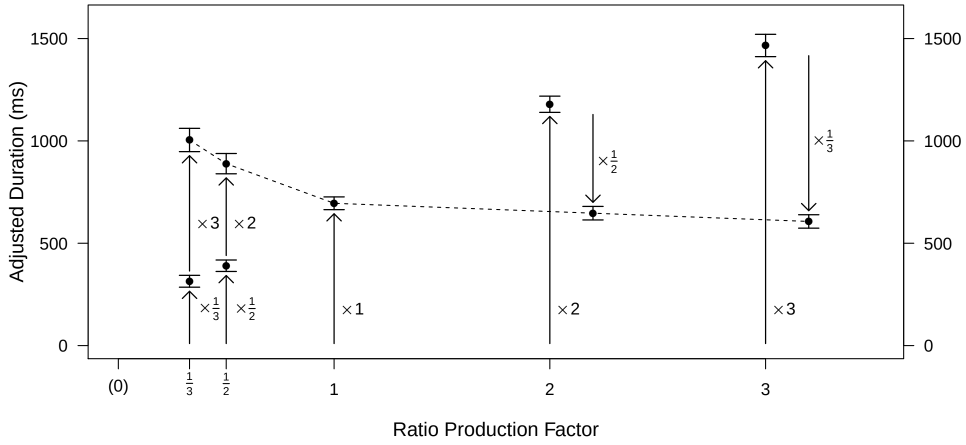

Figure 4.3: Ratio productions obtained in Experiment 2: Arithmetic means and standard deviations of basic and successive trials for $t_3 = 600$ ms and $(N = 11)$ participants. Adjustments connected by dashed lines should coincide, if weak multiplicativity holds.

Invertibility

The axiom of invertibility is satisfied, when the final outcome of the successive $\times\mathbf{p} \times \frac{\mathbf{1}}{\mathbf{p}}$ adjustments is statistically indistinguishable from the duration of the standard, $t_3 = 600$ ms. By conducting nonparametric Mann-Whitney U-tests (two-tailed, $\alpha = .1$), it was tested whether the duration adjustments of the successive trials with $(p, q) = (\frac{1}{3}, 3)$, $(\frac{1}{2}, 2)$, $(2, \frac{1}{2})$, and $(3, \frac{1}{3})$ may be produced by distributions with $\mu = 600$ ms. Tests were performed separately for the four combinations and individually for each of $(N = 11)$ participants, resulting in a total of 44 tests. For the entire sample, 25 violations of 44 tests of the axiom of invertibility were observed. 20 violations of 22 tests were produced in trials with $(p, q) = (\frac{1}{3}, 3)$ and $(p, q) = (\frac{1}{2}, 2)$, while in trials with $(p, q) = (2, \frac{1}{2})$ and $(p, q) = (3, \frac{1}{3})$, only five violations of 22 tests were found; compare Table 4.4, right column.

Model Fitting Procedure

Furthermore, regressions were computed for all participants to estimate the parameters for a linear psychophysical function as well as for a power function.

Table 4.5

Experiment 2: Estimated parameters and squared correlation coefficients for linear model and power function for the standard stimulus of $t_3 = 600$ ms and each ($N = 11$) participant.

	Linear Model			Power Function		
Part.	a	b	R^2	$ln(\alpha)$	β	R^2
ah06	-0.06	1.76	0.89	0.21	1.16	0.88
ar14	-0.27	2.07	0.78	0.22	1.31	0.82
ar18	-0.06	1.47	0.81	0.16	0.93	0.81
cb22	0.03	1.77	0.73	0.23	1.08	0.78
cg26	-0.11	1.92	0.83	0.22	1.07	0.82
dy02	-0.32	2.17	0.78	0.22	1.36	0.81
ek23	-0.59	2.69	0.79	0.29	1.60	0.79
hs29	-0.02	1.63	0.92	0.19	1.00	0.91
ji08	-0.06	1.82	0.78	0.20	0.97	0.76
kw22	-0.30	2.16	0.79	0.22	1.24	0.80
lm15	-0.12	1.77	0.75	0.18	1.05	0.82

The estimated parameters and squared correlation coefficients R^2 for both linear model and power function are shown in Table 4.5. The estimation of the exponent β of the power function revealed a $\beta > 1$ in 9 of 11 cases with an average of $\beta = 1.16$. The comparison between the two models shows no significant difference in their goodness of fit ($t(18.62) = -0.058, p = .53$), but they both explain considerably less variance, 78%, than the models fitted in Experiment 1.

Summary

The analyses showed that the axiom of monotonicity was violated by four participants, i.e., these participants were not able to produce monotonically ordered adjustments for the ratio production factors $\mathbf{p} = \frac{\mathbf{1}}{\mathbf{2}}$ and $\mathbf{p} = \frac{\mathbf{1}}{\mathbf{3}}$. Weak multiplicativity was violated in 55% of all tests. Invertibility showed a comparable violation rate of 57%. For $\times\frac{\mathbf{1}}{\mathbf{p}} \times \mathbf{p}$ adjustments, both axioms were violated more often $(82\%, 91\%)$ than for $\times\mathbf{p} \times \frac{\mathbf{1}}{\mathbf{p}}$ adjustments $(27\%, 23\%)$.

4.4 General Discussion

In two experiments, the present study examined the validity of a number of axioms from representational measurement theory for the ratio production of time intervals. These axioms are fundamental for determining whether subjective duration may be assumed to constitute a ratio scale, and how the numerical scale values obtained may be interpreted. Furthermore, they can confirm the psychological meaningfulness of the function describing the relationship between physical and subjective duration.

4.4.1 Axiomatic Evaluation and Model Fitting

In Experiment 1, multiple analyses revealed that, with all ratio production factors $\mathbf{p} \geq 1$, the axiom of monotonicity was corroborated, indicating that all participants were able to produce monotonically increasing durations in response to appropriate ratio instructions, thus satisfying a basic ordinal requirement for a scale.

The individual evaluation of commutativity and multiplicativity revealed large differences between participants: For some participants, such as mg12, ml16 and mw28, we found almost no axiom violations, whereas others (mh15, ml06) showed as many as five violations in ten tests. This finding implies that some participants were able to deal with the instructions of a ratio production experiment, i.e., they use the numbers presented in the experiment as they are requested to, whereas others were not.

The overall axiomatic evaluation showed the commutative property to hold for most participants (12.5% violations) implying that, generalized, they are capable of processing duration on a ratio scale.

However, the multiplicative property was violated in 32% of all tests showing that the numerals as used by the participants or in the instructions to describe perceived duration cannot always be taken at face value. Thus, Narens' (1996) axioms which are fundamental to Stevens' direct scaling approach could be validated, in that a ratio scale of duration can be assumed, but that there is no obvious way to derive the actual scale values.

The results for commutativity and multiplicativity of the present experiment are comparable with findings for other sensory continua. For the perception of area, Augustin and Maier (2008) reported violation rates of 12% for the axiom of commutativity and 61% for multiplicativity. Ellermeier and Faulhammer (2000) found commutativity to be violated in 11% of all cases and violations of multiplicativity in 94% of the tests, while Zimmer (2005) reported violations rates of 14% and 89%, with both studies examining the perception of loudness. For the perception of pitch, Kattner and Ellermeier (2014) found the axioms to be violated in 22% and 33% of all tests, respectively.

Furthermore, power function exponents fitted to the ratio productions made relative to the two different standards used in Experiment 1 turned out to differ significantly. Therefore, it was tested whether the dependency of the standard can be confirmed by axiomatic testing. The axioms of weak multiplicativity and invertibility, necessary and sufficient conditions for the invariance of the exponent of the power function under changes of the standard, were evaluated in Experiment 2.

The results show the crucial axioms of weak multiplicativity and invertibility to be violated in 55% and 57% of all cases, respectively, suggesting, as already assumed in Experiment 1, the power function exponent to depend on the size of the standard. From a scaling perspective, this might be construed as a context effect due to the use of a fallible method: ratio production. It might be argued that an unconstrained method using "no designated standard, no assigned modulus' disposes of the influence of the standard simply by omitting it. But since there is no axiomatic framework to test this (one stimulus - one response) methodology for internal consistency, we appear to be stuck with ratio production (or estimation) for the time being.

The results of the present axiomatic evaluation are comparable with findings made in the perception of area, where violation rates of 70% for the axiom of weak multiplicativity and 72.5% for the axiom of invertibility were reported (Augustin, 2008). For the perception of loudness and pitch, weak multiplicativity and invertibility were not evaluated yet.

Comparisons between a linear model and a psychophysical power function reveal both types of models to fit the data quite well, with comparably high proportions of variance explained. However, since Experiment 2 has shown that their estimated parameters depend on the size of the standard, the psychophysical functions fitted do not appear to be meaningful.

4.4.2 Implications

The results of the present experiments can be helpful to draw conclusions on the conception of further studies of duration scaling. An interesting question – suggested by one of the reviewers – might be, whether the participants who performed ratio productions of duration without any axiom violations do so for other sensory modalities, as well. That might clarify whether full compliance with the axioms is due to a superior way of handling numbers in general or whether it is specific to a given modality studied.

For successive adjustments, a systematic bias as reported in other studies was found: The final adjustments reached in successive trials, e.g., $\times 2 \times 3$, often exceeded the adjustments made in corresponding basic trials, e.g., $\times 6$. This pattern seems to be systematic, since it was found for other sensory modalities as well, e.g., Augustin (2008) reported a similar bias for area adjustments. Ellermeier and Faulhammer (2000) found $\times 2 \times 3$ loudness adjustments to be systematically higher in level than $\times 6$ adjustments, and Zimmer (2005) found the same pattern for loudness fractionation, i.e., the outcome of a $\times \frac{1}{6}$ adjustment produced less of a level reduction than the outcome of successive $\times \frac{1}{2} \times \frac{1}{3}$ adjustments. Steingrimsson and Luce (2007) investigated this bias and explained it by referring to a "numerical distortion'. They stated that the relationship between scientific numbers and numbers used by the participants is not linear but can be described by another function, e.g., by a power function with an exponent < 1 causing successive adjustments to be greater in size than basic adjustments. So, if multiplicativity as tested in this experiment fails, so-called k-multiplicativity can be tested to examine whether the relationship between scientific numbers and numbers used by the participants follows a power function with a constant exponent.

Furthermore, in Experiment 2, the very basic axiom of monotonicity was found to be violated for 4 of 15 participants. These participants did not produce distinguishable duration adjustments for ratio production factors $\mathbf{p} < 1$, although their adjustments for $\mathbf{p} \geq 1$ clearly follow a monotonic order. It can be ruled out that this finding might be due to a kind of floor effect, because in Experiment 1, even shorter durations were adjusted without any difficulty.

Furthermore, an unpublished experiment conducted in our laboratory investigated whether monotonicity, commutativity and multiplicativity can reliably be shown to hold for the fractionation of time intervals and revealed violation rates comparable to the results of Experiment 1. Thus, it might be assumed that the participants who violated monotonicity in Experiment 2 did not necessarily have difficulties in processing fractions, but might have misconceptions regarding the instructions of the mixed condition itself.

Additionally, a noteworthy order effect was observed when comparing the adjustments of $\times\frac{1}{3}\times 3$ and $\times\frac{1}{2}\times 2$ with the adjustments of $\times 3\times\frac{1}{3}$ and $\times 2\times\frac{1}{2}$: All successive adjustments $\times\mathbf{p}\times\mathbf{q}$ with $\mathbf{p} < 1$ preceding $\mathbf{q} > 1$ resulted in considerably longer outcome durations than successive adjustments with $\mathbf{p} > 1$ followed by $\mathbf{q} < 1$. Augustin (2008) did not report this pattern for the perception of area, so this finding may be assumed to be specific for the perception of duration, but will have to be further investigated.

Furthermore, it might be investigated, how exactly the exponent of the power function varies under changes of the standard stimulus. It might be plausible, as the results of the present experiments assume that the exponent increases with increasing standard duration.

4.4.3 Conclusions

In conclusion, the present experiments show that if using ratio production of temporal intervals, the measurement is based on a ratio scale, although a numerical distortion impedes an unequivocal interpretation of the scale values. Thus, before the shape of the transformation function relating perceived and mathematical numbers is determined, power law fitting using ratio production should be taken with a grain of salt.

Furthermore, the fitting of curves describing the relationship between physical and perceived time, regardless of power function or linear relationship, is difficult: Even if both kinds of models seem to describe the relationship quite well, the estimated parameters depend on the size of the reference stimulus used in the experiment and thus can hardly be interpreted in a psychologically meaningful way.

Chapter 5

Manuscript B: Axiomatic Evaluation of *k*-Multiplicativity: Investigating Numerical Distortion

Abstract

It is a well established empirical observation that most human participants do not process the numerical instructions used in production or estimation tasks veridically. Luce and collaborators (e.g., Luce, 2002; Steingrimsson & Luce, 2007) have analyzed the kind of "numerical distortion" that appears to be operating. They stated the relationship between perceived and mathematical numbers to be described by a power function, if the emprically testable axiom of k-multiplicativity holds. This study examined the validity of k-multiplicativity by testing whether the stimulus intensities resulting from successive adjustments $\times \mathbf{p} \times \mathbf{q}$ multiplied by a constant factor k are equal to the stimulus intensity resulting from single adjustments $\times \mathbf{r}$. Therefore, the data of three different ratio production experiments with a total of $N = 35$ participants were (re-)analyzed. In Experiment 1, integers were used as ratio production factors ($p \geq 1$), while in Experiment 2, only fractions ($p < 1$) were applied. In Experiment 3, both $p \geq 1$ and $p < 1$ were intermixed. In Experiments 1 and 2, k-multiplicativity held for all $n = 20$ participants. Experiment 3 revealed axiom violations

for 13 of $n = 15$ participants. The failure of 1-multiplicativity confirms the common observation that number representation in participants is not veridical. However, the validity of k-multiplicativity shows that the relationship between mathematical and perceived numbers follows a power function of the form $W(p) = kp^{\omega}$ with $k \neq 1$ and $\omega \neq 1$. However, the numerical distortion differs for fractions compared to integers.

5.1 Introduction

Direct scaling methods such as magnitude production or magnitude estimation (S. S. Stevens, 1946, 1956, 1975) typically use numbers to describe the magnitude of one or the perceived ratio of two different stimulus intensities. But it is an established result that participants' understanding of these numbers is not veridical, i.e., the internally represented magnitude of "8" may not be treated like the mathematical magnitude of the number "8".

So just as the relationship between physical and psychological magnitude of stimuli of different sensory modalities can be investigated, the transformation function between mathematical and perceived magnitude of numbers has to be determined. In his development of a theory of global psychophysical judgments, Luce (2002) has formulated this problem as determining a "subjective weighting function", and has theoretically examined several forms it can take, among them a Prelec function, claiming – as have others – that the perception of numbers is *not* generally veridical.

In this article, data collected in three ratio production experiments on the perception of short durations were analysed to examine the numerical distortion and to evaluate the particular functional form of the weighting function, $W(p) = kp^{\omega}$, relating mathematical and perceived magnitude of the numbers used in the experimental instructions.

5.1.1 Theoretical Background

The relationship between perceived and mathematical numbers has been studied in the framework of a separable representation assuming a transformation of the physical stimulus level to a perceived intensity, and subsequently to the numerical response generated in a scaling experiment. In a magnitude estimation experiment, this numerical response would be the numeral the participant assigns to a given stimulus magnitude, whereas in a magnitude production task, a number is given and the participant is required to adjust the corresponding stimulus intensity. However, there is empirical evidence – that will be discussed in this section – showing that most participants are not capable of operating with these numerical values as "scientific numbers". The deviation of perceived numbers from mathematical numbers, sometimes called "numerical distortion", can be described by a transformation or weighting function W.

Attneave (1962) claimed both the transformation between stimulus and psychological magnitude as well as that between psychological magnitude and the number continuum to be power functions. By contrast, Goude (1962) stated a linear relationship between subjective and objective number, which was disputed by Ekman and Hosman (1965): They doubted that numbers as used in magnitude estimations could be interpreted at face value. Thus they developed a model of a numerical distortion, proposing a logarithmic functional form of subjective number, which was supported by a study of Moyer and Landauer (1967).

Investigations by Curtis, Attneave, and Harrington (1968) revealed a decreasing power function to describe the relationship between perceived and mathematical number, as did Curtis and Fox (1969), who showed the psychological function of numbers used as category labels to depart from that of numbers as used in direct scaling experiments. Rule (1971) determined a Thurstone-type discriminability scale of number in an experiment requiring participants to compare the subjective magnitude of weights with that of integers from 1 to 10. He showed their discriminability not to be influenced by whether the numbers were used as fractions or multiple and obtained a power function relationship between mathematical and perceived number with an exponent of 0.49. Rule and Curtis (1973) supported these results using a non-metric conjoint measurement approach.

Following the approach of Attneave, B. Schneider, Parker, Ostrosky, Stein, and Kanow (1974) also assumed the transformation function to described a power relationship and further claimed that the exponent of Stevens' power law for a given sensory modality could be corrected by means of the exponent of the underlying number scale. Furthermore, they claimed differences in the power law exponents for different participants to be caused by differences in their numerical representation and not by differences in their perception. Banks and Hill (1974) as well as Banks and Coleman (1981) used a different methodological paradigm to examine the transformation function, but their analyses confirmed the assumption of a power transformation function of the form $W(\mathbf{p}) = \alpha p^{\omega}$.

This discussion resurfaced, when Narens (1996) formulated a theory of ratio scaling that reframed them in an axiomatic-measurement approach. Narens analyzed the implicit conditions, under which the assumptions he thought to be implied by Stevens' direct scaling procedures are valid, and thus formulated behavioral axioms fundamental to the application of Stevens' paradigm. In addition to two other axioms described in the following, Narens postulated the crucial axiom of multiplicativity (eq. (4)); an axiom, that implies a transformation function of the form $W(\mathbf{p}) = p^{\omega}$. In other words, the validity of multiplicativity is a necessary condition for the numerals used in the experiment to be interpreted as their mathematical equivalent, because the function W implies the identity function $W(\mathbf{p}) = p$.

In several experimental studies, the axiom of multiplicativity has been found to be violated for the investigated sensory modalities, so the assumption of the transformation function to be an identity function was rejected. A complete list of tests of empirical violations of multiplicativity is provided by Steingrimsson, Luce, and Narens (2012). For example, Ellermeier and Faulhammer (2000) investigated Narens' axioms of commutativity and multiplicativity for the perception of loudness and found commutativity to hold in 17 of 19 tests, while multiplicativity was violated in 16 of 17 tests. Zimmer (2005) investigated loudness fractionation and found multiplicativity to be violated, too. Thus, she tested the more general axiom of reduction invariance proposed by Luce (2002) assuming the transformation to follow a Prelec function (Prelec, 1998), but the axiom failed to hold in most of all cases, as well.

Thus, cumulative evidence led to abandoning the assumption that, as a general rule, W could be the identity function. Therefore, Steingrimsson and Luce (2007) extended previous axiomatic analyses and proposed two invariance axioms, k-multiplicativity (eq. (10)) and double reduction invariance (eq. (11)), equivalent to two possible weighting function forms of W that can be empirically tested. For the perception of loudness, the results revealed a power function to fit for most participants and a Prelec function to fit for the others.

In a theoretical approach, Augustin (2010) analyzed different extensions of the axiom of multiplicativity and suggested, as a result, either a power or a logarithmic form of the transformation function excluding a generalized Prelec function and other forms.

5.1.2 Empirical Background

In a previous study, Birkenbusch et al. (2015) investigated, whether the implicit assumptions fundamental to Stevens' direct scaling methods hold for the perception of short durations, i.e., whether individual duration perception is based on a ratio scale and whether the numbers used in the experiment and by the participants can be interpreted at face value. The three fundamental axioms formulated by Narens (1996) were tested: Monotonicity, commutativity and multiplicativity.

The axiom of monotonicity held for all participants and the axiom of commutativity was largely valid (in 87.5% of all tests), demonstrating that perceived duration is measurable on a ratio scale using ratio production. Note that Luce, Steingrimsson, and Narens (2010) showed the commutativity axiom to be both necessary and sufficient condition for a ratio scale to exist. The axiom of multiplicativity failed for 8 of 10 respondents and in 19 of 60 tests demonstrating some discrepancy between mathematical numbers and numbers interpreted by the participants.

Furthermore, the successive adjustments, e.g., $\times 2 \times 3$, were found to exceed the corresponding basic adjustments, e.g., $\times 6$, in 13 of 19 cases. This pattern appears to be systematical, since it was found for all sensory modalities examined by axiomatic methods. Ellermeier and Faulhammer (2000) found the $\times 2 \times 3$ loudness adjustments to be systematically higher in level than the $\times 6$ adjustments, and Zimmer (2005) reported the same pattern for loudness fractionation. Augustin (2008) found a similar bias for area adjustments. A possible explanation for this bias assuming the successive adjustments to produce an overshoot in contrast to the single adjustments, was rejected (Ellermeier & Faulhammer, 2000). Instead, Steingrimsson and Luce (2007) as well as Augustin (2010) claimed that a power transformation function between perceived and mathematical numbers can capture this bias.

Since the previous results demonstrate that $W(p) = p^{\omega}$ cannot be concluded, this study proceeds to explore how well a power function form may describe the relationship between mathematical numbers and numerals observed in human respondents.

5.1.3 Mathematical Background

The focus of this study is to empirically investigate the relationship between perceived and mathematical numbers and therefore, in the following, the axiomatic approach of Steingrimsson and Luce (2007) on which the present experiment is based is recapitulated and explained in detail.In this study, the notation introduced by Narens (1996) is used according to which $(x, \mathbf{p}, t)$ represents a participant's adjustment x, that is perceived to last $\mathbf{p}$ times as long as the standard t. First of all, besides a number of technical conditions concerning the continuity of the physical stimulus values, the axiom of monotonicity has to be tested. It is formulated as:

$$\text{If } (x, \mathbf{p}, t) \in E \text{ and } (y, \mathbf{q}, t) \in E, \text{ then } p > q \Leftrightarrow x \succ y. \tag{5.1}$$

That means if x has been adjusted to be $\mathbf{p}$ times as long as the standard t and another adjustment y is made to be $\mathbf{q}$ times as long as the standard, and p is greater than q, then the adjusted duration x must be longer than the duration y. According to Narens' (1996) theory, if the axiom of monotonicity holds, the stimulus perception can be assumed to occur on a sensory continuum. Furthermore, the axiom of commutativity (Narens, 1996; Axiom 4) can be evaluated, which is formulated as:

$$\text{If } (x, \mathbf{p}, t) \in E, (z, \mathbf{q}, x) \in E, (y, \mathbf{q}, t) \in E, \text{ and } (w, \mathbf{p}, y) \in E, \text{ then } z = w. \tag{5.2}$$

In other words, the commutative property holds, if the stimulus duration resulting from a successive production sequence $\times\mathbf{p} \times \mathbf{q}$ is equal to the stimulus duration resulting from successive adjustments with interchanged ratio production factors $\times\mathbf{q} \times \mathbf{p}$.

Narens (1996) as well as Luce, Steingrimsson, and Narens (2010) showed that if commutativity holds in general, the measurement instrument, ratio production, achieves a ratio-scaled measurement. In his more general axiomatization, Luce (2002) called this property 'threshold-proportion commutativity' and Steingrimsson and Luce (2007, eq. (11)), express this axiom as the *commutative form*:

$$r(p, q) = r(q, p) \tag{5.3}$$

To find evidence for the assumption that the numbers in the experimental instructions and used by the participants to produce various ratios can be interpreted as mathematical numbers, the axiom of multiplicativity (Narens, 1996; Axiom 9) has to be empirically supported. It is formulated as:

$$\text{If } (x, \mathbf{p}, t) \in E, (z, \mathbf{q}, x) \text{ and } r = qp, \text{ then } (z, \mathbf{r}, t) \in E. \tag{5.4}$$

In other words, the multiplicative property holds, if the stimulus duration resulting from the successive adjustments $\times\mathbf{p} \times \mathbf{q}$ is equal to the stimulus intensity resulting from a single adjustment $\times\mathbf{r}$ with r being the mathematical product of p and q. In Luce's (2002) terminology borrowing from his work on utility (Luce, 2001), this is *the probability-reduction property* and Steingrimsson and Luce (2007, eq. (12)) express this axiom as the *multiplicative form*:

$$r = r(p, q) = pq \tag{5.5}$$

Steingrimsson and Luce (2007) refer to this axiom as 1-multiplicativity because it is a special case of k-multiplicativity (eq. (10)) in which $k = 1$. 1-multiplicativity is equivalent to $W(p) = p^{\omega}$, so if 1-multiplicativity holds but $\omega \neq 1$, W is not the identity function. Because for $W(\mathbf{1}) = 1^{\omega} = 1$, the validity of 1-multiplicativity implies:

$$W(\mathbf{1}) = 1 \tag{5.6}$$

That means, the internal weighted magnitude of the ratio production factor $\mathbf{p} = 1$ is equivalent to the mathematical magnitude of 1. If the axiom of 1-multiplicativity (eq. (4) or (5)) holds, it can be expressed using the weighting function (6) to yield equation (7), i.e., the weighted magnitude of two successively presented **p** and **q** is equivalent to the weighted magnitude of the mathematical product of p and q, **r**:

$$W(\mathbf{r}) = W(\mathbf{p})W(\mathbf{q}) \tag{5.7}$$

So far, no numerical distortion ($W(\mathbf{p}) \neq p$) in the treatment of numbers has been assumed. On this basis, several experiments testing Narens' axioms of commutativity and multiplicativity have been performed for sensory modalities such as loudness (Ellermeier & Faulhammer, 2000; Steingrimsson & Luce, 2007; Zimmer, 2005) , pitch (Kattner & Ellermeier, 2014), area (Augustin & Maier, 2008), brightness (Steingrimsson, 2011), and duration (Birkenbusch et al., 2015). They revealed commutativity to hold in most cases, whereas multiplicativity generally failed to hold. Therefore, the assumption that participants have a veridical understanding of numbers had to be rejected. This finding is in accordance with previous findings on the perception of numbers (Attneave, 1962). As a consequence, one has to drop the assumption that number perception is veridical, i.e., multiplicativity holds, and accept another form of weighting function, although this function permits a numerical distortion. For this reason, Steingrimsson and Luce (2007) proposed to test the axiom of k-multiplicativity, if the axiom of 1-multiplicativity fails to hold. Instead of a veridical transformation function, k-multiplicativity confirms a power function weighting of the ratio production factor **p** with a constant factor k and an exponent ω:

$$W(\mathbf{p}) = kp^{\omega} \tag{5.8}$$

This function is further specified as a *general power function form*[1]:

$$W(\mathbf{p}) = W(\mathbf{1}) \begin{cases} p^{\omega}, 0 < p \leq 1 \\ p^{\omega'}, p > 1 \end{cases} \tag{5.9}$$

The general power function form distinguishes adjustments resulting from experimental instructions using fractions ($0 < p \leq 1$) and integers ($p > 1$), because empirical observations have shown fractions to be processed in

[1] $W(\mathbf{1}) = k1^{\omega} \Rightarrow k = W(\mathbf{1})$

a different way than are integers (Bonato et al., 2007; Ganor-Stern, 2012; Steingrimsson & Luce, 2007) and therefore different weighting functions are assumed.

The *multiplicative form* (5) can be combined with the *general power function form* (9) to yield a *general multiplicative form* or the *k-multiplicative property* by multiplication with a constant factor k:

$$r = r(p,q) = pq \begin{cases} k, 0 < p \leq 1 \\ k', p > 1 \end{cases} \tag{5.10}$$

In other words, the k-multiplicative property holds, if the stimulus intensity resulting from the successive adjustments $\times\mathbf{p} \times \mathbf{q}$ multiplied by a factor k is equal to the stimulus intensity resulting from a single adjustment $\times\mathbf{r}$. Of course, it has to be tested whether k is really constant over several combinations of $\times\mathbf{p} \times \mathbf{q}$ and $\times\mathbf{r}$. In a special case, if $k = k' = 1$, the axiom of 1-multiplicativity holds.

Steingrimsson and Luce further formulated the axiom of double reduction invariance equivalent to an even more general Prelec function form of W:

$$W(\mathbf{p}) = W(\mathbf{1}) \begin{cases} \exp[-\omega(-\ln\, p)^{\mu}], 0 < p \leq 1 \\ \exp[\omega'(\ln\, p)^{\mu'}], p > 1 \end{cases} \tag{5.11}$$

In other words, this property holds, if the stimulus intensity resulting from the successive adjustments $\times\mathbf{p} \times \mathbf{q}$ is equal to the double compounding of $\times\mathbf{r}$, i.e., the successive adjustments $\times\mathbf{r} \times \mathbf{r}$. In this case, the combined production of p^N and q^N is indistinguishable from the production of r^N with N being any natural number.

5.1.4 Conclusions

So in conclusion, the validity of 1-multiplicativity supports the assumption of a weighting function of the form $W(\mathbf{p}) = p^{\omega}$. Because testing 1-multiplicativity empirically yielded the axiom to be violated in nearly all tests performed, its fundamental assumption has to be refuted. Thus, the axiom of k-multiplicativity can be tested assuming $W(\mathbf{p}) \neq p$. If this weaker condition holds, the relationship between perceived and mathematical numbers follows an power function with a positive and constant exponent. However, even if the constancy and size of k allows conclusions regarding the validity of k-multiplicativity and the general form of the weighing function, it is not possible to derive the exact function $W(\mathbf{p})$.

Nevertheless, the aim of this experiment was to examine the relationship between mathematical numbers and numerals as interpreted by the participants for three different ratio production experiments on the perception of short durations. The experiments differ in the size of the ratio production factor $\mathbf{p}$: In Experiment 1, all $\mathbf{p}$ are integers ($p > 1$), whereas in Experiment 2, all $\mathbf{p}$ are fractions ($0 < p < 1$). To show the different numerical distortions of fractions and integers, in Experiment 3 fractions and integers are intermixed[2].

Because the axiom of 1-multiplicativity was repeatedly found to be violated, the assumption of a veridical interpretation of numbers is rejected. Therefore, it was tested whether the transformation function can be described by a power relationship with constant exponent by evaluating the axiom of k-multiplicativity.

5.2 Experiments

5.2.1 Methods

Three ratio production experiments were carried out differing in the ratio production factors and the length of the standard durations. In Experiment 1, integer ratio production factors $p \geq 1$ were used, while in Experiment 2, only fractions $p < 1$ were employed. In Experiment 3, both $p \geq 1$ and $p < 1$ were intermixed.

5.2.2 Participants

Both Experiments 1 and 2 were completed by $N = 10$ participants, respectively. $N = 15$ participants took part in Experiment 3. The total sample consisted of 13 male and 22 female participants with a median age of 23 years ranging from

[2]Experiment 1 and 3 have bean reported in Birkenbusch et al. (2015), though with different research goals. In the present report, these data are utilised to assess k-multiplicativity in a larger context of experimental manipulations.

21 to 56. All respondents participated voluntarily or received course credit for their participation. Testing was conducted individually in a double-walled sound-attenuated listening chamber (IAC).

5.2.3 Stimuli and Apparatus

The durations to be perceived or adjusted were marked by pure tones which were generated digitally during the experiment. They had a frequency of 440 Hz (A4 standard pitch) and were converted with a sampling rate of 44.1 kHz and 16-bit resolution by an RME Hammerfall DSP Multiface II sound card. All stimuli had 10 ms cosine-shaped ramps to smooth onsets and offsets. In Experiment 1, standard durations of 100 and 400 ms were used, whereas in Experiment 2, a standard of 2000 ms was chosen. In Experiment 3, a standard of 600 ms was used. The initial duration of the comparison stimuli varied between 1 and 8 times the standard in Experiment 1, $\frac{1}{8}$ and 1 times the standard in Experiment 2 and $\frac{1}{3}$ and 3 times the standard in Experiment 3. The tones were adjusted to a comfortable sound pressure level of 65 dB SPL by means of a headphone amplifier and presented diotically via headphones (Beyerdynamics DT 990 PRO). The experiment was controlled by MATLAB 2012 with PsychToolbox-3 by Brainard (1997) and Pelli (1997).

5.2.4 Procedure

At the beginning of each session and in each experiment, three practice trials were completed before data recording was started, so that the participants could become familiar with the task.

In each trial, two duration intervals marked by continuous tones were presented successively: The first tone (the standard) was of fixed duration, whereas the second tone (the comparison) was of variable starting duration and had to be adjusted by the participants. The tones were separated by a fixed 500-ms silent inter-stimulus interval.

During the presentation of both tones, an instruction presented on the screen including a yellow numeral **p** requested the participant to adjust the duration of the second tone so that it was perceived to be **p**-times as long as the first tone. The adjustment could be made by pressing either the left cursor key for shortening or the right cursor key for lengthening the duration of the comparison tone. The size of the steps was $\frac{1}{20}$ of the duration of the standard interval. To increase the step size, participants could press the shift key together with the cursor key resulting in steps being ten times as long as the original steps. The participants were asked to successively adjust the duration of the comparison tone until they were satisfied with the result. There was no time restriction to performing the task. After each adjustment response, both tones were replayed until the participant pressed the enter key to register the latest adjustment.

Experiment 1

In Experiment 1, participants had to complete 264 trials altogether, which were divided into four sessions each consisting of 3 blocks of 22 trials. In the so-called basic trials, the two different standard durations were combined with the ratio production factors $p = 1, 2, 3, 4, 6$, and 8, resulting in 12 types of basic $\times\mathbf{p}$ adjustments. In the so-called successive trials, the individual adjustments produced by the participants in the basic trials were used as standards. The adjustments of $\times 2$ were combined with the ratio production factors $\mathbf{q} = 2, 3$, and 4, whereas the adjustments of $\times 3$ and $\times 4$ were combined with the ratio production factor $\mathbf{q} = 2$, resulting in 10 different types of successive $\times\mathbf{p} \times \mathbf{q}$ adjustments, altogether. In the entire experiment, each of the 22 types of trials was completed 12 times.

Experiment 2

In Experiment 2, participants had to complete 120 trials, which were divided into two sessions each consisting of 6 blocks of 10 trials. In the basic trials, ratio production factors of $\mathbf{p} = \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{6}$, and $\frac{1}{8}$ were used, resulting in 5 types of basic $\times\mathbf{p}$ adjustments. In the successive trials, the adjustments of $\times\frac{1}{2}$ were combined with the ratio production factors $\mathbf{q} = \frac{1}{2}, \frac{1}{3}$, and $\frac{1}{4}$ and the adjustments of $\times\frac{1}{3}$ and $\times\frac{1}{4}$ were combined with $\mathbf{q} = \frac{1}{2}$, resulting, as well, in 5 types of $\times\mathbf{p} \times \mathbf{q}$ adjustments. As in Experiment 1, each of the 10 types of trials was repeated 12 times.

Experiment 3

In Experiment 3, participants completed 216 trials, which were divided into two sessions each consisting of 12 blocks of 9 trials. In the basic trials, ratio production factors of $p = \frac{1}{3}, \frac{1}{2}, 1, 2,$ and 3 were used, resulting in 5 types of basic adjustments. The 4 types of successive trials are composed by the $\times\frac{1}{3}$ adjustments combined with the ratio production factor $\mathbf{q} = 3$, the $\times\frac{1}{2}$ adjustments combined with $\mathbf{q} = 2$, the $\times 2$ adjustments combined with $\mathbf{q} = \frac{1}{2}$ and the $\times 3$ adjustment combined with $\mathbf{q} = \frac{1}{3}$. Each of the 9 types of adjustments was repeated 24 times.

5.2.5 Computing k-Multiplicativity

In Experiment 1, the axiom of k-multiplicativity was evaluated based on eq. (10) by dividing the mean outcome duration of the single adjustments $\times\mathbf{r}$ by the mean outcome duration of the corresponding successive adjustments $\times\mathbf{p} \times \mathbf{q}$ to calculate k. For example, the mean outcome duration of 12 basic-trial $\times 6$ adjustments was divided by the mean outcome duration of the 12 corresponding $\times 2 \times 3$ successive-trial adjustments. Thus, if a given participant's 6-times adjustment fell short of his/her successive 2-times-3-times adjustment by 10%, the resulting k was 0.90. The calculation of k was performed separately for each participant, each standard duration, and each of the trial types $(p, q) = (2, 2), (2, 3)$ and $(2, 4)$, resulting in three different values of k per participant and standard. For Experiment 2 and 3, an analogous calculation method was chosen.

To check whether k is constant across trial types, a linear regression of the form $k_{n,r} = apq_n$ was computed, with $k_{n,r}$ being the calculated k for participant r and trial type n and pq_n being the product of ratio production factors p and q in the trial type n.

To compute this linear regression with an intercept $y = 0$, both variables were normalized. To that effect, each $\overline{k_r}$, being the mean of the three $k_{n,r}$ for participant r, was subtracted from each $k_{n,r}$. Each pq_n was also normalized by subtracting $\overline{pq}$ being the mean of the ratio production factors p and q of all trial types n. By proceeding this way, the regressions are more accurate than regressions with an intercept term.

Table 5.1

Percentage of axiom violations for the different Experiments 1, 2 and 3 with different ratio production factors. Datasets of participants violating the axiom of monotonicity were excluded from the further analyses.

	Monot.	Commut.	1-MP	k-MP
Experiment 1 $p > 1$ ($n = 10$)	0%	12.5%	32%	0%
Experiment 2 $0 < p < 1$ ($n = 10$)	10%	5.5%	33%	0%
Experiment 3 Mixed ($n = 15$)	27%	91%	55%	82%

5.3 Results

The data of the three experiments were analyzed separately and thus, they are described in three separate sections. Furthermore, the validity of the axioms was tested individually for each participant and each trial type. In the following, the results for the axiomatic analyses of monotonicity, commutativity and 1-multiplicativity are recapped for each experiment. Afterwards, the axioms of k-multiplicativity is tested, respectively. Finally, the results are briefly summarized. A summary on the validity of each of the four axioms in each of the three different experiments is given in Table 5.1. In Figure 5.1, it is shown whether the averaged k values are constant for integers, fractions and the intermixed condition.

For axiomatic testing, a standard significance level of $\alpha = .1$ was used. Since the aim of the axiomatic analyses was to maintain a statistical null hypothesis, the elevated significance level makes it harder to assert the validity of an axiom in a particular comparison. Corrections for multiple comparisons were not conducted for the same reason. For the other tests aiming at rejecting null hypotheses, a standard significance level of $\alpha = .01$ was used. Moreover, α was corrected for multiple comparisons.

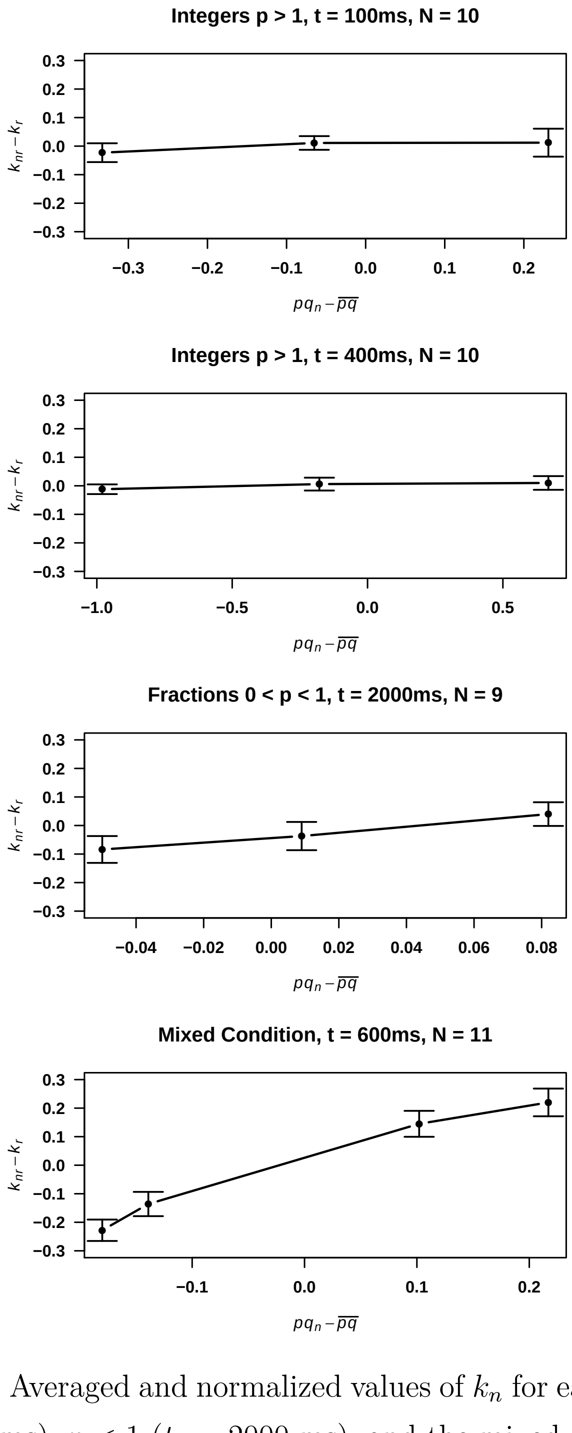

Figure 5.1: Averaged and normalized values of k_n for each $p \geq 1$ ($t_1 = 100$ ms, $t_2 = 400$ ms), $p < 1$ ($t_3 = 2000$ ms), and the mixed condition ($t_4 = 600$ ms) and the corresponding pairs of ratio production factors pq_n. If the slopes of the regressions are indifferent from zero, the axiom of k-multiplicativity holds.

5.3.1 Experiment 1

Monotonicity, Commutativity and 1-Multiplicativity

The axiom of monotonicity holds, if the adjustment of $\times\mathbf{p}$ exceeds the adjustment of $\times\mathbf{q}$ with $p > q$. Monotonicity was tested by conducting an ordinal analysis based on cumulative sums proposed by Augustin and Maier (2008). For each participant, the cumulative sums of the duration adjustments for the different ratio production factors were analyzed. If the cumulative sums of two proximate ratio production factors in a trial n do not differ significantly, this gives evidence for a violation of monotonicity. Because none of such violations was found, the analysis showed the axiom of monotonicity to hold for all participants and all trial types.

The axiom of commutativity holds, if the mean outcome duration of the successive $\times\mathbf{p}\times\mathbf{q}$ adjustment is statistically indistinguishable from the mean outcome of the $\times\mathbf{q}\times\mathbf{p}$ adjustments. The axiom was tested by conducting nonparametric Mann-Whitney U-tests (two-tailed, $\alpha = .1$) for both pairs $(p, q) = (2, 3)$ and $(2, 4)$ and both standards, which results in four tests per participant and a total of 40 tests for the entire sample. Five violations in the 40 tests were observed, which correspond to a proportion of 12.5% of all tests.

The axiom of 1-multiplicativity holds, if the duration reached by the combined $\times\mathbf{p}\times\mathbf{q}$ (respectively, $\times\mathbf{q}\times\mathbf{p}$) adjustments is statistically indistinguishable from the $\times\mathbf{r}$ adjustments, with $r = pq$. The axiom was tested by conducting Mann-Whitney U-tests (two-tailed, $\alpha = .1$) for the three pairs $(p, q) = (2, 2), (2, 3)$ and $(2, 4)$ and both standards, which results in six tests per participant and a total of 60 tests for the entire sample. Altogether, 19 violations of 60 comparisons for the axiom of 1-multiplicativity were observed, corresponding to a proportion of 32% of all tests (Birkenbusch et al., 2015).

k-Multiplicativity

The axiom of k-multiplicativity holds, if the adjusted duration of a successive trial $\times\mathbf{p}\times\mathbf{q}$ multiplied by a constant factor k is statistically indistinguishable from a single adjustments $\times\mathbf{r}$ and if k is indistinguishable over several associated $\times\mathbf{p}\times\mathbf{q}$ and $\times\mathbf{r}$ adjustments. The axiom was tested by dividing the mean outcome duration of single trials $\times\mathbf{r}$ by the mean outcome duration of the corresponding successive $\times\mathbf{p}\times\mathbf{q}$ trials to compute k. This was done for the three pairs $(p, q) = (2, 2), (2, 3)$, and $(2, 4)$ and both standards, which results in six calculations per participant and a total of 60 calculations for the entire sample. After normalizing the k values for the three different pairs of standard durations, respectively, a linear regression was computed for each participant and each standard, resulting in 20 comparisons altogether. If the slope of the resulting function was statistically indistinguishable from 0, k was assumed to be constant.

For both standards of 100 and 400 ms, all linear regressions had slopes not differing from zero and therefore, the axiom of k-multiplicativity was assumed to hold for all participants and both standard durations.

The overall mean of the computed k was $\overline{k} = 0.96$ with individual $\overline{k_r}$ ranging from 0.67 to 1.15 for the standard of 100 ms and $\overline{k} = 0.96$ with individual $\overline{k_r}$ ranging from 0.83 to 1.08 for the standard of 400 ms. So for integers being used as ratio production factors, it turned out that $k < 1$ meaning that the concatenated adjustments $\times\mathbf{p}\times\mathbf{q}$ typically exceed the single adjustments $\times\mathbf{r}$. The individual results for the computed slope of the linear regression, the p-value showing whether this slope is different from 0, the resulting statistical trend and the $\overline{k_r}$ are shown in the upper section of Table 5.2 for the standard of 100 ms and in the lower section for the standard of 400 ms.

5.3.2 Experiment 2

Monotonicity, Commutativity, and 1-Multiplicativity

The analysis based on cumulative sums showed the axiom of monotonicity to be violated for only one participant. Usually, such are excluded from further analyses, because they indicate that single adjustments of two proximate ratio production factors do not differ significantly and thus it is questionable, whether the participants can distinguish between two durations corresponding to these ratio production factors. Therefore, the data of this participant were excluded from further analyses.

The axiom of commutativity was tested by conducting Mann-Whitney U-tests (two-tailed, $\alpha = .1$) for both pairs $(p, q) = (\frac{1}{2}, \frac{1}{3})$ and $(\frac{1}{2}, \frac{1}{4})$ resulting in two tests per participant and thus 18 tests for the entire sample. One violation

Table 5.2

Experiment 1: Results for testing the k-multiplicative property for $p > 1$ and both standards $t_1 = 100$ ms and $t_2 = 400$ ms.

$t_1 = 100$ ms				
Participant	Slope a	p_{stat}	Stat. trend	$\overline{k_r}$
as11	-0.15	.63	$a = 0$	0.95
jb13	0.63	.11	$a = 0$	0.97
mg12	0.26	.32	$a = 0$	0.93
mh15	0.37	.15	$a = 0$	0.81
ml06	0.05	.12	$a = 0$	0.67
ml16	-0.12	.76	$a = 0$	0.99
mn21	0.08	.46	$a = 0$	1.01
mw28	-0.30	.15	$a = 0$	1.00
tb01	-0.18	.11	$a = 0$	1.15
we28	-0.03	.85	$a = 0$	1.13
				0.96
$t_2 = 400$ ms				
Participant	Slope a	p_{stat}	Stat. trend	$\overline{k_r}$
as11	0.04	.31	$a = 0$	0.83
jb13	0.03	.30	$a = 0$	1.02
mg12	-0.01	.78	$a = 0$	0.96
mh15	0.02	.33	$a = 0$	0.85
ml06	-0.04	.14	$a = 0$	0.99
ml16	-0.04	.62	$a = 0$	0.95
mn21	0.04	.33	$a = 0$	0.99
mw28	0.11	.38	$a = 0$	0.88
tb01	0.01	.84	$a = 0$	1.02
we28	-0.01	.94	$a = 0$	1.07
				0.96

in 18 tests was found for the axiom of commutativity, which corresponds to a proportion of 5.5% of these tests.

The axiom of 1-multiplicativity was also tested by conducting Mann-Whitney U-tests (two-tailed, $\alpha = .1$) comparing the three pairs $(p, q) = (\frac{1}{2}, \frac{1}{2}), (\frac{1}{2}, \frac{1}{3})$ and $(\frac{1}{2}, \frac{1}{4})$ of consecutive adjustments against single adjustments of $\times\frac{1}{4}, \times\frac{1}{6}$ and $\times\frac{1}{8}$, respectively. Given three tests per participant and a total of 27 tests, 9 violations were observed corresponding to a proportion of 33% of all comparisons.

k-Multiplicativity

The axiom of k-multiplicativity was evaluated by dividing the mean outcome duration in single trials by the mean outcome duration in the corresponding successive trials to compute k. This was done for the three pairs $(p, q) = (\frac{1}{2}, \frac{1}{2}), (\frac{1}{2}, \frac{1}{3})$ and $(\frac{1}{2}, \frac{1}{4})$, which results in three calculations per participant and a total of 27 calculations for the entire sample. After normalizing each participant's k values for the three different pairs, a linear regression on each triple of k values was computed. All linear regressions revealed slopes statistically indistinguishable from 0 and therefore, the axiom of k-multiplicativity was assumed to hold for all participants.

The overall mean of the computed k was $\overline{k} = 1.22$ with individual $\overline{k_r}$ ranging from 0.97 to 1.47. Thus, if fractions are used as ratio production factors, $k > 1$ meaning that consecutive adjustments $\times\frac{1}{\mathbf{p}} \times \frac{1}{\mathbf{q}}$ understroot (or produce even smaller fractions than) the single adjustments $\times\frac{1}{\mathbf{r}}$ with $r = pq$. The individual results for the slope of the linear regression, the p-value showing whether this slope is differing from 0, the resulting statistical trend and the $\overline{k_r}$ are shown in Table 5.3.

5.3.3 Experiment 3

Monotonicity, Commutativity and 1-Multiplicativity

The analysis of the cumulative sums showed the axiom of monotonicity to be violated by four of 15 participants corresponding to 27% of all participants. The data of these four participants were excluded from further analyses.

Table 5.3

Experiment 2: Results for testing the k-multiplicative property for $0 < p < 1$ *and the standard* $t_3 = 2000$ *ms.*

$t_3 = 2000$ ms				
Participant	Slope a	p_{stat}	Stat. trend	$\overline{k_r}$
chpe06	0.73	.66	$a = 0$	0.99
hawo03	1.63	.45	$a = 0$	1.29
kafr04	-0.03	.98	$a = 0$	1.06
lini10	2.21	.49	$a = 0$	1.25
mage07	2.70	.28	$a = 0$	1.32
mami04	2.87	.31	$a = 0$	1.24
simi04	2.44	.31	$a = 0$	1.42
stja03	2.78	.21	$a = 0$	1.47
urwo12	2.94	.11	$a = 0$	0.97
utwi10	0.48	.63	$a = 0$	1.21
				1.22

The axiom of commutativity was tested by conducting Mann-Whitney U-tests (two-tailed, $\alpha = .1$) for both pairs $(p, q) = (\frac{1}{2}, 2)$ and $(\frac{1}{3}, 3)$ resulting in two tests per participant and 22 tests for the entire sample. Commutativity was violated in 20 of 22 tests, which corresponds to a proportion of 91% of all comparisons.

In Experiment 3, the axiom of 1-multiplicativity is identical to the axiom of weak multiplicativity (Augustin, 2010), because $pq = 1$. Because the axiom of commutativity turned out to be violated for most of the tests, denoting, e.g., the mean outcome duration of the $\times\frac{1}{2}\times 2$ adjustments not to correspond to the mean outcome of the $\times 2 \times \frac{1}{2}$ adjustments, the axiom of (weak) 1-multiplicativity was tested separately for $(p, q) = (\frac{1}{2}, 2)$ and $(\frac{1}{3}, 3)$ and $(p, q) = (2, \frac{1}{2})$ and $(3, \frac{1}{3})$. Therefore, Mann-Whitney U-tests (two-tailed, $\alpha = .1$) were conducted for the pairs $(p, q) = (\frac{1}{2}, 2)$ and $(\frac{1}{3}, 3)$ resulting in four tests per participant and 44 tests for the entire sample. The axiom of (weak) 1-multiplicativity was violated in 24 of 44 tests corresponding to a proportion of 55% of all tests (for details, see Birkenbusch et al. (2015)). For the pairs $(p, q) = (\frac{1}{2}, 2)$ and $(\frac{1}{3}, 3)$, 1-multiplicativity was violated in 18 of 22 tests (82%), whereas for the pairs

$(p,q) = (2, \frac{1}{2})$ and $(3, \frac{1}{3})$, the axiom was violated in 6 of 22 tests (27%).

k-Multiplicativity

With regards to the previous axiomatic testing, the evaluation of k-multiplicativity in Experiment 3 turned out to be problematic, because the axiom of commutativity was violated in most cases. Usually, if there is no violation of commutativity, the adjustments of $\times\mathbf{p}\times\mathbf{q}$ and $\times\mathbf{q}\times\mathbf{p}$ are averaged for the evaluation of 1-multiplicativity. If commutativity is violated and 1-multiplicativity should be computed nevertheless, the $\times\mathbf{p}\times\mathbf{q}$ and $\times\mathbf{q}\times\mathbf{p}$ will have to be treated separately. This method is applied in testing k-multiplicativity, as well.

As before, for the different pairs of $(p,q) = (\frac{1}{2}, 2), (\frac{1}{3}, 3), (2, \frac{1}{2})$ and $(3, \frac{1}{3})$, $k_{n,r}$ can be computed by dividing their mean outcomes by the outcome of the $\times 1$ adjustment.

Instead of averaging the individual $k_{n,r}$ of the different trial types n and computing $\overline{k_r}$, however, the different $\overline{k_n}$ for each trial type n and over all participants r were computed. The means over all $\overline{k_n}$ revealed $\overline{k} = 0.79$ for the pair $(p,q) = (\frac{1}{2}, 2)$, $\overline{k} = 0.70$ for the pair $(p,q) = (\frac{1}{3}, 3)$, $\overline{k} = 1.07$ for the pair $(p,q) = (2, \frac{1}{2})$ and $\overline{k} = 1.15$ for the pair $(p,q) = (3, \frac{1}{3})$. It appears that in the consecutive adjustments, fractions followed by integer multiples overshoot the target adjustment of 1.0, while the opposite sequence falls short of 1.0.

To check whether the $\overline{k_n}$ for the different trial types are statistically different, a repeated measures ANOVA was computed with the four different pairs being the independent and the $\overline{k_n}$ being the dependent variable, $F(3, 42) = 68.58, p < .001$. Therefore, the axiom of k-multiplicativity can be assumed to be violated, because $\overline{k_n}$ is not constant over different types of trials.

Furthermore, there is a remarkable difference between successive trials, in which integers are followed by fractions ($\overline{k} = 1.11, k > 1$) in contrast to trials in which fractions are followed by integers ($\overline{k} = 0.75, k < 1$). This difference is statistically significant, $t(57) = 10.47, p < .001$. Clearly, to further investigate this issue, experiment with $pq \neq 1$ should be performed.

5.3.4 Summary

In Experiment 1 employing integers and in Experiment 2 using fractions as ratio production factors, the axiom of 1-multiplicativity was found to be violated in about 30% of all tests. Therefore, the weaker axiom of k-multiplicativity was tested. Because the slope of each of the 19 individual linear regressions on each participant's triple of k values was equal to zero, the axiom was shown to hold for all participants. In Experiment 3 intermixing fractions and integers, the axiomatic tests revealed 1-multiplicativity to be violated in 55% of all tests and k-multiplicativity to be violated as well, because the computed $\overline{k_n}$ were not constant over different types of trials n.

5.4 Discussion and Implications

5.4.1 Axiomatic testing

The aim of this experiment was to investigate the relationship between perceived and mathematical numbers by testing the axioms of 1-multiplicativity and k-multiplicativity using data from three different duration production experiments.

In Experiment 1 and 2, the axiomatic testing procedure as proposed by Narens (1996) yielded the fundamental axioms of monotonicity and commutativity to hold for most participants. Thus, it may be assumed that participants operate on a sensory continuum and make quantitative statements on a ratio scale. However, testing the axiom of 1-multiplicativity revealed violations in about 30% of the comparisons and therefore, the assumption that our participants treat the numerical instructions used in duration production like mathematical numbers has to be rejected.

That is why, in the next step, the "weaker" axiom of k-multiplicativity, permitting a larger class of function form for W than 1-multiplicativity, formulated by Steingrimsson and Luce (2007) was tested. That k-multiplicativity holds means the relationship between numbers and their perception is well described by a power function with a constant exponent. Testing of k-multiplicativity yielded the axiom to hold for all participants in Experiment 1 and 2, i.e., a power relationship between mathematical and perceived number was shown to exist.

Furthermore, the fact that k-multiplicativity was valid in both Experiment 1 and 2 showed that this relationship holds for integers ($p \geq 1$, Experiment 1) as well as for fractions ($0 < p < 1$, Experiment 2). As proposed by Steingrimsson and Luce (2007) and others, the numerical distortion differs for integers and fractions, because for the vast majority of participants, the obtained values of k follow the predicted pattern of $k < 1$ for integers and $k > 1$ for fractions.

Experiment 3 was conducted to examine the differences in processing fractions and integers by intermixing them as ratio production factors and confirmed the assumption that the numerical distortion differs for $p \geq 1$ and $p < 1$. In addition to the axioms of monotonicity and commutativity being violated in a far larger number of tests than in the Experiments 1 and 2, both crucial axioms of 1-multiplicativity and k-multiplicativity failed to hold. Consequently, there is no single parametric instantiation describing the relationship between perceived and mathematical numbers from 0 to ∞. There are, however, two separate parametrizations of the investigated form: One for fractions and another one for integers.

5.4.2 Weighting Function

After determining the values of k for fractions and integers, one can specify the form of the weighting function $W(\mathbf{p})$ with greater certainty. The results of the present experiments confirm what was found in previous studies on other sensory modes (Ellermeier & Faulhammer, 2000; Steingrimsson & Luce, 2007; Zimmer, 2005), i.e, the assumptions of $W(\mathbf{p}) = p$ and $k = 1$ have to be rejected, because the axiom of 1-multiplicativity is violated, and thus, $W(\mathbf{p}) \neq p$.

By finding k to be constant for both fractions and integers, W can be further specified by considering the exact value of k: Because for integers, $k < 1$, one can assume $W(\mathbf{1}) < 1$, whereas for fractions, $k > 1$ and thus, $W(\mathbf{1}) > 1$ can be assumed. But even though the validity of k-multiplicativity is equivalent to a single functional form of W, the results reject the existence of a single parametric instantiation for numbers > 0 and rather propose different parameters for positive numbers < 1 and ≥ 1. This finding is in line with the theory of reference points (Luce et al., 2010) accounting for the discrepancy arising from using integers and fractions: The authors state that the axiom of commutativity in the mixed condition is predicted only if the reference points for $p < 1$ and $p \geq 1$ are equal. However, this proposition was rejected by their theory and data as well as by the results of the present study indicating different reference points for fractions and integers.

Other studies of the perception and processing of fractions propose numerators and denominators to be represented as separate integers instead of integrated representations (M. Schneider & Siegler, 2010). This finding might be a well-fitting explanation for the present results: The analyses of k-multiplicativity revealed "opposite" values of k for fractions and integers by using fractions like $\frac{1}{\mathbf{p}}$ with $\mathbf{p}$ being an integer between 2 and 8 like the $\mathbf{p}$ used in Experiment 1.

Moreover, the differential processing of fractions and integers can explain a bias observed in the majority of ratio production experiments: When integers are used as ratio production factors, the adjusted magnitudes of successive trials often exceed the adjustments of single trials (Augustin & Maier, 2008; Birkenbusch et al., 2015; Ellermeier & Faulhammer, 2000) whereas when using fractions, by contrast, the outcomes of successive trials fall below the outcomes of single trials (Steingrimsson & Luce, 2007; Zimmer, 2005).

5.4.3 Implications for Stevens' Power Law

An assumption implicit in Stevens (1954) concerning the application of direct scaling methods and formally stated by Narens (1996), namely that numerals as used in the experimental instructions (or by the participants) can be interpreted as mathematical numbers, may be considered rejected. When conducting a magnitude production or magnitude estimation experiment to determine the parameters of Stevens's power law, it is therefore difficult to interpret the "direct" scale values, because they do not necessarily correspond to the numbers they mathematically express.

Fortunately, the present re-analysis of a set of experiments yielded that the "numerical distortion" does not reflect an entirely arbitrary or indefinable interpretation of numbers, but a well-characterized mathematical relationship – a power function with a constant exponent. Knowing the size of k and the approximate form of the weighting function $W(\mathbf{p})$ does not allow to accurately determine the function relating mathematical and perceived numbers, even though this might be worthwile: An approach by Schneider B. Schneider et al.

(1974) suggests that the effect of non-veridical perception of numbers can be captured in the shape of the psychophysical power function. In contrast to that, Luce's model of global psychophysics assumes separable representations of the weighting function $W(\mathbf{p})$ and the psychophysical function $\psi(t)$, i.e., the perception of a stimulus' intensity $\psi(t)$ does not change because the task indirectly requires the processing of a number, $W(\mathbf{p})$. For example, the representation $W(\mathbf{p})\psi(t)$ implies that for an observer and a given number p, the function W is a number that multiplies the output of the psychophysical function $\psi(t)$ without directly changing it. If the psychophysical function is a power function, this multiplication only means changing the function's intercept.

Chapter 6

Manuscript C: Quantifying Subjective Duration: Both Power Function Exponents and Weber Fractions Vary With the Standard

Abstract

Since it is important for the results of psychophysical scaling experiments to remain invariant with changes of the standard, the dependency of magnitude productions of short auditory durations was studied as a function of standard duration. $N = 10$ participants were required to adjust the duration of a comparison tone to specific ratios $\mathbf{p}$ ($\times 2$, $\times 3$, and $\times 6$) of six different standard durations t ($0.1, 0.2, 0.3, 0.4, 0.5$ and 0.6 s). Furthermore, they completed an adaptive procedure to determine the size of the corresponding Weber fractions in order to rule out that the influence arises due to the method of ratio production. The results show a positive exponential relationship between the duration of the standard t and the estimated exponent of Stevens' power law β, which can be described by the function $\beta = 0.13t^{0.3}$ with an adjusted R^2 of 99%. Between the Weber fractions and the size of the standard, a negative exponential relationship of the form $W = 0.84t^{-0.3}$ with an adjusted R^2 of 88% was found. Thus, it

can be assumed that in the range of 0.1 and 0.6 s, the sensitivity of perceived duration increases with increasing standard. A bias due to the ratio production procedure was ruled out.

6.1 Introduction

Axiomatic analyses (Birkenbusch et al., 2015) found the exponent of Stevens' power law (1956) calculated for the perception of short durations to depend on the size of the standard used in the experiment. Therefore, the aim of the present study was to investigate the functional relationship between the standard duration and the size of the exponent and to check, whether this relationship is systematic. Furthermore, it was examined, whether the standard dependency is induced by the procedure of ratio production and thus, the results of the first experiment were validated by a second experiment applying an adaptive procedure to calculate Weber fractions, which may be interpreted, like the exponent of the power function, as a measure of sensitivity.

6.1.1 The Importance of Stevens' Power Law

Stevens' method of direct scaling (1956, 1975) is a very popular and widespread way to describe the relationship between a stimulus' physical intensity and its perceived magnitude and to compare this relationship across different sensory modalities. Indeed, Stevens provides a very easy and straightforward procedure to investigate the internal perceptual process by assuming that participants are able to directly describe the perceived magnitude of a stimulus. Especially in contrast to the Fechnarian approach using the indirect route via discriminability (e.g., Dzhafarov & Colonius, 1999), direct scaling provides several economic and conceptual advantages: It usually requires a smaller number of trials to determine the participant's sensitivity of a certain sensory modality. Furthermore, it uses ratios of perceived stimulus intensities rather than just noticeable differences (JNDs) as the perceptional basis for the estimated psychophysical function. The output of a perceptual process, i.e., the perceived magnitude is not measured in the units of the input, i.e., the underlying physical continuum – as is in indirect scaling – but in sensation units and thereby provides a more appropriate picture of the investigated sensory system (Gescheider, 1997).

Stevens' power law, formulated as $\varphi(t) = \alpha t^{\beta}, t > 0$ states that the perceived magnitude $\varphi(t)$ of a standard stimulus t is described by the power function αt^{β}, in which α is a proportionality factor depending on the physical units used to measure stimulus intensity and β is the gradient of the function depending on the sensory modality examined.

Typically, the data extracted from direct scaling experiments are used to determine the form of the power function: By fitting an exponential relationship between the stimulus intensities as independent and the assigned numbers as dependent variable, the parameters α and β can be determined. If then $\beta > 1$, the perceived magnitude of the stimulus grows disproportionally faster than the corresponding physical intensity, whereas $\beta < 1$ indicates that the increments in perceived stimulus magnitude become smaller with increasing physical intensity.

Furthermore, the exponent of the power function, β, can be used to compare the gradients of different sensory modalities to draw conclusions about their sensitivity (H. E. Ross, 1997; Ward et al., 1996), i.e., if the exponent of a sensory modality A is larger than the exponent of a modality B, it can be assumed that the observer is more sensitive to modality A than to modality B. In relation to indirect scaling methods, it was found that the ratio of two power function exponents β_A and β_B of two different sensory modalities A and B corresponds to the inverse ratio of the Weber fractions of these modalities W_A and W_B (R. Teghtsoonian, 2012; Ward, 1995).

For the perception of duration, numerous studies applying different scaling methods (Bobko et al., 1977; H. Eisler, 1975) determined the parameters of Stevens' power function. H. Eisler (1976) analyzed data resulting from 111 different experiments and found an average exponent $\beta = 0.90$ to best represent the relationship between physical and perceived duration.

6.1.2 The Influence of the Standard

Since the size of the exponent β of the power function plays a very important role in psychophysics, its determination should be as unambiguous and straightforward as possible. Nonetheless, Stevens' direct scaling are faced with a difficulty: It can be questioned, whether the estimated parameters are invariant under certain context effects (Gescheider, 1997; Poulton, 1989; M. Teghtsoonian & R. Teghtsoonian, 2003), such as, in this study, the size of the standard duration. For example, in a ratio production experiment on duration perception, the exponent computed for a series of adjustments based on the standard $t_1 = 100$ ms should not differ from the exponent computed for a standard of $t_2 = 400$ ms. If it does, the exponents thus determined are difficult to interpret or lack, as often stated, psychological "meaningfulness" (Luce, 1978; Narens, 1981; S. S. Stevens, 1946).

Augustin (2008) formulated two mathematical axioms, i.e., empirically testable conditions to check, whether the size of the exponent depends on the size of the size of the standard. These axioms are called *weak multiplicativity* and *invertibility.* For the perception of short durations, Birkenbusch et al. (2015) found both axioms to be violated by most participants. Furthermore, they computed the exponents of Stevens' power law for each of $N = 10$ participants and each of two standard durations ($t_1 = 100$, $t_2 = 400$ ms) and found them to differ significantly: Exponents based on the shorter standard duration were smaller ($\overline{\beta} = 0.87$) than exponents resulting from adjustments based on the longer standard ($\overline{\beta} = 1.02$). A further experiment with a standard $t_3 = 600$ ms revealed an exponent $\overline{\beta} = 1.16$.

Because these experiments differed in methodology and were based on different samples of participants, it was deemed necessary to explicitly investigate the influence of the standard on the size of the exponent by using different standard durations and establishing their relationship with the resulting exponents. Based on the data collected by Birkenbusch et al. (2015), a positive relationship between the size of the standard and the size of the exponent is assumed to exist.

Indeed, there is only one study on time perception directly addressing this question: Kane and Lown (1986) used two different standard durations of 30 and 180 s, but did not find the standard duration to affect the size of the power law exponent. Although in most of the other experiments on duration perception, different standards ranging from 50 ms to 3000 s were used and different exponents were found (H. Eisler, 1976), the direct influence of the size of the standard on the size of the exponent was not further investigated.

One might argue that the method of ratio production, in which the participants are asked to adjust the duration of a comparison stimulus until it corresponds to a certain ratio of the standard stimulus, is an inappropriate method to determine the size of the power function exponent. One might even assert that the standard-dependency is "caused" by this method. Therefore, a secondary goal of the present study was to check whether standard dependencies can be found for other indices of sensitivity such as psychophysical measures of discrimination. If so, it can be assumed that the effect is due to a change in sensitivity for different standards and not caused by the experimental method.

6.1.3 In Summary

Experiment 1 was conducted to check the assumption that the power function exponent grows with increasing standard duration. In this ratio production experiment, the participants had to adjust the duration of a comparison tone to a certain ratio to a standard duration. Six different standard durations were investigated ranging from 100 to 600 ms.

In Experiment 2, an adaptive procedure measuring discrimination "thresholds" (Kaernbach, 1991) was used to determine Weber fractions at the same standard durations and for the same participants. The task required them to indicate which of two successively presented durations was longer and converged to a discrimination threshold of 75%. In this case, decreasing Weber fractions were expected as the standard duration was increased.

6.2 Method

6.2.1 Participants

$N = 10$ students took part in the experiment. The sample consisted of 3 male and 7 female participants with a median age of 22.5 years ranging from 19 to 28. The testing was conducted individually in a double-walled sound-attenuated listening chamber (IAC).

6.2.2 Stimuli and Apparatus

Pure tones of a frequency of 440 Hz (A4 standard pitch) converted with 16 bits resolution and a sampling rate of 44.1 kHz were used as stimuli. To smooth onsets and offsets, the tones contained 10 ms cosine-shaped rise and decay ramps. In both experiments, the six different standard tones were set to 100, 200, 300, 400, 500 and 600 ms. The tones, which were generated in MATLAB, were adjusted to a fixed, comfortable sound pressure level of 65 dB SPL. After passing through a headphone amplifier (Behringer HA 800 Powerplay Pro 8), the tones were presented diotically via headphones (Beyerdynamics DT 990 Pro). The experimental program was generated in MATLAB 2012 with PsychToolbox-3 by Brainard (1997) and Pelli (1997).

6.2.3 Procedure

Experiment 1

In the first experiment, the participants had to complete a ratio production procedure in four identical test sessions. Each session was composed of three blocks of 18 trials, resulting in 45 trials per session and in a total of 216 trials altogether. Breaks of three minutes were programmed after the completion of each block. At the beginning of each session, the participants could become familiar with the ratio production task during three practice trials, after which the recording of the data started.

In each trial, two filled duration intervals in the form of continuous tones were presented in succession. The standard interval, which was presented first, was of fixed duration, i.e., of 100, 200, 300, 400, 500 or 600 ms. After a 500 ms silent inter-stimulus interval, the adjustable comparison interval was presented. Its initial duration was randomly chosen from the interval between 1 and 6 times as long as the preceding standard. During the presentation of both tones, a yellow numeral $\mathbf{p}$ ($\mathbf{p}$ = 2, 3, 6) was displayed on the screen. This numeral instructed the participant to adjust the duration of the second tone, so that it was perceived to be p-times as long as the duration of the first tone. The duration of the comparison tone could either be increased by pressing the right cursor key or decreased by pressing the left cursor key. Adjusting the duration was performed in discrete steps, i.e., one key press resulted in incrementing or decrementing the comparison by $\frac{1}{20}$ of the standard interval. To increase step size, participants could press the shift key together with the cursor key resulting in steps being ten times as long as the original. After each key press, the current standard and the altered comparison were replayed, and the participant was asked to adjust the comparison again, until he or she was satisfied with the result. The trial ended, when the participant registered the final adjusted duration by pressing the enter key. The next trial started after an inter-trial interval of 2000 ms. There was no time restriction to performing the task.

Each of the six different standard durations was combined with each of the three different ratio production factors. Each of these 18 combinations was presented once in a block and repeated 12 times during the entire experiment.

Experiment 2

In the second experiment, the participants had to complete a weighted up-down procedure (Kaernbach, 1991) in two different test sessions in order to measure just noticeable differences (JNDs) in duration. The procedure chosen accomplishes that by using asymmetric step sizes for "up" and "down" changes, thereby converging on the 75%-correct point of the psychometric function, i.e., that duration that can be distinguished from that of the standard 75% of the time. Each session consisted of six blocks of 60 trials each measuring one JND, resulting in 360 trials per session and 720 trials altogether. Breaks of three minutes were programmed after the completion of two blocks.

In each trial, the standard and the comparison interval were presented successively in random order. They were separated by a 500 ms silent inter-stimulus interval. The participants were asked to indicate, which of the tones was perceived to be longer: If the first tone was longer, they were asked to press "1", and if the second tone was longer, they were to press "2". Immediately afterwards, a visual feedback either saying "Correct!" or "False!" was presented on the screen for 1500 ms. The next trial started after an inter-trial interval of 2000 ms.

Due to the weighted up-down method, the response on the current trial determined the difficulty of the next trial: If the response was correct, the distance between the standard and the comparison tone in the next trial was decreased by a value of C, otherwise, if the response was incorrect, the distance was increased by a value of $3C$. The stepsize of C, by which the distance between standard and comparison was decreased or increased, depended on the size of the standard in the current block and the number of the trial: In trials

1 to 10, C was 6, 12, 18, 24, 30 or 36 ms for the standards of 100, 200, 300, 400, 500 and 600 ms. In trials 11 to 30, the C was 2, 4, 6, 8, 10 or 12 ms and in trials 31 to 60, C was 1, 2, 3, 4, 5 or 6 ms for the corresponding standards.

In the first session, the initial durations of the comparison interval were 50, 100, 150, 200, 250, and 300 ms, i.e., shorter than the standard. Thus, the duration of the comparison tone increased after a correct response. In the second session, the initial duration of the comparisons were 150, 300, 450, 600, 750, and 900 ms, i.e., longer than the standard. Therefore, after a correct response, the duration of the comparison tone was decreased.

Each of the six different standard durations was combined with both the ascending and descending comparison durations thus determining an "upward" and a "downward" JND from a given standard. The order of the standards in each session was randomly chosen.

6.3 Results

6.3.1 Experiment 1

The mean adjustments and standard errors for the different standards and the three different ratio production factors for $N = 10$ participants are shown in Figure 6.1. The mean number of adjustments made in one trial was $M = 9.85$. In 64% of the adjustments, small steps were used, whereas in the other 36%, large steps were used.

Linear regressions were computed for the six different standard durations to estimate the parameters a and b for a simple linear function of the form $\varphi(t) = a+bt$. It was assumed that the mean adjustment $\overline{x}$ of the individual single adjustments with a fixed standard t and a ratio production factor $\mathbf{p}$ is perceived to be $\mathbf{p}$ times as long as the standard. Thus, for a linear "psychophysical" model, a linear regression between the ratio production factor $\mathbf{p}$ constituting the dependent variable and the mean adjustment $\overline{x}$ constituting the independent variable was computed.

Furthermore, the parameters α and β for the power law of the form $\varphi(t) = \alpha t^{\beta}$ were computed. Analogously to calculating the parameters of the simple linear function, for the power law, a linear regression was computed with the

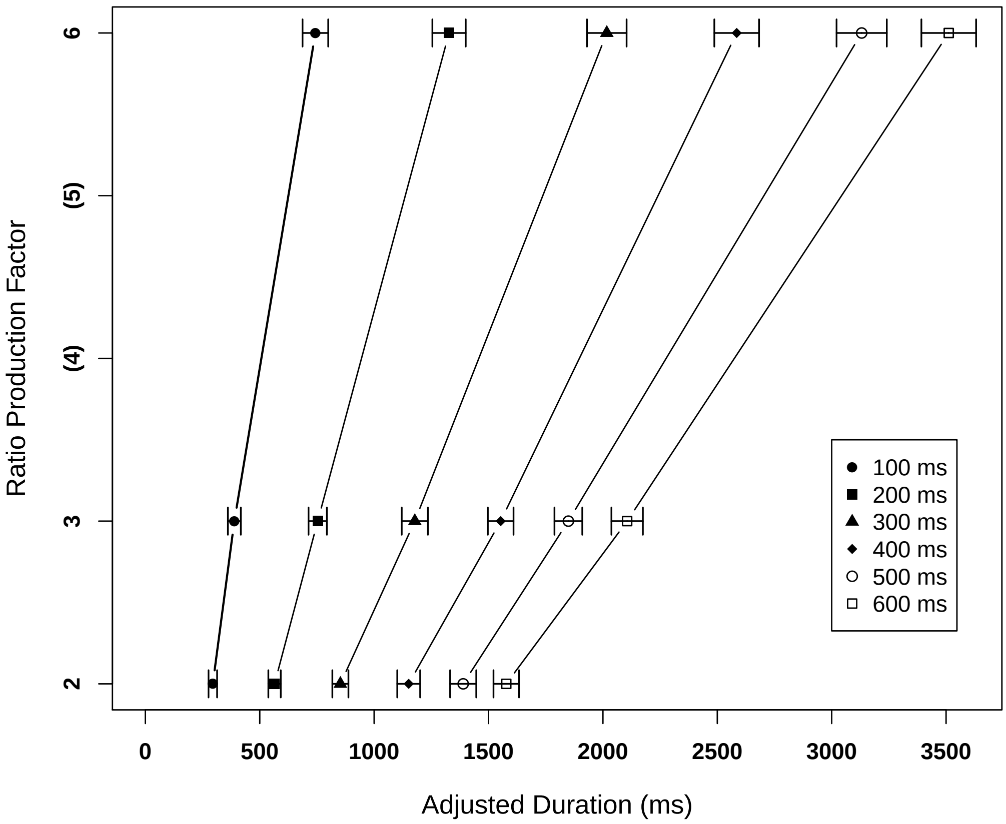

Figure 6.1: Means and standard deviations of adjusted duration in the basic trials for the six different standards and the three different ratio production factors.

Table 6.1

Experiment 1: Estimated parameters and squared correlation coefficients for linear model and power function for the six different standards.

	Linear Model			Power Function		
Standard	a	b	R^2	$ln(\alpha)$	β	R^2
100	1.24	1.59	0.35	0.47	0.54	0.41
200	0.91	1.23	0.47	0.33	0.66	0.47
300	0.68	0.98	0.57	0.20	0.76	0.56
400	0.51	0.85	0.60	0.09	0.84	0.60
500	0.45	0.73	0.64	0.01	0.87	0.62
600	0.39	0.67	0.65	-0.06	0.91	0.63

logarithm of the ratio production factor **p** as the dependent variable and the logarithm of the mean adjustments $\overline{x}$ serving as the independent variable. The estimated parameters and squared correlation coefficients R^2 for both linear and power function model are shown in Table 6.1. For the liner model, the parameters b computed for the different standards vary between 0.67 and 1.59, the slope estimated a vary between 0.39 and 1.25. Both parameters decrease with increasing standard duration. For Stevens' power law, the exponents β vary between 0.54 and 0.91 and the parameters α vary between -0.06 and 0.47. Parameters α decrease with increasing standard duration, whereas parameters β increase with the standard.

A linear model with the standard duration t as the independent and the estimated exponent β as the dependent variable reveals a positive linear relationship of the form $\beta = 0.75t + 0.5$ with an adjusted R^2 of 94%. An exponential model using the log-transformed standard duration as the independent and the log-transformed exponents β as the dependent variable reveals a function of the form $\beta = 0.13t^{0.3}$ with a slightly better fit of R^2 of 99%.

The size of the estimated exponent as a function of the duration of the standard, the exponential model for estimating the exponents as well as the confidence interval for this model are depicted in Figure 6.2. For the parameter α, a linear function of the form $\alpha = -1.06t + 0.54$ with an adjusted R^2 of 97% was found.

Furthermore, individual power function models were computed for each

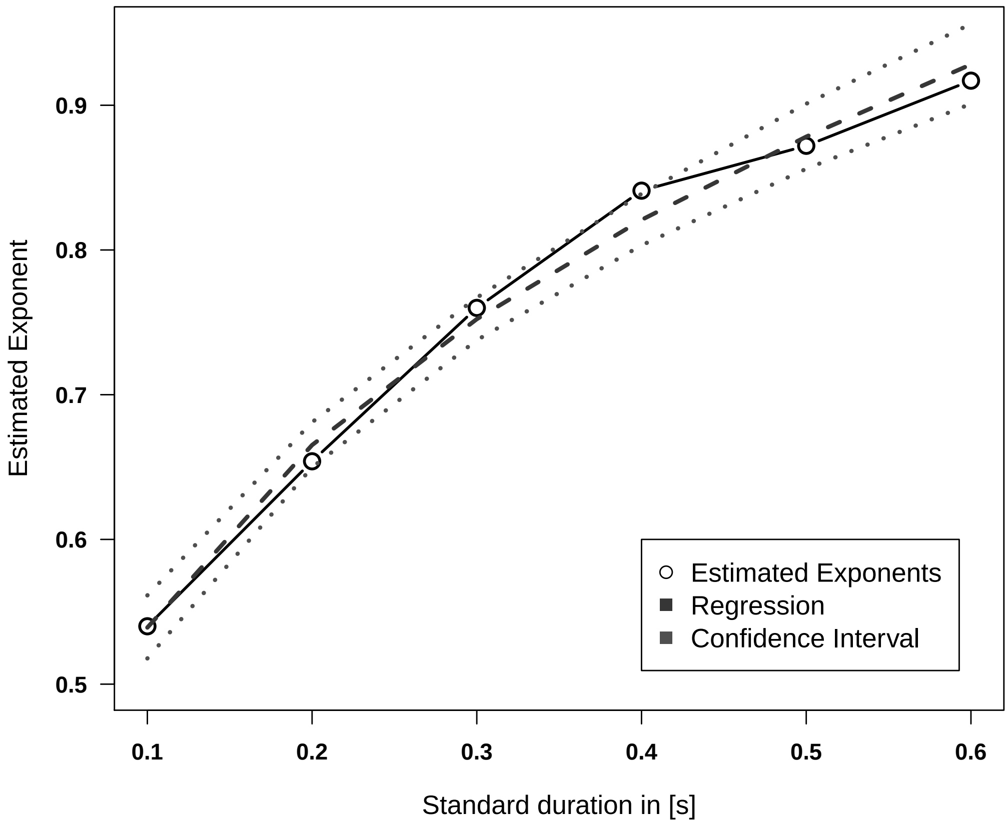

Figure 6.2: Size of the exponent of Stevens' Power Law as a function of the size of the standard. The black line shows the computed exponents for the entire sample, the gray lines show the exponential function to estimate the β-exponents and the confidence interval of the estimation.

participant and each standard. Regressions with the standard durations t as independent and the individually estimated exponents β as dependent variables revealed positive exponential relationships with a statistically significant gradient (slope) for 7 of 10 participants ($p < .001$ for sz1312, $p < .01$ for fl2909, lb2607, ls0907, ls1403, and lw2807, and $p < .1$ for kf1507), while for the other participants, the exponent β did not depend on the size of the standard (all $p > .1$).

6.3.2 Experiment 2

The Weber fractions were calculated by first estimating the difference threshold (JND) across all $N = 10$ participants. It was calculated as the difference between the average of the durations of the comparison interval in the last 10 trials and the standard, i.e., by subtracting the average comparison from the standard in ascending trials (first session) and by subtracting the standard from the average comparison in the descending trials (second session) (Grondin, Ouellet, & Roussel, 2001). The ratio of the difference threshold on the standard constitutes the Weber fraction.

The Weber fractions resulting from the ascending and descending discrimination series did not differ significantly ($t = .73, df = 8.5, p = .48$). Thus, the mean Weber fractions are reported in the following.

The difference thresholds estimated for the six different standard durations are 13.83 ms, 22.33 ms, 26.88 ms, 33.90 ms, 43.05 ms, and 51.78 ms. According to these values, the Weber fractions were 13.83%, 11.17%, 8.96%, 8.48%, 8.61%, and 8.63%.

A linear model with the standard duration t as the independent and the estimated Weber fraction W as the dependent variable reveals a positive linear relationship of the form $W = -9,76t + 13.36$ with an adjusted R^2 of 65%. An exponential model using the logarithmically transformed standard duration as the independent and the logarithmically transformed Weber fraction W as dependent variable reveals a function of the form $W = 0.84t^{-0.3}$ with a clearly better fit of R^2 of 88%.

The Weber fractions, the values for ascending and descending JNDs, as well as the exponential regression and its confidence interval are shown in Figure 6.3. Due to the small number of threshold estimated per participant, an individual analysis of the relationship between Weber fractions and standard durations was not conducted.

A negative correlation of $r = -0.94$ ($p = .004$) was found between the Weber fraction and the power function exponent.

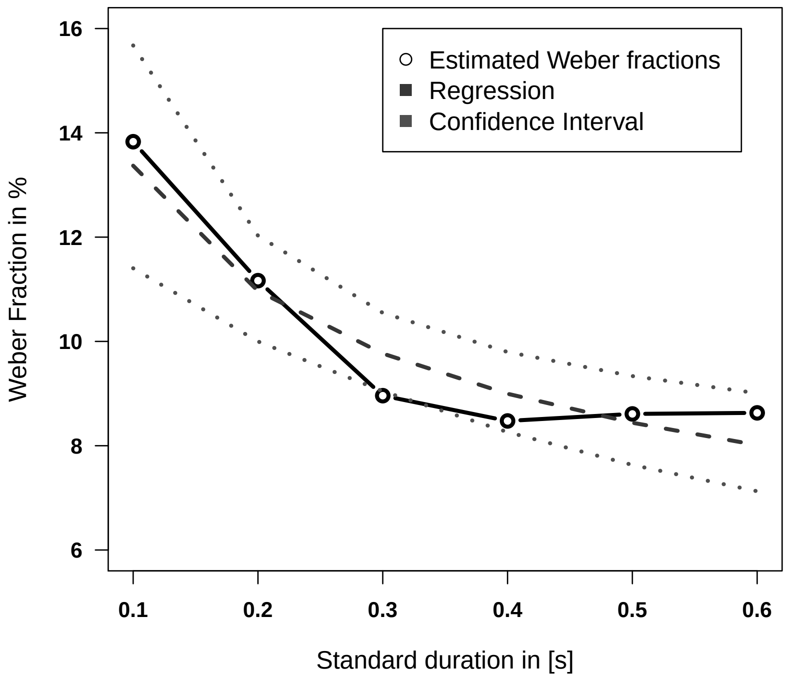

Figure 6.3: Size of the Weber fraction as a function of the size of the standard. The black line shows the computed Weber fraction for the entire sample, the gray lines show the exponential function to estimate the Weber fractions and the confidence interval of the estimation.

6.4 General Discussion

For the perception of short durations, the exponent of Stevens' power law estimated in a ratio production experiment depends on the size of the standard stimulus. The aim of this study was to determine the exact functional relationship between the size of the exponent and the size of the standard. Thus, six different standard durations between 100 and 600 ms were investigated. A further aim of the study was to validate this result by means of an alternative measure of sensitivity, i.e., by the Weber fractions determined in an adaptive (weighted up-down) prodecure determining JNDs at exactly the same standard durations.

6.4.1 Power Law

To estimate the exponents of the psychophysical function for each of the standard durations, six separate exponential regressions were computed on the overall data. These analyses revealed increasing exponents for increasing standards, which was confirmed by conducting a regression on the estimated exponents as a function of the standard size indicating an exponential relationship with a satisfactory fit. However, the slope of this function is steeper between 100 and 400 ms and decreases between 400 and 600 ms.

The individual analyses showed the same pattern, i.e., increasing exponents with increasing standards for seven of ten participants, whereas for the other three participants, a constant exponent was found. Thus, it can be assumed some participants' duration perception is capable of being influenced by the standard size, whereas for other participants it is not.

These current results mostly fit to observations made in other studies, e.g., Treisman (1963) found similar power law exponents of $\beta = 0.89$ (and Weber fractions of $W = 6.1\%$).

Because greater power function exponents are often associated with a higher differential sensitivity (R. Teghtsoonian, 2012; Ward et al., 1996), one might interpret the functional relationship as an increase in sensitivity with increasing standard duration.

6.4.2 Weber Fractions

To estimate the Weber fractions for each of the six standard durations, the difference thresholds (JNDs) were calculated by subtracting from the corresponding standard. Dividing the JND by the respective standard durations, the Weber fractions obtained were found to decrease with increasing standard duration. This observation was confirmed by conducting a regression on the estimated Weber fractions as a function of the standard indicating a well-fitting exponential relationship. Particularly between 100 and 300 ms, the slope of the function in very steep, whereas between 300 and 600 ms, the function converges to an asymptote of $W = 9\%$.

Results comparable to those of the current study are reported in Getty (1975) finding high Weber fractions (13 %) for short standards up to 200 ms and a tendency to remain constant at a level of about 7% for standards beyond 200 ms. Constant Weber fractions for standards between 400 and 1200 ms implying a validity of Weber's law in this stimulus range were reported as well (Ehrlé & Samson, 2005; Grondin, 2001, 2010). An investigation using the same weighted up-down procedure as in the present study showed that the Weber fractions remain constant for standard durations between 400 and 600 ms (Grondin, Ouellet, & Roussel, 2001).

Since low Weber fractions reflect a high differential sensitivity, interpreting the course of the function as an increase in sensitivity with increasing standard duration appears plausible. However, the generalized form of Weber's law assuming a constant sensory noise interfering with the process of duration perception predicts higher Weber fractions for short standard durations: The noise as a duration-independent source of timing variability especially influences short standard durations whereas longer durations remain unaffected (Getty, 1975; Rammsayer, 2010a).

6.4.3 Implications

Several conclusions can be drawn from these results. First of all, there are consequences for handling psychophysical functions. Estimating power function exponents for duration perception requires reference to the duration of the standard, because the power law parameters are not invariant under different standards. If the standard duration is not specified, the interpretation of the exponent, e.g., for comparing different sensory modalities, remains difficult.

Furthermore, treating the standard dependency of the exponent an inconvenient bias specific to the method of ratio production – or to methods using a standard stimulus in general – can be precluded. The parallel analyses of power function exponents and Weber fractions for the same participants showed compatible results, i.e., a strong standard-dependency between 100 and 300 to 400 ms and a plateau beyond about 400 ms. The interpretation of both indices implies an increasing sensitivity with increasing standard for durations shorter than 400 ms and a largely constant sensitivity for longer standard durations.

Moreover, the results might be relevant for the debate about the relationship between Stevens' power law and Weber fractions. Although it was qualitatively argued that there should not be a correlation between the Weber fractions as a measure of resolving power and the power function exponent as a measure of sensation magnitudes (Laming, 1997; S. S. Stevens, 1961), R. Teghtsoonian (1971) found a negative Spearman rank correlation coefficient of $r = -0.44$ between Weber fractions and power law exponents when aggregating data across several sensory modalities. The results of the current study support the finding of a negative correlation rather than assuming no systematic relationship.

The present results can be related to findings on scalar timing properties: According to the conformity of scalar expectancy theory (SET), the timing behavior is required to exhibit two properties: *Mean accuracy* requires the means of the adjusted durations to vary linearly with increasing standards, whereas the *scalar property of variance* necessitates timing sensitivity to remain constant with increasing standard (Wearden & Lejeune, 2008). For Experiment 1, scalar timing properties seem to hold, because the *coefficient of variation* (CV = SD/M) stays constant for each standard duration. For Experiment 2, one might assume scalar properties not to be valid, because the Weber fraction decreases up to 300 ms and remains constant beyond 300 ms. These results contrast the findings of Wearden and Lejeune (2008), who found scalar properties to hold for most discrimination tasks, but not for "classical" timing tasks such as interval production, temporal reduction and verbal estimation. Grondin (2010) found the Weber fractions for the standards of 200 and 1000 ms not to conform to the requirements of scalar properties, but found lower Weber fractions for the shorter standard. Apparently, this issue has to be investigated in more detail.

6.4.4 Limitations

To examine whether the current finding is generalizable across different sensory modalities, it should be investigated, whether the relationship between standard duration and exponent revealed in this study can be found for other modalities, as well. If the detected pattern, i.e., an increasing power law exponent with increasing standard could be found for multiple perceptual continua, the gradients of the determined functions might offer a possibility to compare the modalities and their sensitivities with each other. This might be valuable, since comparisons between sensory modalities using empirical power functions exponents - even if the standards employed are denoted - cannot be conducted without an element of arbitrariness, because there is no plausible way to find a comparable standard size in the other modality.

Further research might investigate whether the influence of the standard size on the exponent interacts with other context effects, e.g., the range of the standard stimuli (Ward et al., 1996) or the numbers assigned to them (Beck & Shaw, 1965), which were examined in previous studies.

6.4.5 Conclusion

The fundamental outcome of this experiment is the determination of the functional relationship between the size of the standard used in a ratio production experiment and the power function exponent derived from the participants' adjustments: Increasing standard durations ranging from 100 to 600 ms revealed increasing exponents. This result confirms former axiomatic investigations stating the size of the exponent to depend on the size of the standard. To rule out that the influence is due to the ratio production method, an adaptive procedure was used to determine JNDs and subsequently Weber fractions for the six different standard durations. The Weber fractions decreased with increasing standard durations and thus, combining both findings, it can be assumed that the sensitivity for duration perception increases between 100 and 400 ms and remains at a constant level between 400 and 600 ms.

Bibliography

Abel, S. M. (1972a). Discrimination of temporal gaps. *The Journal of the Acoustical Society of America*, *52*(2B), 519–524.
DOI: 10.1121/1.1913139

Abel, S. M. (1972b). Duration discrimination of noise and tone bursts. *The Journal of the Acoustical Society of America*, *51*(4B), 1219–1223.
DOI: 10.1121/1.1912963

Allan, L. G. (1977). The time-order error in judgments of duration. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *31*(1), 24–31.
DOI: 10.1037/h0081647

Allan, L. G. (1979). The perception of time. *Perception & Psychophysics*, *26*(5), 340–354.
DOI: 10.3758/BF03204158

Allan, L. G., Kristofferson, A. B., & Wiens, E. W. (1971). Duration discrimination of brief light flashes. *Perception & Psychophysics*, *9*(3), 327–334.
DOI: 10.3758/BF03212659

Anderson, N. H. (1970). Functional measurement and psychophysical judgment. *Psychological Review*, *77*(3), 153–170.
DOI: 10.1037/h0029064

Angrilli, A., Cherubini, P., Pavese, A., & Manfredini, S. (1997). The influence of affective factors on time perception. *Perception & Psychophysics*, *59*(6), 972–982.
DOI: 10.3758/BF03205512

Attneave, F. (1962). Perception and related areas. In S. Koch (Ed.), *Psychology: A study of science, Volume 4. Biologically oriented fields: Their place in psychology and in biological science* (pp. 619–659). New York, NY:

McGraw-Hill.
DOI: 10.1037/10541-013

Augustin, T. (2008). Stevens' power law and the problem of meaningfulness. *Acta Psychologica*, *128*(1), 176–185.
DOI: 10.1016/j.actpsy.2007.12.005

Augustin, T. (2010). Extensions of Narens' theory of ratio magnitude estimation. *British Journal of Mathematical and Statistical Psychology*, *63*(3), 647–663.
DOI: 10.1348/000711010X486075

Augustin, T. & Maier, K. (2008). Empirical evaluation of the axioms of multiplicativity, commutativity, and monotonicity in ratio production of area. *Acta Psychologica*, *129*(1), 208–216.
DOI: 10.1016/j.actpsy.2008.06.004

Banks, W. P. & Coleman, M. J. (1981). Two subjective scales of number. *Perception & Psychophysics*, *29*(2), 95–105.
DOI: 10.3758/BF03207272

Banks, W. P. & Hill, D. K. (1974). The apparent magnitude of number scaled by random production. *Journal of Experimental Psychology*, *102*(2), 353.
DOI: 10.1037/h0035850

Beck, J. & Shaw, W. A. (1965). Magnitude of the standard, numerical value of the standard, and stimulus spacing in the estimation of loudness. *Perceptual and Motor Skills*, *21*(1), 151–156.
DOI: 10.2466/pms.1965.21.1.151

Behar, I. & Bevan, W. (1961). The perceived duration of auditory and visual intervals: Cross-modal comparison and interaction. *The American Journal of Psychology*, *74*(1), 17–26.
DOI: 10.2307/1419819

Bernoulli, D. (1738). Specimen theoriae novae de mensura sortis. *Commentarii Academiae Scientarium Imperiales Petropolitanae*, *5*, 175–192.

Birkenbusch, J., Ellermeier, W., & Kattner, F. (2015). Octuplicate this interval! Axiomatic examination of the ratio properties of duration perception. *Attention, Perception & Psychophysics*, *77*(5), 1767–1780.
DOI: 10.3758/s13414-015-0846-0

Birnbaum, M. H. (1982). Controversies in psychophysical measurement. In B. Wegener (Ed.), *Social attitudes and psychophysical measurement* (pp. 401–485). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bisson, N., Tobin, S., & Grondin, S. (2008). Remembering the duration of joyful and sad musical excerpts: Assessment with three estimation methods. *NeuroQuantology*, *7*(1).
DOI: 10.14704/nq.2009.7.1.206

Björkman, M. & Holmkvist, O. (1960). The time-order error in the construction of a subjective time scale. *Skandinavian Journal of Psychology*, *1*, 7–13.
DOI: 10.1111/j.1467-9450.1960.tb01275.x

Block, R. A. & Zakay, D. (1997). Prospective and retrospective duration judgments: A meta-analytic review. *Psychonomic Bulletin & Review*, *4*(2), 184–197.
DOI: 10.3758/BF03209393

Bobko, D. J., Thompson, J. G., & Schiffman, H. R. (1977). The perception of brief temporal intervals: Power functions for auditory and visual stimulus intervals. *Perception*, *6*(6), 703–709.
DOI: 10.1068/p060703

Bonato, M., Fabbri, S., Umiltà, C., & Zorzi, M. (2007). The mental representation of numerical fractions: Real or integer? *Journal of Experimental Psychology: Human Perception and Performance*, *33*(6), 1410–1419.
DOI: 10.1037/0096-1523.33.6.1410

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436.
DOI: 10.1163/156856897X00357

Brown, D. R. & Hitchcock, L. (1965). Time estimation: Dependence and independence of modality-specific effects. *Perceptual and Motor Skills*, *21*(3), 727–734.
DOI: 10.2466/pms.1965.21.3.727

Brown, S. W. (1985). Time perception and attention: The effects of prospective versus retrospective paradigms and task demands on perceived duration. *Perception, & Psychophysics*, *38*(2), 115–124.
DOI: 10.3758/BF03198848

Brown, S. W. & Boltz, M. G. (2002). Attentional processes in time perception: Effects of mental workload and event structure. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(3), 600–615. DOI: 10.1037/0096-1523.28.3.600

Brown, S. W. & Stubbs, D. A. (1992). Attention and interference in prospective and retrospective timing. *PERCEPTION-LONDON*, *21*, 545–545. DOI: 10.1068/p210545

Buhusi, C. V. & Meck, W. H. (2009). Relative time sharing: New findings and an extension of the resource allocation model of temporal processing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1525), 1875–1885. DOI: 10.1098/rstb.2009.0022

Churcher, B. G. (1935). A loudness scale for industrial noise measurements. *The Journal of the Acoustical Society of America*, *6*(4), 216–225. DOI: 10.1121/1.1915740

Creelman, C. D. (1962). Human discrimination of auditory duration. *The Journal of the Acoustical Society of America*, *34*(5), 582–593. DOI: 10.1121/1.1918172

Curtis, D. W., Attneave, F., & Harrington, T. L. (1968). A test of a two-stage model of magnitude judgment. *Perception & Psychophysics*, *3*(1), 25–31. DOI: 10.3758/BF03212707

Curtis, D. W. & Fox, B. E. (1969). Direct quantitative judgments of sums and a two-stage model for psychophysical judgments. *Perception & Psychophysics*, *5*(2), 89–93. DOI: 10.3758/BF03210528

Divenyi, P. L. & Danner, W. F. (1977). Discrimination of time intervals marked by brief acoustic pulses of various intensities and spectra. *Perception & Psychophysics*, *21*(2), 125–142. DOI: 10.3758/BF03198716

Drake, C. & Botte, M.-C. (1993). Tempo sensitivity in auditory sequences: Evidence for a multiple-look model. *Perception & Psychophysics*, *54*(3), 277–286. DOI: 10.3758/BF03205262

Droit-Volet, S. & Gil, S. (2009). The time–emotion paradox. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1525), 1943–1953.
DOI: 10.1098/rstb.2009.0013

Dzhafarov, E. N. & Colonius, H. (1999). Fechnerian metrics in unidimensional and multidimensional stimulus spaces. *Psychonomic Bulletin & Review*, *6*(2), 239–268.
DOI: 10.3758/BF03212329

Ehrlé, N. & Samson, S. (2005). Auditory discrimination of anisochrony: Influence of the tempo and musical backgrounds of listeners. *Brain and Cognition*, *58*(1), 133–147.
DOI: 10.1016/j.bandc.2004.09.014

Eisler, A. D. (1995). Cross-cultural differences in time perception: Comparison of African immigrants and native Swedes. *Perception and Psychophysics in Theory and Application*, 137–145.

Eisler, A. D., Eisler, H., & Montgomery, H. (2007). A quantitative model for retrospective subjective duration. *NeuroQuantology*, *2*(4), 263–291.
DOI: 10.14704/nq.2004.2.4.51

Eisler, A. D., Eisler, H., & Mori, S. (2001). Time perception: Comparison of schizophrenic and healthy Japanese males. In *Abstract Book of the VIIth European Congress of Psychology, The British Psychological Society, London* (p. 120).

Eisler, H. (1975). Subjective duration and psychophysics. *Psychological Review*, *82*(6), 429–450.
DOI: 10.1037/0033-295X.82.6.429

Eisler, H. (1976). Experiments on subjective duration 1868-1975: A collection of power function exponents. *Psychological Bulletin*, *83*(6), 1154–1171.
DOI: 10.1037/0033-2909.83.6.1154

Eisler, H. & Eisler, A. D. (1992). Time perception: Effects of sex and sound intensity on scales of subjective duration. *Scandinavian Journal of Psychology*, *33*(4), 339–358.
DOI: 10.1111/j.1467-9450.1992.tb00923.x

Eisler, H., Eisler, A. D., & Hellström, Å. (2008). Psychophysical issues in the study of time perception. In S. Grondin (Ed.), (pp. 75–109). Bingley, UK: Emerald Group Publishing.

Ekman, G. (1958). Two generalized ratio scaling methods. *The Journal of Psychology*, *45*(2), 287–295.
DOI: 10.1080/00223980.1958.9916259

Ekman, G. & Hosman, B. (1965). Note on subjective scales of number. *Perceptual and Motor Skills*, *21*(1), 101–102.
DOI: 10.2466/pms.1965.21.1.101

Ellermeier, W. & Faulhammer, G. (2000). Empirical evaluation of axioms fundamental to Stevens's ratio-scaling approach: I. Loudness production. *Perception & Psychophysics*, *62*(8), 1505–1511.
DOI: 10.3758/BF03212151

Engen, T. & Tulunay, U. (1957). Some sources of error in half-heaviness judgments. *Journal of Experimental Psychology*, *54*(3), 208–212.
DOI: 10.1037/h0044868

Estel, V. (1885). Neue Versuche über den Zeitsinn (New studies on the sense of time). *Philosophische Studien*, (2), 37–65.

Fechner, G. T. (1860). *Elemente der Psychophysik (Elements of psychophysics).* Leipzig, Germany: Breitkopf und Härtel.

Fishburn, P. C. (1988). *Utility theory.* Wiley Online Library.

Fletcher, H. & Munson, W. A. (1933). Loudness, its definition, measurement and calculation. *Bell System Technical Journal*, *12*(4), 377–430.
DOI: 10.1002/j.1538-7305.1933.tb00403.x

Fraisse, P. (1963). *The psychology of time.* Oxford, UK: Harper & Row.

Fraisse, P. (1967). *Psychologie du temps (Psychology of time).* Paris, France: Presses Universitaires de France.

Fraisse, P. (1978). Time and rhythm perception. *Handbook of perception: Vol. 8. Perceptual coding*, *8*, 203–254.

Fraisse, P. (1984). Perception and estimation of time. *Annual Review of Psychology*, *35*(1), 1–36.
DOI: 10.1146/annurev.ps.35.020184.000245

Frankenhaeuser, M. (1960). Subjective time as affected by gravitational stress. *Skandinavian Journal of Psychology*, *1*, 1–6. DOI: 10.1111/j.1467-9450.1960.tb01274.x

Friberg, A. & Sundberg, J. (1995). Time discrimination in a monotonic, isochronous sequence. *The Journal of the Acoustical Society of America*, *98*(5), 2524–2531. DOI: 10.1121/1.413218

Fullerton, G. S. & Cattell, J. M. (1892). *On the perception of small differences.*

Ganor-Stern, D. (2012). Fractions but not negative numbers are represented on the mental number line. *Acta Psychologica*, *139*(2), 350–357. DOI: 10.1016/j.actpsy.2011.11.008

Garner, W. R. (1954). Context effects and the validity of loudness scales. *Journal of Experimental Psychology*, *48*(3), 218–224. DOI: 10.1037/h0061514

Gescheider, G. A. (1997). *Psychophysics: The Fundamentals.* Mahwah, NJ: Lawrence Erlbaum Associates.

Getty, D. J. (1975). Discrimination of short temporal intervals: A comparison of two models. *Perception & Psychophysics*, *18*(1), 1–8. DOI: 10.3758/BF03199358

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, *84*(3), 279–325. DOI: 10.1037/0033-295X.84.3.279

Goldstone, S. & Goldfarb, J. L. (1964). Direct comparisons of auditory and visual durations. *Journal of Experimental Psychology*, *67*(5), 483–485. DOI: 10.1037/h0046997

Goldstone, S. & Lhamon, W. T. (1974). Studies of auditory-visual differences in human time judgment: 1. Sounds are judged longer than lights. *Perceptual and Motor Skills*, *39*(1), 63–82. DOI: /10.2466/pms.1974.39.1.63

Goude, G. (1962). *On fundamental measurement in psychology.* Oxford, UK: Almqvist & Wirksell.

Grondin, S. (1993). Duration discrimination of empty and filled intervals marked by auditory and visual signals. *Perception & Psychophysics*, *54*(3), 383–

394.
DOI: 10.3758/BF03205274

Grondin, S. (2001). Discriminating time intervals presented in sequences marked by visual signals. *Perception & Psychophysics*, *63*(7), 1214–1228.
DOI: 10.3758/BF03194535

Grondin, S. (2008). *The psychology of time*. Bingley, UK: Emerald Group Publishing.

Grondin, S. (2010). Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions. *Attention, Perception, & Psychophysics*, *72*(3), 561–582.
DOI: 10.3758/APP.72.3.561

Grondin, S., Meilleur-Wells, G., & Lachance, R. (1999). When to start explicit counting in a time-intervals discrimination task: A critical point in the timing process of humans. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(4), 993–1004.
DOI: 10.1037/0096-1523.25.4.993

Grondin, S., Ouellet, B., & Roussel, M. (2001). About optimal timing and stability of Weber fraction for duration discrimination. *Acoustical Science and Technology*, *22*(5), 370–372.
DOI: 10.1250/ast.22.370

Guilford, J. P. (1932). A generalized psychophysical law. *Psychological Review*, *39*(1), 73–85.
DOI: 10.1037/h0070969

Haldemann, J., Stauffer, C., Troche, S., & Rammsayer, T. H. (2012). Performance on auditory and visual temporal information processing is related to psychometric intelligence. *Personality and Individual Differences*, *52*(1), 9–14.
DOI: 10.1016/j.paid.2011.08.032

Hellman, R. P. & Zwislocki, J. J. (1968). Loudness determination at low sound frequencies. *The Journal of the Acoustical Society of America*, *43*(1), 60–64.
DOI: 10.1121/1.1910763

Hellström, Å. (1985). The time-order error and its relatives: Mirrors of cognitive processes in comparing. *Psychological Bulletin*, *97*(1), 35–61. DOI: 10.1037/0033-2909.97.1.35

Hellström, Å. & Rammsayer, T. H. (2004). Effects of time-order, interstimulus interval, and feedback in duration discrimination of noise bursts in the 50- and 1000-ms ranges. *Acta Psychologica*, *116*(1), 1–20. DOI: 10.1016/j.actpsy.2003.11.003

Henry, F. M. (1948). Discrimination of the duration of a sound. *Journal of Experimental Psychology*, *38*(6), 734–743. DOI: 10.1037/h0058552

Holway, A. H. & Pratt, C. C. (1936). The Weber ratio for intensive discrimination. *Psychological Review*, *43*(4), 322–340. DOI: 10.1037/h0059748

Jamieson, D. G. & Petrusic, W. M. (1975). The dependence of time-order error direction on stimulus range. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *29*(3), 175–182. DOI: 10.1037/h0082023

Jones, F. N. (1974). Overview of psychophysical scaling methods. In E. C. Carterette & M. P. Friedman (Eds.), *Psychophysical Judgment and Measurement* (pp. 343–360). New York, NY: Academic Press.

Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Attention, Perception, & Psychophysics*, *49*(3), 227–229. DOI: 10.3758/BF03214307

Kane, L. S. & Lown, B. A. (1986). Stevens' power law and time perception: Effect of filled intervals, duration of the standard, and number of presentations of the standard. *Perceptual and Motor Skills*, *62*(1), 35–38. DOI: 10.2466/pms.1986.62.1.35

Kattner, F. & Ellermeier, W. (2014). Fractionation of pitch intervals: An axiomatic study testing monotonicity, commutativity, and multiplicativity in musicians and non-musicians. *Attention, Perception, & Psychophysics*, *76*(8), 2508–2521. DOI: 10.3758/s13414-014-0674-7

Killeen, P. R. & Weiss, N. A. (1987). Optimal timing and the Weber function. *Psychological Review*, *94*(4), 455–468.
DOI: 10.1037/0033-295X.94.4.455

Kinchla, J. (1972). Duration discrimination of acoustically defined intervals in the 1- to 8-sec range. *Perception & Psychophysics*, *12*(3), 318–320.
DOI: 10.3758/BF03207213

Kornbrot, D. E., Msetfi, R. M., & Grimwood, M. J. (2013). Time perception and depressive realism: Judgment type, psychophysical functions and bias. *PloS one*, *8*(8).
DOI: 10.1371/journal.pone.0071585

Kristofferson, A. B. (1977). A real-time criterion theory of duration discrimination. *Perception & Psychophysics*, *21*(2), 105–117.
DOI: 10.3758/BF03198714

Kristofferson, A. B. (1980). A quantal step function in duration discrimination. *Perception & Psychophysics*, *27*(4), 300–306.
DOI: 10.3758/BF03206118

Laming, D. R. (1997). *The measurement of sensation*. New York, NY: Oxford University Press.

Lockhead, G. R. (1992). Psychophysical scaling: Judgments of attributes or objects? *Behavioral and Brain Sciences*, *15*(3), 543–558.
DOI: 10.1017/S0140525X00069934

Loeb, M., Behar, I., & Warm, J. S. (1966). Cross-modal correlations of the perceived durations of auditory and visual stimuli. *Psychonomic Science*, *6*(2), 87–87.
DOI: 10.3758/BF03327970

Long, G. M. & Beaton, R. J. (1980). The contribution of visual persistence to the perceived duration of brief targets. *Perception & Psychophysics*, *28*(5), 422–430.
DOI: 10.1037/h0043178

Luce, R. D. (1959). On the possible psychophysical laws. *Psychological Review*, *66*(2), 81.

Luce, R. D. (1978). Dimensionally invariant numerical laws correspond to meaningful qualitative relations. *Philosophy of Science*, 1–16.
DOI: 10.2307/186895

Luce, R. D. (2001). Reduction invariance and Prelec's weighting functions. *Journal of Mathematical Psychology*, *45*(1), 167–179.
DOI: 10.1006/jmps.1999.1301

Luce, R. D. (2002). A psychophysical theory of intensity proportions, joint presentations, and matches. *Psychological Review*, *109*(3), 520–532.
DOI: 10.1037/0033-295X.109.3.520

Luce, R. D. (2008). Symmetric and asymmetric matching of joint presentations: Correction to Luce (2004). *Psychological Review*, *115*(3), 601.
DOI: 10.1037/0033-295X.115.3.601

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*(4), 1247–1258.
DOI: 10.1037/a0020174

Marks, L. E. & Gescheider, G. A. (2002). *Psychophysical scaling. Stevens' Handbook of Experimental Psychology.* Wiley Online Library.
DOI: 10.1002/0471214426.pas0403

Mehner, M. (1885). Die Lehre vom Zeitsinn (The sense of time). *Philosophische Studien*, *2*, 546–602.

Mellers, B. A. & Birnbaum, M. H. (1982). Loci of contextual effects in judgment. *Journal of Experimental Psychology: Human Perception and Performance*, *8*(4), 582–601.
DOI: 10.1037/0096-1523.8.4.582

Merkel, J. (1888). Das psychophysische Grundgesetz in Bezug auf Schallstärken (The psychophysical law of perceived sound strength). *Philosophische Studien*, *4*, 117–160.

Michon, J. A. (1967). *Timing in temporal tracking.* Soesterberg, The Netherlands: Institute for Perception RVO-TNO.

Moyer, R. S. & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, *216*, 159–160.
DOI: 10.1038/2151519a0

Narens, L. (1981). A general theory of ratio scalability with remarks about the measurement-theoretic concept of meaningfulness. *Theory and Decision*,

13(1), 1–70.
DOI: 10.1007/BF02342603

Narens, L. (1996). A theory of ratio magnitude estimation. *Journal of Mathematical Psychology*, *40*(2), 109–129.
DOI: 10.1006/jmps.1996.0011

Narens, L. & Mausfeld, R. (1992). On the relationship of the psychological and the physical in psychophysics. *Psychological Review*, *99*(3), 467–479.
DOI: 10.1037/0033-295X.99.3.467

Orth, B. (1974). *Einführung in die Theorie des Messens (Introduction to the theory of measurement)*. Stuttgart, Germany: Kohlhammer.

Painton, S. W., Cullinan, W. L., & Mencke, E. O. (1977). Individual pitch functions and pitch-duration cross-dimensional matching. *Perception & Psychophysics*, *21*(5), 469–476.
DOI: 10.3758/BF03199504

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.
DOI: 10.1163/156856897X00366

Penney, T. B. & Tourret, S. (2005). Les effets de la modalité sensorielle sur la perception du temps (the effects of the investigated sensory modality on the perception of time). *Psychologie Française*, *50*(1), 131–143.
DOI: 10.1016/j.psfr.2004.10.011

Plateau, M. J. (1872). Sur la measure des sensations physiques, et sur la loi qui lie l'intensite des ces sensations a l'intensite de la cause excitante (On the measurement of perceived intensities and on the law that relates perceived intensities to the physical magnitude). *Bulletin de I'Academie Royale des Sciences, des Lettres et des Beaux Arts de Belgique*, *33*, 376–388.

Pöppel, E. (2004). Lost in time: A historical frame, elementary processing units and the 3-second window. *Acta Neurobiologiae Experimentalis*, *64*(3), 295–302.

Poulton, E. C. (1989). *Bias in quantifying judgements*. East Sussex, UK: Lawrence Erlbaum Associates.

Pradham, P. L. & Hoffman, P. J. (1963). Effect of spacing and range of stimuli on magnitude estimation judgements. *Journal of Experimental Psychology*,

66(6), 533–541.
DOI: 10.1037/h0041492

Prelec, D. (1998). The probability weighting function. *Econometrica*, *66*(3), 497–527.
DOI: 10.2307/2998573

Rammsayer, T. H. (1994). A cognitive-neuroscience approach for elucidation of mechanisms underlying temporal information processing. *International Journal of Neuroscience*, *77*(1-2), 61–76.
DOI: 10.3109/00207459408986019

Rammsayer, T. H. (2010a). Differences in duration discrimination of filled and empty auditory intervals as a function of base duration. *Attention, Perception, & Psychophysics*, *72*(6), 1591–1600.
DOI: 10.3758/APP.72.6.1591

Rammsayer, T. H. (2010b). The effects of sensory modality and type of task on discrimination of durations ranging from 400 to 1,000 milliseconds. In A. Bastianelli & G. Vidotto (Eds.), *Fechner Day 2010* (pp. 439–444). Padova, Italy: International Society for Psychophysics.

Rammsayer, T. H. & Lima, S. D. (1991). Duration discrimination of filled and empty auditory intervals: Cognitive and perceptual factors. *Perception & Psychophysics*, *50*(6), 565–574.
DOI: 10.3758/BF03207541

Richards, W. (1964). Time estimates measured by reproduction. *Perceptual and Motor Skills*, *18*(3), 929–943.
DOI: 10.2466/pms.1964.18.3.929

Richardson, L. F. & Ross, J. S. (1930). Loudness and telephone current. *The Journal of General Psychology*, *3*(2), 288–306.
DOI: 10.1080/00221309.1930.9918206

Robinson, G. H. (1976). Biasing power law exponents by magnitude estimation instructions. *Perception & Psychophysics*, *19*(1), 80–84.
DOI: 10.3758/BF03199390

Ross, H. E. (1997). On the possible relations between discriminability and apparent magnitude. *British Journal of Mathematical and Statistical Psychology*, *50*(2), 187–203.
DOI: 10.1111/j.2044-8317.1997.tb01140.x

Rule, S. J. (1971). Discriminability scales of number for multiple and fractional estimates. *Acta Psychologica*, *35*(4), 328–333. DOI: 10.1016/0001-6918(71)90041-2

Rule, S. J. & Curtis, D. W. (1973). Conjoint scaling of subjective number and weight. *Journal of Experimental Psychology*, *97*(3), 305–309. DOI: 10.1037/h0034095

Schneider, B., Parker, S., Ostrosky, D., Stein, D., & Kanow, G. (1974). A scale for the psychological magnitude of number. *Perception & Psychophysics*, *16*(1), 43–46. DOI: 10.3758/BF03203247

Schneider, M. & Siegler, R. S. (2010). Representations of the magnitudes of fractions. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(5), 1227–1238. DOI: 10.1037/a0018170

Small, A. M. & Campbell, R. A. (1962). Temporal differential sensitivity for auditory stimuli. *The American Journal of Psychology*, *75*(3), 401–410. DOI: 10.2307/1419863

Steingrimsson, R. (2011). Evaluating a model of global psychophysical judgments for brightness: II. Behavioral properties linking summations and productions. *Attention, Perception, & Psychophysics*, *73*(3), 872–885. DOI: 10.3758/s13414-010-0067-5

Steingrimsson, R. & Luce, R. D. (2005a). Evaluating a model of global psychophysical judgements - I: Behavioral properties of summations and productions. *Journal of Mathematical Psychology*, *49*(4), 290–307. DOI: 10.1016/j.jmp.2005.03.003

Steingrimsson, R. & Luce, R. D. (2005b). Evaluating a model of global psychophysical judgements - II: Behavioral properties linking summations and productions. *Journal of Mathematical Psychology*, *49*(4), 308–319. DOI: 10.1016/j.jmp.2005.03.001

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*(1), 29–44. DOI: 10.1016/j.jmp.2006.08.001

Steingrimsson, R., Luce, R. D., & Narens, L. (2012). Brightness of different hues is a single psychophysical ratio scale of intensity. *The American Journal of Psychology*, *125*(3), 321–333.
DOI: 10.5406/amerjpsyc.125.3.0321

Stevens, J. C. & Guirao, M. (1964). Individual loudness functions. *Journal of the Acoustical Society of America*, *36*(11), 2210–2213.

Stevens, S. S. (1936). A scale for the measurement of a psychological magnitude: Loudness. *Psychological Review*, *43*(5), 405–416.
DOI: 10.1037/h0058773

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, *103*, 677–680.

Stevens, S. S. (1953). On the brightness of lights and the loudness of sounds. *Science*, *118*, 576.

Stevens, S. S. (1956). The direct estimation of sensory magnitudes: Loudness. *American Journal of Psychology*, *69*(1), 1–15.
DOI: 10.2307/1418112

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, *64*(3), 153–181.
DOI: 10.1037/h0046162

Stevens, S. S. (1958). Problems and methods of psychophysics. *Psychological Bulletin*, *55*(4), 177–196.
DOI: 10.1037/h0044251

Stevens, S. S. (1961). To honor Fechner and repeal his law. *Science*, *133*, 80–86.
DOI: 10.1126/science.133.3446.80

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

Stevens, S. S. & Galanter, E. H. (1957). Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology*, *54*(6), 377–411.
DOI: 10.1037/h0043680

Stevens, S. S. & Greenbaum, H. B. (1966). Regression effect in psychophysical judgment. *Perception & Psychophysics*, *1*(12), 439–446.
DOI: 10.3758/BF03215821

Stevens, S. S. & Guirao, M. (1962). Loudness, reciprocality, and partition scales. *The Journal of the Acoustical Society of America*, *34*(9B), 1466–1471. DOI: 10.1121/1.1918370

Stott, L. H. (1935). Time-order errors in the discrimination of short tonal durations. *Journal of Experimental Psychology*, *18*(6), 741–766. DOI: 10.1037/h0057714

Stubbs, D. A. (1988). The psychophysics of retrospective and prospective timing. *Perception*, *17*, 297–310. DOI: 10.1068/p170297

Teghtsoonian, M. & Teghtsoonian, R. (1983). Consistency of individual exponents in cross-modal matching. *Perception & Psychophysics*, *33*(3), 203–214. DOI: 10.3758/BF03202857

Teghtsoonian, M. & Teghtsoonian, R. (2003). Putting context effects into context. In B. Berglund & E. Borg (Eds.), *Fechner Day 2003* (pp. 309–314). Larnaca, Cyprus: International Society for Psychophysics.

Teghtsoonian, R. (1971). On the exponents in Stevens' law and the constant in Ekman's law. *Psychological Review*, *78*(1), 71–80. DOI: 10.1037/h0030300

Teghtsoonian, R. (2012). The standard model for perceived magnitude: A framework for (almost) everything known about it. *The American Journal of Psychology*, *125*(2), 165–174. DOI: 10.5406/amerjpsyc.125.2.0165

Thomas, E. A. C. & Brown, I. (1974). Time perception and the filled-duration illusion. *Perception & Psychophysics*, *16*(3), 449–458. DOI: 10.3758/BF03198571

Thomas, E. A. C. & Weaver, W. B. (1975). Cognitive processing and time perception. *Perception & Psychophysics*, *17*(4), 363–367. DOI: 10.3758/BF03199347

Treisman, M. (1963). Temporal discrimination and the indifference interval: Implications for a model of the "internal clock". *Psychological Monographs: General and Applied*, *77*(13), 1–31. DOI: 10.1037/h0093864

van Wassenhove, V. (2009). Minding time in an amodal representational space. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1525), 1815–1830.
DOI: 10.1098/rstb.2009.0023

Vierordt, C. (1868). *Der Zeitsinn nach Versuchen (Time sense examined in experiments).* Tübingen, Germany: H. Laupp.

Wackermann, J. (2007). Inner and outer horizons of time experience. *The Spanish Journal of Psychology*, *10*(01), 20–32.
DOI: 10.1017/S1138741600006284

Walker, J. T. & Scott, K. J. (1981). Auditory–visual conflicts in the perceived duration of lights, tones, and gaps. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(6), 1327–1339.
DOI: 10.1037/0096-1523.7.6.1327

Ward, L. M. (1995). On the role of discriminability in psychophysics. In C.-A. Possamai (Ed.), *Fechner Day 1995* (pp. 61–66). Cassis, France: International Society for Psychophysics.

Ward, L. M., Armstrong, J., & Golestani, N. (1996). Intensity resolution and subjective magnitude in psychophysical scaling. *Perception & Psychophysics*, *58*(5), 793–801.
DOI: 10.3758/BF03213110

Warm, J. S., Stutz, R., & Vassolo, P. A. (1975). Intermodal transfer in temporal discrimination. *Perception & Psychophysics*, *18*(4), 281–286.
DOI: 10.3758/BF03199375

Wearden, J. H. & Lejeune, H. (2008). Scalar properties in human timing: Conformity and violations. *The Quarterly Journal of Experimental Psychology*, *61*(4), 569–587.
DOI: 10.1080/17470210701282576

Zimmer, K. (2005). Examining the validity of numerical ratios in loudness fractionation. *Perception & Psychophysics*, *67*(4), 569–579.
DOI: 10.3758/BF03193515

List of Manuscripts

Manuscript A:
Birkenbusch, J., Ellermeier, W., & Kattner, F. (2015). Octuplicate This Interval! Axiomatic Examination of the Ratio Properties of Duration Perception. *Attention, Perception, & Psychophysics, 77(5),* 1767–1780. DOI: 10.3758/s13414-015-0846-0.

Manuscript B:
Birkenbusch, J., & Ellermeier, W. (accepted for publication). Axiomatic Evaluation of k-Multiplicativity in Ratio Scaling: Investigating Numerical Distortion. *Special Issue of the Journal of Mathematical Psychology in Honor of R. Duncan Luce*, 29 pages.

Manuscript C:
Birkenbusch, J., & Ellermeier, W. (to be revised and resubmitted). Quantifying Subjective Duration: Both Power Function Exponents and Weber Fractions Vary With the Standard. *Attention, Perception, & Psychophysics*, 23 pages.

Scientific CV

Jana Birkenbusch, M. Sc.
Date of Birth: 13/04/1989
Place of Birth: Darmstadt, Germany
Nationality: German

Since 10/2012	Research Assistant, Applied Cognitive Psychology, Technische Universität Darmstadt
09/2012	Master of Science Psychology, Technische Universität Darmstadt
10/2011-09/2012	Student Assistant, Fraunhofer-Institut für Graphische Datenverarbeitung (IGD), Darmstadt
10/2009-09/2011	Student Assistant, Applied Cognitive Psychology, Technische Universität Darmstadt
10/2008-09/2009	Student Assistant, Staatliche Materialprüfungsanstalt, Technische Universität Darmstadt
10/2008-09/2012	Studies of Psychology, Technische Universität Darmstadt
06/2008	Diploma, Albert-Einstein-Schule, Groß-Bieberau

Obligatory Declaration

I declare that I have developed and written the enclosed doctoral thesis entiteled "Evaluating the Ratio Scalability of Perceived Duration - An Axiomatic Study" completly by myself, and have not used sources or means without declaration in the text. Any thoughts from others or literal quotations are clearly marked. This thesis was not used in the same or in a similar version to achieve an academic grading or is being published elsewhere.

Darmstadt, July 24, 2015

....
Jana Birkenbusch, M. Sc.