PERCEPTUAL RATIOS, DIFFERENCES, AND THE UNDERLYING SCALE

Wolfgang Ellermeier, Louis Narens and Benno Dielmann
Department of Acoustics, Aalborg University, Denmark
Institute for Mathematical Behavioral Sciences, University of California, Irvine, CA, U.S.A.
Institut für Experimentelle Psychologie, Universität Regensburg, Germany
we@acoustics.dk, lnarens@uci.edu, mail@benno-dielmann.de

Abstract

Torgerson's (1961) conjecture that subjects cannot meaningfully distinguish perceptual ratios and differences has triggered a continuing debate in psychophysics on what constitutes adequate scaling methodology. It is shown that an axiomatic theory of magnitude estimation proposed by Narens (1996, 1997) sheds new light on this problem by dropping the assumption that subjects can report sensation magnitudes veridically. Furthermore, this theory provides qualitative predictions suited to test Torgerson's conjecture, by checking whether ratio and difference productions may be combined commutatively, which through the theory implies that the same psychological operation underlies both tasks. A loudness production experiment in which six subjects were instructed to double, or triple loudness, and to add a small, or a large loudness difference to a standard revealed that the two kinds of instructions were indeed commutative in most instances, consistent with Torgerson (1961) and other empirical results. Only two subjects showed small, but significant violations of commutativity. These results, when interpreted in the context of Narens' (1997) axiomatic theory, suggest that most individuals' ratio and difference adjustments can be simultaneously represented as ratios. This conclusion, which implies Torgerson's conjecture, is derived using methods that are more rigorous than those employed previously to tackle the problem.

In psychophysics, scaling methods supposedly based on the evaluation of sensory differences, such as equisection or categorisation, appear to be incompatible with methods supposedly based on the evaluation of sensory ratios, such as fractionation, or magnitude estimation. As early as 1961, however, Torgerson - explicitly refering to these two classes of scaling methods - proposed a provocative hypothesis about the relation between *perceptual ratios* and *differences* that initiated an ongoing research tradition of its own. Based on his work on scaling the lightness of Munsell grays, Torgerson observed that category ratings (on a scale from 0 to 10) and ratio estimates (magnitude estimation with a fixed standard and modulus) of the same stimuli were non-linearly related, suggesting that subjects do not evaluate differences and ratios on the same underlying scale "in the manner required by the number system" (Torgerson, 1961, p. 203). Rather, the difference scale exhibited a linear relationship with the logarithm of the ratio scale, suggesting that stimuli judged to be separated by equal subjective distances under one instruction were also separated by equal subjective ratios under the other instruction. This lead Torgerson to hypothesize that "the subject perceives or appreciates but a single quantitative relation between a pair of stimuli" (p. 203), which he/she reports as a ratio or a difference, depending on what the instructions require. This claim has been known as *Torgerson's conjecture* in the research literature (e.g. Schneider, 1980; Birnbaum, 1982).

Given the many assumptions implicit in Torgerson's conjecture, it appears desirable (a) to have a more formalized theory of judgments of ratios and differences, and (b) to derive tests from it that do not require more than qualitative comparisons. Precisely that has been accomplished based on an axiomatic theory of magnitude scaling proposed by Narens (1996) and later extended to Torgerson's problem (Narens, 1997).

Narens' theory of subjective intensity

Narens' (1996) "Theory of Ratio Magnitude Estimation" formulates the assumptions inherent in S.S. Stevens' direct scaling approach by specifying them in the form of an axiomatic theory. This theory takes care to treat the *numerals* uttered by the subject in a magnitude estimation task as distinct from (scientific) *numbers*, of which the subject may or may not have a sound understanding. Two of the axioms (*commutativity* and *multiplicativity*) formulated by Narens (1996) are crucial to both an empirical test of the theory and the interpretation of subjects' scaling behavior. These axioms were empirically evaluated by Ellermeier & Faulhammer (2000) by having subjects produce loudness ratios. They showed that subjects' loudness adjustments were consistent with the weaker *commutativity* axiom (e.g., making a reference sound twice as loud and tripling the loudness of the resultant yields the same sound-pressure level as first tripling and then doubling the loudness). However, the multiplicativity axiom (e.g., the fact that consecutive doubling and tripling of loudness should be equivalent to making the starting intensity six times as loud), was violated in a significant number of cases. According to Narens' axiomatization, this outcome implies that though in principle a ratio scale of loudness exists, the numerals used by subjects in order to describe sensation ratios may not be taken at face value.
Narens' (1997) extension of the theory is based on the following idea: A person's intensity judgments on a continuum of stimuli result from "calculations" (e.g., algorithms) based on an inner psychological structure that is ratio scalable. Using this, and results from Narens theory of meaningfulness (published later in Narens, 2002), it follows that for functions f and g on the stimuli that are produced using ratio production or difference production or both kinds of productions the following *Generalized Commutativity Principle* holds:

$$f * g = g * f,$$

where, of course, $*$ is the operation of functional composition.
Because it has been empirically established by Ellermeier and Faulhammer (2000) that ratio production functions commute, Torgerson's conjecture that subjective differences are subjective ratios suggests that difference production functions should commute with ratio production functions.
Note, however, that Togerson's conjecture was based on a theory of measurement due to S. S. Stevens which identifies subjects' responses to ratio and difference instructions veridically; that is, in the ratio condition, the "*numeral* ratio $\mathbf{3}$" was interpreted by the measurement procedure as the *numerical* ratio represented by the *real number* 3, and the "*numeral* difference $\mathbf{2}$" was interpreted as the *numerical* difference represented by the *real number* 2. Such veridicality is not part of Narens (1997) theory nor its deduction of the Generalized Commutativity Principle.

Operationalization of Ratio and Difference Production

While ratio productions are easily implemented by asking the subject to "make the second tone $\mathbf{p}$ times as loud" as the first one (denoted by $R_{\mathbf{p}}(x) = y$, with x being the level of the first (reference) tone, and y being the result of the adjustment), we cannot simply say for difference productions "adjust the second tone so that the loudness difference is $\mathbf{p}$" without providing a unit. Therefore, the following *difference matching* instruction $D_{a,b}$ was implemented where $D_{a,b}(x) = y$ holds if and only if the subject adjusts a stimulus y such that "the difference in loudness between y and x is the same as the difference between b and a."
The above observation leads to the following specification of the General Commutativity Principle (GCP): For behavioral functions $R_{\mathbf{p}}$ and $D_{a,b}$ and all stimuli x in X:

$$D_{a,b}[R_{\mathbf{p}}(x)] = R_{\mathbf{p}}[D_{a,b}(x)]. \quad (1)$$

For each such test of the GCP, the subject has to perform two ratio productions (here specified as making the comparison tone three times as loud), and two interval matches as illustrated in Figure 1. If the GCP holds, the two orders of chaining these operations should not make a difference, and multiple productions of y and v (as specified in Figure 1) should be statistically indistinguishable.

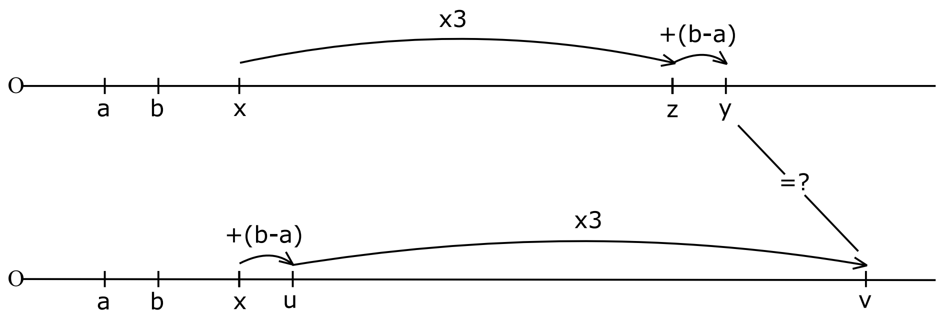

Figure 1: Illustration of the paradigm used to investigate the validity of the General Commutativity Principle (GCP). For an explanation, see text.

Method

Participants

The first author and five staff members at the Department of Acoustics, University of Aalborg, participated in the experiments. This sample ranged in age from 22 to 45 years, and consisted of 4 male and 2 female participants. All subjects had normal hearing thresholds in the frequency range from 0.25 to 8 kHz. None - except for the author - had prior knowledge of the hypotheses being investigated.

Apparatus

The stimuli were 1-kHz sinusoids of 500 ms duration, including 10-ms rise and decay ramps. They were computed via a Tucker-Davis-Technologies (TDT) signal processor card (model AP2), and played from a 16-bit digital-analogue converter (TDT model DD1) at a sampling rate of 50 kHz. After passing through a low-pass filter set at 10 kHz (TDT model FT5), the signal was set to the proper level by means of a programmable attenuator having 0.1 dB resolution (TDT model PA4). Subsequently, the signal was fed to a headphone buffer (TDT HB6) from which it was diotically delivered to the subject seated in a double-walled sound-attenuated chamber via AKG-K 501 headphones.

Stimuli and experimental design

To test whether the General Commutativity Principle (GCP) holds, the ratio production and interval addition had to be chained in two ways (see Figure 1) in order to check whether the two operations commute. Both to increase the generality of this test and to discourage stereotypical adjustments, two different ratio instructions were emloyed, making a sound twice $R_{\mathbf{2}}$, and making it three times as loud $R_{\mathbf{3}}$. Likewise, two different intervals had to be added: one specified by the loudness difference between 50 and 58 dB SPL (denoted as $+[50, 58]$), and the other one by the loudness difference between 50 and 62 dB SPL ($+[50, 62]$). Combining the two ratios with the two differences yielded 4 tests of the GCP for each participant, requiring a total of (a) four “simple” adjustments, all made from a base starting level of 65 dB SPL, (b) four “consecutive” productions according to the left side of eq. 1, and (c) four consecutive productions of the right side of eq. 1. These 12 trial types were randomized within each block of trials, leading to sufficient variability in starting levels and target values to reduce memory effects, and to obscure the participant’s view with respect to the theoretically “critical” comparisons.

Procedure

Each trial proceeded as follows: One of three appropriately labelled LEDs mounted on a hand-held response unit signalled the subject which instruction to follow: (1) Making a standard tone ”twice as loud”, (2) making a standard ”three times as loud”, or (3) adding a specified loudness difference to a standard.

Ratio instruction: On ”ratio” trials, strictly following the procedure, developed by Ellermeier & Faulhammer (2000), the subject was presented with a repeated sequence of two tones, and instructed to make the second tone twice (or three times) as loud as the first one. As soon as the subject pressed a ”ready” button, the standard stimulus kept alternating with the (variable) comparison, which the subject was asked to adjust. The inter-stimulus interval between standard and comparison was 500 ms, and the subject could adjust the level of the comparison tone during the 2 s which elapsed before the next pair was presented. Details of the adjustment procedure which included features borrowed from adaptive procedures (such as decreasing stepsizes), are described in Ellermeier & Faulhammer (2000).

Difference instruction: Difference productions followed the same scheme, except that the reference interval to be added was played before each presentation. Thus, the subject was repeatedly presented with a sequence of *four* tones. The first pair of tones, separated by a 500-ms ISI, marked the ”reference” loudness difference ($[50, 58]$ or $[50, 62]$) to be added on a given trial. The second pair of tones constituted the difference to be matched, with the third tone in the sequence being the fixed standard, and the fourth tone being the variable comparison. The subject was asked to adjust the level of this tone so that the loudness difference perceived in the second pair equalled the loudness difference perceived in the first pair.

In addition to a practice session consisting of two blocks in which each of the 12 trial types occured, each subject completed 15 blocks of trials in 4 sessions, thus producing a total of 15 adjustments of each type.

Results

Precision and Monotonicity of Adjustments

The participants took considerable care to make their adjustments: They spent a median time of 50 s on each adjustment, and made a median number of 9 level changes to achieve it. In the vast majority of trials, they reached the minimum stepsize of 0.5 dB. The only difference emerging between ratio and difference trials appears to be that the latter are slightly more reliable: While simple ratio instructions ($R_2[x]$ or $R_3[x]$) yielded an average individual standard error of 0.68 dB, adding loudness differences ($D_{50,58}[x]$ or $D_{50,62}[x]$) produced an average error of only 0.35 dB. Thus, there is no indication that the new difference matching task was more difficult for the participants than ratio production. Self-reports, in fact, argue that difference matching was the easier task.

Furthermore, the mean outcomes of doubling vs. tripling instructions (as well as those of adding small vs. large loudness intervals) were monotonically ordered for each individual subject, thus increasing confidence in the validity of their adjustments.

Testing the General Commutativity Principle (GCP)

Figure 2 illustrates the outcome of one particular test of the GCP for an individual subject (KA). The sequence $D_{50,58}[R_2(x)]$ is depicted by the left concatenation of an arrow illustrating the effect of doubling the loudness of the standard (set at 65 dB SPL), and a line segment illustrating the effect of subsequently adding a (small) loudness interval to the outcome of the first operation. The reverse concatenation on the right side of the graph (arrow on top of a line segment) depicts the sequence $R_2[D_{50,58}(x)]$. As is evident in Figure 2, both sequences produced nearly identical average outcomes differing by a mere 0.2 dB.

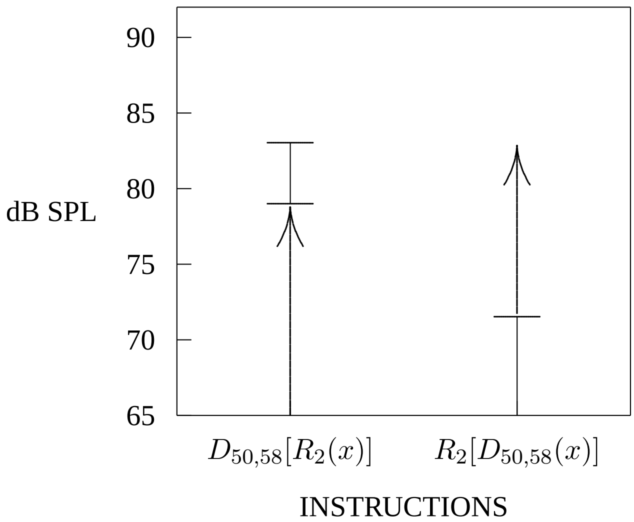

Figure 2: Outcome of four tests of the GCP for an individual participant (KA). The entries on the abscissae give the order of instructions, with the "inner" function being applied first. In the panels, arrows mark ratio productions, line segments with horizontal tops mark the outcome of adding a loudness interval. The sound-pressure levels plotted are averages of 15 adjustments per condition.

Because two ratio instructions ($R_2[x]$ and $R_3[x]$) were combined with two difference instructions ($D_{50,58}[x]$ and $D_{50,62}[x]$), four tests of the GCP were performed on each subject's data. For our 6 subjects, that resulted in a total of 24 tests. The outcome of these tests is given in Table 1. It is obvious, that the decibel differences produced by adding a difference subsequent to generating a loudness ratio, as opposed to applying the two operations in the reverse order, are relatively small, the average discrepancy amounting to less than a decibel. Two-tailed U-tests were performed to determine whether the two sets of adjustments differed significantly. Assuming $\alpha = 0.05$, in 24 tests, only 4 instances were found, in which the GPC was violated (indicated by asterisks in Table 1).
Looking at the overall outcome broken down by subjects, it turns out that two (of 6) subjects (*vs* and *jg*) showed significant violations of the GPC in half of the conditions investigated. Two more violations emerge for one of these subjects (*jg*), if the probability of detecting a violation is increased by adopting $\alpha = 0.10$ as is often done, when the goal of the investigation is to maintain the null hypothesis.

Discussion

The demonstration that ratio productions and difference productions of loudness commute, confirms Narens' (1997) theory of subjective intensity, and agrees with Torgerson's conjecture. The fact that two subjects are not in line with this outcome, may be consistent with the occasional observation that some subjects sometimes distinguish perceptual ratios and differences (e.g. Schneider, 1980; Birnbaum, 1982). Note, however, that the deviations from the Generalized Commutativity Principle are small, and that the direction of the effect (positive differences in Table 1) is the opposite of what would be intuitively expected when numerical ratios and differences are concatenated (see Figure 1).
The present study further stresses the importance of distinguishing between *numeral* ratios or differences (i.e. the "number words") used or interpreted by subjects, from true *numerical* ratios or differences. Also note that Narens (1996, 1997) allow for distortions of the usual numerical interpretations of ratio and difference productions, because these productions are characterized in terms of qualitative properties. Determining the function h that translates subjects' behavior involving *numerals* $\mathbf{p}$ into *numbers* $h(\mathbf{p})$ is at present an unresolved issue. Ellermeier and Faulhammer (2000) have shown empirically that h cannot have the form $h(\mathbf{p}) = p$ where p is the number that mathematically corresponds

Table 1: Tests of the General Commutativity Principle (GCP)

S.	$2 \times +D_{50,58}$ vs. $+D_{50,58} \times 2$ diff[dB]	p	$2 \times +D_{50,62}$ vs. $+D_{50,62} \times 2$ diff[dB]	p	$3 \times +D_{50,58}$ vs. $+D_{50,58} \times 3$ diff[dB]	p	$3 \times +D_{50,62}$ vs. $+D_{50,62} \times 3$ diff[dB]	p
we	+ 0.67	0.441	- 0.50	0.950	+ 0.30	0.787	+ 1.40	0.203
vs	+ 2.07	$0.036^{\star}$	+ 1.53	$0.039^{\star}$	+ 0.63	0.560	+ 0.50	0.467
sc	+ 0.47	0.662	+ 0.16	0.724	+ 0.20	0.724	+ 0.43	0.561
ka	+ 0.20	0.917	+ 0.03	0.693	+ 0.70	0.739	+ 0.66	0.604
jg	+ 1.73	$0.007^{\star\star}$	+ 1.30	$0.044^{\star}$	+ 1.30	0.092	+ 1.10	0.064
az	- 0.07	0.708	- 0.26	0.917	+ 0.46	0.464	- 0.77	0.561

Note. Mean decibel differences produced by concatenating a ratio and a difference instruction in two different orders are given, followed by the p-value for the significance of the difference (U-test, two-tailed). Since two ratio-production factors ($\times 2$ and $\times 3$) were combined with the addition of two loudness differences ($D_{50,58}(x)$ and $D_{50,62}(x)$), four tests of the General Commutitivity Principle resulted.

to the numeral $\mathbf{p}$. In a related but different kind of paradigm, Luce and Steingrimsson (2003) provide a theory of a special form of the function relating numerals and numbers. It is currently being put to an empirical test by these authors as well as by Zimmer and Baumann (2003).

References

Birnbaum, M.H. (1982). Controversies in psychological measurement. In B. Wegener (Ed.), *Social attitudes and psychological measurement* (pp. 401-485). Hillsdale, N.J.: Erlbaum.

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

Luce, R.D. & Steingrimsson, R. (2003). A model of ratio production and estimation and some behavioral predictions. This volume.

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

Narens, L. (1997). On subjective intensity and its measurement. In A.A.J. Marley (Ed.), *Choice, decision, and measurement: Essays in honor of R. Duncan Luce* (pp. 189-205). Mahwah, N.J.: Erlbaum.

Narens, L. (2002). *Theories of Meaningfulness.* Mahwah, N.J.: Lawrence Erlbaum.

Schneider, B.A. (1980). A technique for the nonmetric analysis of paired comparisons of psychological intervals. *Psychometrika, 45,* 357-372.

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

Zimmer, K. & Baumann, O. (2003). Validity of Luce's transformation function relating numerals and numbers in loudness fractionation. This volume.