Cognition 169 (2017) 36–45

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Original Articles

Spontaneous, modality-general abstraction of a ratio scale

Cory D. Bonn *, Jessica F. Cantlon

Department of Brain and Cognitive Sciences, 358 Meliora Hall, PO Box 270268, University of Rochester, Rochester, NY 14627-0258, United States

ARTICLE INFO

Article history:
Received 25 September 2016
Revised 26 July 2017
Accepted 29 July 2017

Keywords:
Generalized magnitude representation
Ratio scale

ABSTRACT

The existence of a generalized magnitude system in the human mind and brain has been studied extensively but remains elusive because it has not been clearly defined. Here we show that one possibility is the representation of relative magnitudes via ratio calculations: ratios are a naturally dimensionless or abstract quantity that could qualify as a common currency for magnitudes measured on vastly different psychophysical scales and in different sensory modalities like size, number, duration, and loudness. In a series of demonstrations based on comparisons of item sequences, we demonstrate that subjects spontaneously use knowledge of inter-item ratios within and across sensory modalities and across magnitude domains to rate sequences as more or less similar on a sliding scale. Moreover, they rate ratio-preserved sequences as more similar to each other than sequences in which only ordinal relations are preserved, indicating that subjects are aware of differences in levels of relative-magnitude information preservation. The ubiquity of this ability across many different magnitude pairs, even those sharing no sensory information, suggests a highly general code that could qualify as a candidate for a generalized magnitude representation.

1. Introduction

Magnitudes such as size, duration, and number share similar psychophysical signatures, appear to use overlapping neural resources, and can influence each other in dual tasks. These observations are consistent with the existence of a shared analog code or generalized magnitude representation (Gallistel & Gelman, 2000; Holyoak & Glass, 1978; Pinel, Piazza, Le Bihan, & Dehaene, 2004; Walsh, 2003; see Bonn & Cantlon, 2012; Bueti & Walsh, 2009; Cantlon, Platt, & Brannon, 2009; Cohen Kadosh, Lammertyn, & Izard, 2008; and Lourenco, 2015 for extensive reviews). However, the shared-code hypothesis remains underspecified because the existing data has not revealed much about the code's internal structure. In addition, it remains unclear whether the many studies demonstrating interactions between magnitude dimensions are tapping into more than one possible mechanism.

At a minimum, a shared code should be inherently meaningful across many magnitude domains and across sensory modalities. Two types of relative-magnitude representation—ratios and ranks—automatically offer such generality at different levels of granularity. Ratios and ranks are dimensionless quantities that abstract away from original metrics; for example, the ratio of 1:2 is meaningful on any intensity scale such as loudness or size. Some recent evidence suggests that ratios, represented by pairs of lines of different length or subsets of dot arrays painted in a particular color, are spontaneously represented in a fronto-parietal network in adult humans and macaques (Vallentin & Nieder, 2008, 2010; Jacob & Nieder, 2009; Jacob, Vallentin, & Nieder, 2012), but it is unclear whether these representations are restricted to their particular dimensions. In principle, ratios could support cross-dimension mapping between pairs of structurally similar analog magnitudes (Srinivasan & Carey, 2010), but current evidence for such transfer is limited.

We explore the possibility that humans spontaneously represent fine-grained information about ratios and ranks in a format that can be compared across modalities and dimensions, providing a candidate for a generalized magnitude representation.

1.1. Current evidence for cross-dimension transfer of relative magnitudes

An abstract representation of relative magnitude should allow observers to transfer information about a set of two or more stimuli from one dimension to another without presenting both dimensions simultaneously. Evidence for the transfer of representations of relative magnitude across dimensions or across sensory

* Corresponding author at: Laboratoire Psychologie de la Perception, UMR 8242, Centre National de Recherche Scientifique, Université Paris Descartes, 45 rue des Saints-Pères, 75006 Paris, France.

E-mail addresses: corydbonn@gmail.com (C.D. Bonn), jcantlon@rcbi.rochester.edu (J.F. Cantlon).

http://dx.doi.org/10.1016/j.cognition.2017.07.012
0010-0277/© 2017 Elsevier B.V. All rights reserved.

modalities is scattered across several literatures; here we review a selection of representative examples.

Magnitude-estimation experiments designed to measure subjective sensation demonstrated that after observing a change in magnitude in one dimension relative to an anchor stimulus, subjects are able to generate equivalent proportional changes in other dimensions, given explicit instruction (Luce, 1990, 2002; Shepard, 1981; Stevens, 1975; Stevens, Mack, & Stevens, 1960). However, more ambiguous instructions in this task could elicit a wider range of magnitude estimates from unconstrained, heterogeneous transformation rules. Using a more constrained bisection task, Balci and Gallistel (2006) found that within-dimension calculation of proportions likely explained the transfer of duration discrimination to numerical discrimination behavior in humans, but it is unknown how generalizable this result is across multiple dimensions, and whether subjects spontaneously represent rank or proportion relations from sequences.

Cross-dimension transfer of relative magnitudes has been shown in infants for more imprecise representations resembling the concepts of *more* or *less*. For example, Lourenco and Longo (2010) showed that when infants learned to associate arbitrary features with large and small object sizes, they expected a similar association between those same features and large and small numerosities or durations. In another study, de Hevia and Spelke (2010) showed that after exposure to a series of stimuli with increasing or decreasing numerosities, 8-month-olds failed to dishabituate to sequences of lines changing length in the same direction, but dishabituated to sequences proceeding in the opposite direction. These studies leave open the question of whether infants generate more precise representations such as ratios and multi-item ranks.

In audition, humans and macaques retain representations of pitch-height changes in sequences of tones ('melodic contour'; Brosch, Selezneva, Bucks, & Scheich, 2004; Dowling & Fujitani, 1971; Marvin, 1997; Marvin & Laprade, 1987; Trehub, Thorpe, & Morrongiello, 1987). One study found that these representations can be constructed from and transferred across other auditory continua such as brightness and loudness (McDermott, Lehr, & Oxenham, 2008). Other studies found that adults can compare melodies to line drawings that represent long sequences of pitch-height changes (Prince, Schmuckler, & Thompson, 2009), suggesting a modality-independent representation of height. However, the granularity of these abstract representations of pitch and other auditory contours remains unknown.

In summary, previous studies are consistent with the existence of precise, spontaneous, relative-magnitude representations that can be transferred or mapped across diverse dimensions, but no series of experiments has demonstrated precision, spontaneity, and generality of these underlying representations simultaneously while keeping the behavioral methodology constant. Moreover, to our knowledge, no study has yet explicitly distinguished between ratio-based and rank-order-based representations of magnitude sets as potential candidates for generalized magnitude representations.

1.2. Overview of experiments

We provide evidence that human adults use precise, relative-magnitude information to compare sequences within and across sensory modalities and dimensions. Using a sequence-comparison method, we tested the specific hypotheses that subjects (1) can automatically extract ratio information within visual and auditory modalities and (2) can use it to compare sequences across sensory modalities and across the dimensions of space, time and number.

We created pairs of stimulus sequences containing a randomly generated, standard sequence and a comparison sequence that preserved the standard's abstract structure with varying levels of precision. The comparison could be the same sequence (*Same* sequences, for within-dimension comparisons only), a sequence in which between-item ratios were preserved (*Ratio* sequences), a sequence in which only the between-item ranks was preserved (*Rank* sequences), and a pseudorandom sequence that violated the rank-ordering of the standard (*Different* sequences); see Fig. 1 for an illustration. We predicted that perceived similarity of patterns would decrease as a function of increased information loss from standard to comparison: *Same* > *Ratio* > *Rank* > *Different*.

2. Experiment 1: within-dimension sequence comparisons

In this experiment, sequence pairs were presented in the same stimulus dimension, with separate groups of subjects tested in each. Visual sequences consisted of three squares varying in the dimensions of height or surface area. Auditory sequences consisted of three, band-pass-filtered samples of white noise varying in the dimensions of brightness (center frequency) or loudness (bandwidth and gain). These particular dimensions were chosen for the following reasons: (1) loudness and brightness are subsets of dimensions used in McDermott et al. (2008) and the visual continua of size and height provide intuitive, simple analogues in another sensory modality; (2) they also provide samples of both quantitative dimensions (amounts or intensities) and qualitative dimensions (continuous or categorical features; see Stevens, 1975, and Gati & Tversky, 1982 for further theoretical discussion). Size and loudness are examples of quantitative dimensions while auditory brightness is an example of a qualitative dimension. Object height is interpretable as either, depending upon whether it is measured as a distance from an anchor or a location detected using a filter bank.

2.1. Method

2.1.1. Subjects

Adults from the United States were recruited on Amazon Mechanical Turk; $n = 15$ for each of the stimulus dimensions. Subjects were paid \$3.50 (≈\$8/h).

2.1.2. Stimuli

Visual stimuli appeared in a white, 600-by-600-pixel (px) window with a black, 1-px-thick border. On each trial, 'Sequence 1' or 'Sequence 2' was printed in the middle of the viewing area prior to each sequence for 750 ms. Sequences consisted of three, 500-ms stimulus intervals separated by 250 ms inter-stimulus intervals.

Magnitude ratios of adjacent stimuli were constrained to be no smaller than 7:8. In the size sequences, squares were constrained

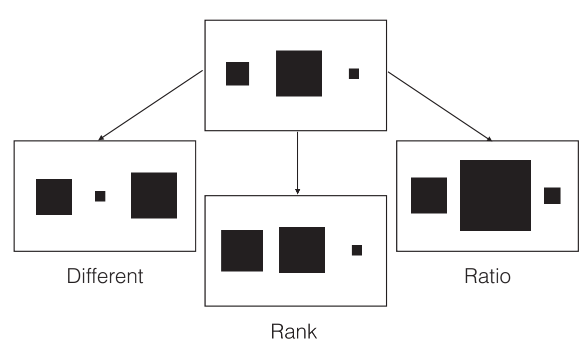

Fig. 1. Examples of sequence types. At the top is the standard sequence with the horizontal axis representing time (left to right). Arrows lead to possible comparison sequences in the *Different*, *Rank*, and *Ratio* pairs.

to be from 50 to 500 px in width (and height) and were centered in the stimulus window. In the height sequences, they were constrained to be located within a 500-px vertical range in the center of the stimulus window.

Auditory stimuli were presented with a similar visual layout, though during sound presentation a speaker icon appeared in the center of the stimulus window. Sounds consisted of a white-noise sample generated offline but filtered in the browser. Each sound consisted of 5-ms ramp-up and ramp-down periods to minimize transients.

In brightness sequences, subjects heard sounds in which the filter's center frequency was varied. Center frequencies were constrained to be between 1000 Hz and 10,000 Hz and scaled in equivalent-rectangular-bandwidth units. For the loudness dimension, subjects heard sounds varying in filter width and amplitude, but not center frequency. The simultaneous filter-width and amplitude manipulations allowed for a wide range of differences in subjective loudness without generating uncomfortable stimulus levels.

2.1.3. Procedure

Before testing, subjects were instructed to pay attention to the patterns of variation along each dimension in each sequence without any specific instruction for the *kind* of pattern, which is the standard instruction in melodic-contour experiments (Dowling & Fujitani, 1971).

At the end of each trial, subjects adjusted a 10-by-400-px, vertical slider on the right side of the screen to rate how similar the patterns were. The lower end and upper ends of the slider were labeled 'Very Different' and 'Very Similar,' respectively. The slider handle always began in the middle (200 px). Subjects recorded their final response by clicking a button to begin the next trial. Subjects received no feedback.

The full experiment consisted of two practice trials (one each of *Same* and *Different* sequence types), then 20 trials of each sequence type at test presented in random order, yielding 82 trials per subject. Subjects proceeded from practice trials to test trials without interruption. In addition to test trials, 5 catch trials where the visual stimulus was replaced with the words 'CATCH TRIAL' were included. On these trials, subjects were required to press a button labeled 'Catch Trial' instead of move the slider.

2.2. Results

2.2.1. Subject inclusion

Separate pilot experiments indicated subjects could potentially rate at random while still paying attention to the catch trials, presumably to maximize pay for the least effort. To reduce this source of data contamination, subjects were included in the analysis if (1) they missed fewer than 3 catch trials *and* (2) rated *Same* sequences significantly above the midline (one-sample, one-tailed t-tests on un-transformed ratings). There were 51 subjects included in the analysis (Size: $n = 14$, Height: $n = 12$; Loudness: $n = 11$; Brightness: $n = 14$).

2.2.2. Analytical approach

We fitted a linear, multilevel, regression model (Gelman & Hill, 2007) using restricted maximum likelihood to estimate effects of stimulus type and dimension on ratings. We fitted models with the maximal random effects structure, which included random intercepts and slopes by subject for sequence type and random intercepts by stimulus (see Barr, Levy, Scheepers, & Tily, 2013). We used Satterthwaite's (1946) approximate degrees of freedom for assessing significance of the test statistics as implemented in the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2016).

2.2.3. Rating transformation

Ratings within each dimension were compressed near the range limits, so we rescaled the dependent measure from 0 to 400 px to be in the interval [0, 1] and transformed them using the empirical logit to account for the presence of zeros and ones: *empirical logit* $(p) = \log((rating + 0.5)/(1 - rating + 0.5))$.

2.2.4. Predictors

We coded sequence type using a difference-coding scheme. Each coefficient thus represents the mean difference in rating (on the empirical logit scale) between the following sequence-type pairs: *Different* vs. *Rank*, *Rank* vs. *Ratio*, and *Ratio* vs. *Same*.

We contrast coded dimensions to test for differences in performance across sensory modalities, dimension types (qualitative vs. quantitative), and their interaction. We labeled the height dimension as qualitative (despite its ambiguous status), expecting that variation in ratings caused by differences between height and brightness would be captured indirectly by the interaction term.

2.2.5. Model results

Due to the large number of coefficients to report, in this paper we most often present and interpret the important results in the text without numbers but refer to and display the complete set of statistics in tables.

Fig. 2 displays the results in rating space. The details of the main model are presented in Table 1.

The model revealed a main effect of sequence type, indicated as SequenceType in the ANOVA table, with the corresponding coefficients indicating significant differences between each successive pair of levels, indicated as Different.vs.Rank, Rank.vs.Ratio, and Ratio.vs.Same in the regression table. As predicted, *Same* sequences were rated highest, followed by *Ratio*, then *Rank*, and finally *Different*.

The model failed to reveal a main effect of dimension but indicated a significant sequence-type-by-dimension interaction, indicated as Dimension and SequenceType:Dimension in the ANOVA table, respectively. The corresponding set of coefficients indicated the following: (1) the difference in ratings between the *Different* and *Rank* types were smaller for auditory than visual stimuli, as indicated by the Different.vs.Rank:Audition.vs.Vision coefficient; (2) the differences in ratings between the *Ratio* and *Rank* types were smaller for qualitative than for quantitative stimuli, as indicated by the Rank.vs.Ratio:Qualitative.vs.Quantitative coefficient, and (3) the differences in ratings between *Ratio* and *Same* stimuli were larger for qualitative than for quantitative stimuli, as indicated by the Ratio.vs.Same:Qualitative.vs.Quantitative coefficient. Supplementary analyses probing simple differences between *Rank* and *Ratio* ratings for each dimension confirmed the significance of the corresponding coefficient in all but the brightness dimension [$B = 0.0317$, 95% CI = (-0.0513, 0.1146)] (see Supplementary Materials, S1, for a full table of simple slopes).

2.3. Interim discussion

The overall results of Experiment 1 showed the predicted ratings pattern: when the comparison sequence preserves less information about the standard, similarity ratings decrease. Moreover, each level of information preservation was statistically distinguishable from the next in the ratings, indicating an awareness of violations of absolute magnitude levels and sensitivity to changes to inter-stimulus ratios and ranks. Most importantly, the ratings show that subjects distinguish between *Ratio* and *Rank* sequences, indicating the presence of a spontaneous yet precise level of abstract magnitude representation within each dimension.

The difference in ratings between *Different* and *Rank* sequences varied by sensory modality, indicating either increased difficulty in

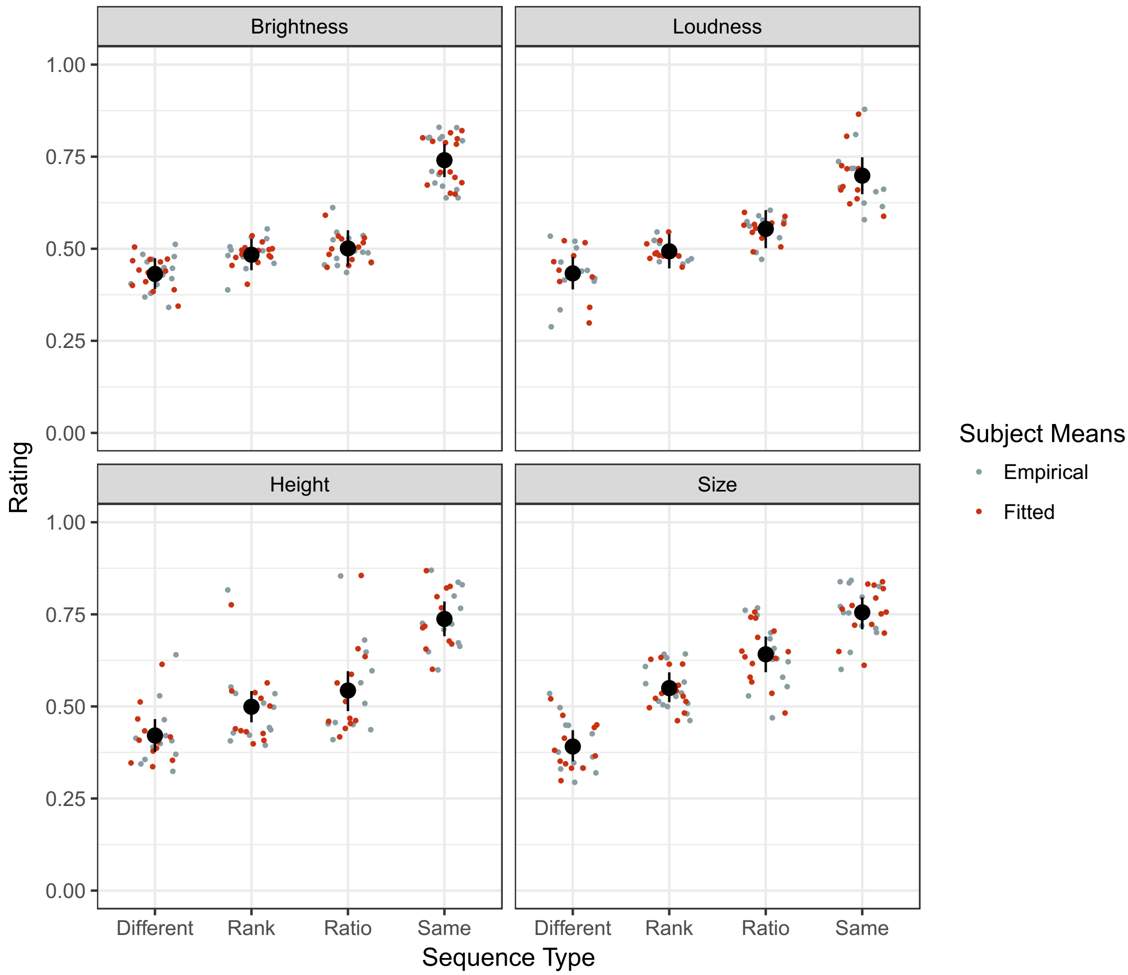

Fig. 2. Experiment 1 results. All displayed mean ratings are back-transformed from empirical logit space to the data space (0 = most dissimilar, 1 = most similar). Each panel corresponds to a subject group tested in one dimension. Black dots with 95% confidence intervals represent mean ratings generated from 1000 simulations from the fitted model. The small, horizontally jittered dots represent individual subject means, the gray being the raw means and the red being model predictions (fitted means), which take into account the random-effects structure. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the auditory dimensions, increased variability in the rendering of stimuli across subjects within those domains due to uncontrollable differences in equipment, or both.

Greater differences in ratings between *Ratio* and *Same* sequences and smaller differences in ratings between *Rank* and *Ratio* sequences for qualitative dimensions of brightness and height indicated either a decreased sensitivity to inter-item ratios overall for these dimensions, an over-weighting of absolute magnitude cues, or both. In particular, though the ratings follow the general pattern, the difference between *Ratio* and *Rank* ratings was not significant for the brightness dimension.

To test the hypothesis that the presence of *Same* trials led to the over-weighting of absolute cues in the brightness condition that masked observable sensitivity to inter-item ratios, we ran an additional set of 12 subjects in that dimension who were not exposed to *Same* trials and 30 trials of each sequence type.

The data from 10 subjects passing criterion (see Experiment 2a for procedure) are displayed in Fig. 3. There was a main effect of sequence type [$F(25.906) = 17.375$, $p < 0.00002$] and coefficients corresponding to the differences in ratings between *Different* and *Rank* [$B = 0.220$, $SE = 0.0522$, $t(20.860) = 4.204$, $p < 0.001$] conditions and between *Rank* and *Ratio* sequences [$B = 0.158$, $SE = 0.041$, $t(48.890) = 3.885$, $p < 0.001$] were significantly above 0, indicating that subjects were indeed sensitive to the differences in *Rank* and *Ratio* sequences, but possibly weighed absolute cues too highly in the original experiment. In addition, a regression comparing these results to the *Ratio* sequence ratings for the brightness condition in Experiment 1 revealed that ratings were significantly higher in the supplementary group [$B = 0.215$, $SE = 0.059$, $t(48.880) = 3.615$, $p < 0.001$].

3. Experiment 2: cross-dimension sequence comparisons

A domain-general code, whether shared as a common resource or simply a common code generated by all systems representing magnitudes, should be able to support comparisons across dimensions. In Experiment 2a, we demonstrate that fine-grained representations of relative magnitudes extracted in Experiment 1 can be compared across vision and audition. In Experiment 2b, we show that similar cross-dimension comparison behavior extends to time (interval duration) and number (Arabic numerals).

3.1. Experiment 2a

We asked subjects to compare object-height sequences to noise-brightness sequences and to compare object-size sequences to noise-loudness sequences.

3.1.1. Method

3.1.1.1. Subjects. Twelve subjects per dimension pair (48 total) were recruited via Mechanical Turk and paid $4.

Table 1
Model results for Experiment 1. The top part of the table indicates the results for each individual regression coefficient while the bottom indicates the associated ANOVA table for assessing the significance of each batch of coefficients overall. Coefficients for main effects are interpretable as the mean difference between sequence type or dimension group in empirical logit space. For example, the positive Different.vs.Rank coefficient indicates that *Rank* sequence ratings were higher than *Different* ratings. Generally, contrasts are indicated with 'vs.' Multiplicative interactions are indicated with a colon (:), with the exception of the lower-order term 'Modality.by.DimType' (sensory modality by dimension type) that completes the contrast coding scheme for Dimension. Results discussed in the main text are indicated in bold.

Experiment 1: Regression Model					
Coefficient	*B*	*SE*	*df*	*t*	*p*
(Intercept)	0.11179	0.0146	60.87	7.659	1.72E−10***
Different.vs.Rank	0.17395	0.02733	87.74	6.365	8.66E−09***
Rank.vs.Ratio	0.10781	0.02176	131.22	4.954	2.20E−06***
Ratio.vs.Same	0.3557	0.03213	72.94	11.071	2.00E−16***
Audition.vs.Vision	0.1043	0.05838	60.87	1.787	0.079·
Qualitative.vs.Quantitative	0.07625	0.05838	60.87	1.306	0.19649
Modality.by.DimType	0.2391	0.23354	60.87	1.024	0.30997
Different.vs.Rank:Audition.vs.Vision	0.24838	0.10933	87.74	2.272	0.02554*
Rank.vs.Ratio:Audition.vs.Vision	0.12604	0.08704	131.22	1.448	0.15
Ratio.vs.Same:Audition.vs.Vision	−0.15179	0.12851	72.94	−1.181	0.24139
Different.vs.Rank:Qualitative.vs.Quantitative	0.17407	0.10933	87.74	1.592	0.11493
Rank.vs.Ratio:Qualitative.vs.Quantitative	0.18651	0.08704	131.22	2.143	0.03398*
Ratio.vs.Same:Qualitative.vs.Quantitative	−0.35541	0.12851	72.94	−2.766	0.00719**
Different.vs.Rank:Modality.by.DimType	0.59064	0.4373	87.74	1.351	0.18028
Rank.vs.Ratio:Modality.by.DimType	0.03223	0.34816	131.22	0.093	0.92639
Ratio.vs.Same:Modality.by.DimType	0.12987	0.51405	72.94	0.253	0.80126
Experiment 1: ANOVA Summary					
Factor	*SS*	*MS*	*df*	*F*	*p*
SequenceType	14.8498	4.9499	83.718	86.152	2.00E−16***
Dimension	0.3661	0.122	60.953	2.124	0.10639
SequenceType:Dimension	1.3536	0.1504	83.843	2.618	0.01023*

B = slope, *SE* = standard error, *SS* = sum of squares, *MS* = mean square, *df* = degrees of freedom.
* Significance level: $p < 0.05$.
** Significance level: $p < 0.01$.
*** Significance level: $p < 0.001$.
· Significance level: $0.05 < p \leq 0.1$.

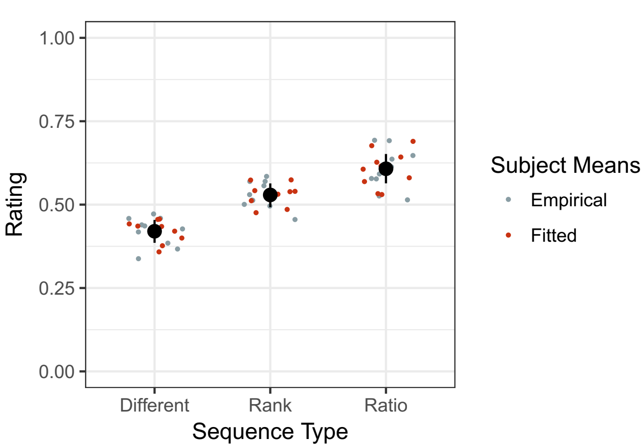

Fig. 3. Supplementary *Brightness* dimension data. See Fig. 2 for details on calculation of subject means and sequence-type means.

3.1.1.2. Stimuli. Stimuli were generated in the same manner as in Experiments 1a and 1b.

There were 4, between-subject stimulus cross-dimension conditions: object-height to pitch-height, object-size to loudness, and their opposite comparison orders.

There were three sequence-pair types per condition generated as in experiment 1a: *Ratio*, *Rank*, and *Different* sequence pairs. No *Same* sequence type was possible.

Stimulus values for comparison sequences were generated in the scale of the standard then linearly transformed to the comparison scales using the range of values and psychophysical scales explained in Section 2.2.

3.1.1.3. Procedure. The procedure was the same, except subjects first heard or saw a standard sequence and then a comparison sequence from the other modality. The full experiment consisted of two practice trials, 30 trials of each pair type at test, as well as 5 catch trials, yielding 97 total trials per subject.

3.1.2. Results

3.1.2.1. Inclusion of subjects. A priori, we excluded subjects based on (1) performance in catch trials as in Experiment 1 and (2) whether or not each subject rated *Different* sequences significantly *below* the slider midpoint in independent, one-tailed, one-sample *t*-tests, independently validated in a set of pilot subjects. A total of 28 subjects passed criterion (Height/Brightness: $n = 11$ ($n = 6$ in Height first); Size/Loudness: $n = 17$ ($n = 8$ in Size first)).

3.1.2.2. Analytical approach. We followed a multilevel modeling approach with maximal random effects similar to experiment 1, with difference codes for sequence type and contrast codes for dimension-pair condition. For dimension pair, we coded for which modality was presented first, whether the pair contained qualitative or quantitative dimensions, and their interaction.

3.1.2.3. Model results. Detailed results are presented in Table 2 and the data in Fig. 4. The model revealed a main effect of sequence type, but no main effects of dimension pair or a sequence-type-by-dimension-pair interaction. As predicted, coefficients corresponding to the differences between *Different* and *Rank* conditions (Different.vs.Rank coefficient) as well as the *Rank* and *Ratio* conditions (Rank.vs.Ratio coefficient) were significantly above 0.

3.1.3. Interim discussion

The results of this experiment support the conclusion that subjects can use abstract ratio and rank relations to compare

Table 2
Cross-modality comparisons. As with Table 1, the top portion indicates the results for each coefficient while the bottom indicates the associated ANOVA table summarizing each batch of coefficients. Coefficient names including 'vs.' indicate a group contrast. The additional coefficient, *DimType.by.Order*, reflects the lower-order interaction between dimension type (qualitative vs. quantitative) and order of modality presentation (auditory vs. visual).

Experiment 2a: Regression Model					
Coefficient	*B*	*SE*	*df*	*t*	*p*
(Intercept)	0.09044	0.01711	29.72	5.287	1.06E−05***
Different.vs.Rank	0.31876	0.03533	32.58	9.024	2.23E−10***
Rank.vs.Ratio	0.24518	0.0239	48.6	10.259	9.41E−14***
Qualitative.vs.Quantitative	0.07684	0.06842	29.72	1.123	0.27
AuditoryFirst.vs.VisualFirst	0.06527	0.06842	29.72	0.954	0.348
DimType.by.Order	−0.13366	0.2737	29.72	−0.488	0.629
Different.vs.Rank:Qualitative.vs.Quantitative	0.13083	0.1413	32.58	0.926	0.361
Rank.vs.Ratio:Qualitative.vs.Quantitative	0.13453	0.0956	48.6	1.407	0.166
Different.vs.Rank:AuditoryFirst.vs.VisualFirst	0.05606	0.1413	32.58	0.397	0.694
Rank.vs.Ratio:AuditoryFirst.vs.VisualFirst	0.05903	0.0956	48.6	0.617	0.54
Different.vs.Rank:DimType.by.Order	0.11134	0.56521	32.58	0.197	0.845
Rank.vs.Ratio:DimType.by.Order	−0.26768	0.3824	48.6	−0.7	0.487
Experiment 2a: ANOVA Summary					
Factor	*SS*	*MS*	*df*	*F*	*p*
SequenceType	10.966	5.483	39.074	85.613	5.22E−15***
Condition	0.1355	0.0452	29.904	0.705	0.5563
SequenceType:Condition	0.2211	0.0369	39.371	0.575	0.7474

B = slope, *SE* = standard error, *SS* = sum of squares, *MS* = mean square, *df* = degrees of freedom.
* Significance level: $p < 0.05$.
** Significance level: $p < 0.01$.
*** Significance level: $p < 0.001$.
· Significance level: $0.05 < p \leq 0.1$.

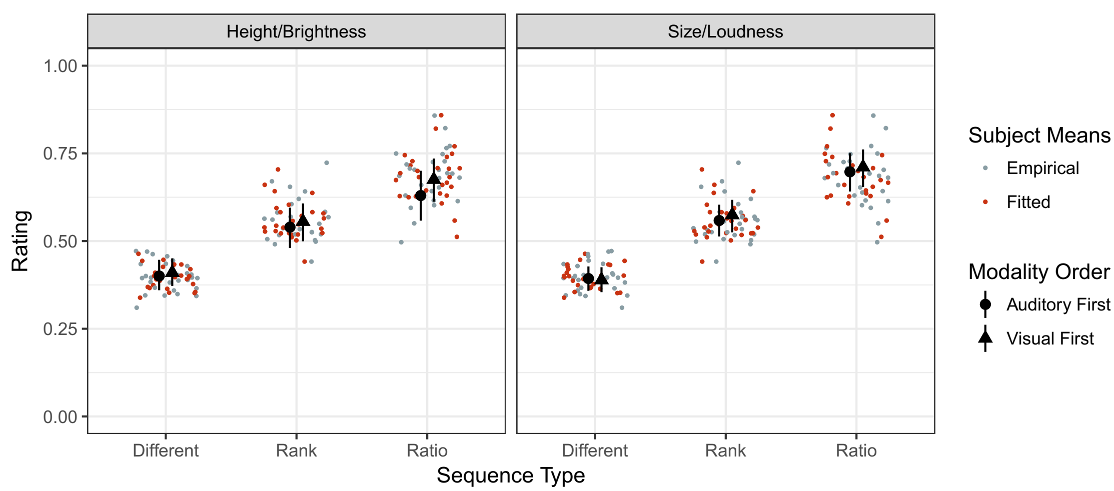

Fig. 4. Cross-modality comparisons. See Fig. 2 for details on calculation of means. Here the overall means are additionally broken down by the order in which each modality was presented.

sequences across sensory modalities. In addition, as in Experiment 1, subjects are sensitive to the differences in the amount of information preserved between standard and comparison stimuli in the *Ratio* and *Rank* sequence types, suggesting a spontaneous, modality-general, ratio representation.

3.2. Experiment 2b

Sequences in this experiment could be squares of different sizes, tones synthesized in the browser paired with a concurrent visual stimulus at varying durations, or Arabic numerals. We used 2-digit Arabic numerals rather than dot arrays to minimize perceptual differences among stimuli and to avoid known difficulties with counterbalancing non-numerical stimulus features, which would make the experiment prohibitively long.

3.2.1. Method

3.2.1.1. Subjects. We recruited 20 subjects on Mechanical Turk for the size-number and size-duration conditions, 10 for each order of dimension presentation Because the size-duration condition took longer than expected for subjects to complete, for the number-duration condition we recruited 30 subjects, with 15 experiencing each order, allowing a reduction in the number of trials within-subject (see *Procedure* below). All other recruitment specifications and payment procedures were identical to previous experiments.

3.2.1.2. Stimuli. Numbers were Arabic numerals ranging from 5 to 50 selected randomly from a uniform distribution on a logarithmic scale and rounded to the nearest integer. They were presented in a bluish color (hexadecimal code #0066FF) in the center of the screen for 500 ms each.

Duration stimuli were presented bimodally to reduce task difficulty: they consisted of (1) a red circle with a 25-px radius in the center of the stimulus window and a synchronized sinusoid of 440 Hz synthesized in the browser, with 5-ms ramp-up and ramp-down times. Duration intervals were restricted to last between 500 and 5000 ms on a logarithmic scale.

Size stimuli were created with the same restrictions used in previous experiments.

3.2.1.3. Procedure. The procedure was the same as in previous experiments. For the size-number and size-duration conditions, there were 25 stimuli for each sequence type, 2 practice trials, and 5 catch trials (82 total). In the number-duration condition, we decreased the number of trials per sequence type to 12 to decrease the overall duration of the experiment to approximately match the size-number condition, as mentioned above in response to the unexpected duration of the size-duration condition (43 total).

3.2.2. Results

3.2.2.1. Subject inclusion. The small number of trials per subject in the *Different* sequence type for the number-duration condition obviated our previous exclusion criteria, so we only included subjects on the basis of the catch-trial criterion from the previous experiments. The number of subjects passing criterion was 60 (Number-Duration: $n = 24$, Size-Duration: $n = 16$, Size-Number: $n = 20$).

3.2.2.2. Analytical approach. We applied the same multilevel modeling approach with maximal random effects. Sequence type was difference coded. Condition was simple coded; the number-duration condition served as the (arbitrary) reference group. Condition was pooled across orders of presentation.

3.2.2.3. Model results. The detailed results are indicated in Table 3 and data displayed in Fig. 5. The model indicated a significant main effect of sequence type, with no main effect of condition or sequence-type-by-condition interaction. The individual coefficients for sequence type indicated significant differences between the *Different* and *Rank* sequence ratings as well as the *Rank* and *Ratio* ratings (Different.vs.Rank coefficient and Rank.vs.Ratio coefficient).

Though the set of interaction coefficients did not account for a significant proportion of variance in the model, indicating that between-sequence-type comparisons did not vary significantly across conditions, one of the interaction coefficients indicated that the difference between *Different* and *Rank* ratings was slightly larger for the size-number condition than for the number-duration condition. The simple slope for the Different vs. Rank coefficient in the number-duration condition was still significantly above 0 [$B = 0.2046$, $SE = 0.0294$, 95% CI = {0.1470, 0.2621}], corroborating the fact that this condition followed the same pattern of results as the other two conditions.

3.3. Interim discussion

These experiments demonstrated sensitivity to the difference between *Rank* and *Ratio* and between *Different* and *Rank* sequences when comparing across sensory modalities and across canonical magnitude domains. We interpret these results as consistent with the hypothesis that a domain-general encoding of ratios and ranks supports comparisons of magnitude sequences. Ratios preserve more within-dimension information about a sequence than ranks only; thus, similarity ratings are higher when ratio information is preserved across sequences compared to preserved rank alone.

The set of Sequence-Type-by-Condition interaction terms did not account for a significant proportion of variance in the model, though examination of the individual coefficients in the set of interaction terms revealed a difference between the number-duration condition and the size-number condition in Experiment 2b for discriminating between the *Different* and *Rank* sequences. This indicates that the difference is likely a result of noise or a statistical artifact of greater shrinkage toward the overall mean due to fewer trials in the Number-Duration condition. In other words, the effect of partial pooling is drawing the cells with comparatively fewer data points more strongly to the overall mean. However, if the difference is not just a statistical artifact, then it is likely due to the high working memory demands involving duration stimuli. Both sets of results involving duration stimuli appear quite similar. Regardless, these proposed explanations do not conflict with the overall result, which extends the finding from Experiment 2a.

Table 3
Cross-dimension comparison model results: space, time, and number. As with previous tables, 'vs.' indicates a contrast. Backslashes separate the individual dimensions of a particular pair. Items in bold are discussed in the text.

Experiment 2b: Regression Model					
Coefficient	*B*	*SE*	*df*	*t*	*p*
(Intercept)	0.053096	0.013859	75.29	3.831	0.000262***
Different.vs.Rank	0.18547	0.028996	83.25	6.396	8.81E−09***
Rank.vs.Ratio	0.195284	0.026704	88.14	7.313	1.14E−10***
Number/Duration.vs.Size/Duration	0.052115	0.035081	82.21	1.486	0.141225
Number/Duration.vs.Size/Number	0.004824	0.03177	72.57	0.152	0.879732
Different.vs.Rank:Number/Duration.vs.Size/Duration	0.051003	0.074081	92.69	0.688	0.492875
Rank.vs.Ratio:Number/Duration.vs.Size/Duration	0.026117	0.068649	99.01	0.38	0.704426
Different.vs.Rank:Number/Duration.vs.Size/Number	0.147571	0.066247	79.13	2.228	0.028747*
Rank.vs.Ratio:Number/Duration.vs.Size/Number	0.003755	0.060866	82.96	0.062	0.950962
Experiment 2b: ANOVA Summary					
Factor	*SS*	*MS*	*df*	*F*	*p*
SequenceType	4.9314	2.4657	82.119	38.013	2.06E−12***
Condition	0.1662	0.0831	74.617	1.281	0.2838
SequenceType:Condition	0.3834	0.09586	80.494	1.478	0.2166

B = slope, *SE* = standard error, *SS* = sum of squares, *MS* = mean square, *df* = degrees of freedom.
* Significance level: $p < 0.05$.
** Significance level: $p < 0.01$.
*** Significance level: $p < 0.001$.
· Significance level: $0.05 < p \leq 0.1$.

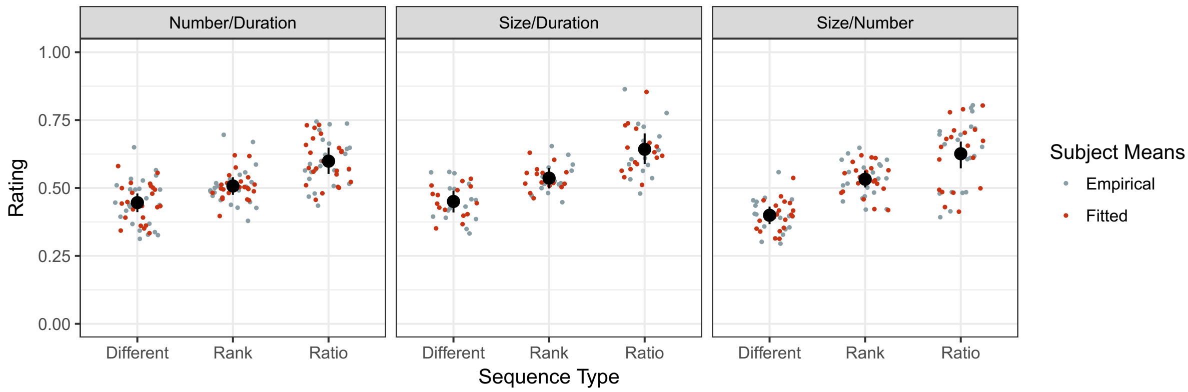

Fig. 5. Cross-dimension comparisons: space, time, and number. See Fig. 2 for details on mean calculations. Panels reflect the particular dimension pairs, as indicated in the panel label.

4. Omnibus analysis

To compare within-dimension and across-dimension ratings, we ran an omnibus analysis in which we combined all data, excluding *Same* trials from Experiment 1. As with previous analyses, sequence type was difference-coded. We simple-coded for 3 condition types: *Within* conditions that included *Same* trials (labeled Within(+Same) in the coefficient names), the *Within* condition that excluded *Same* trials (labeled Within(ØSame) in the coefficient names), and *Across* conditions. In addition to the random effects previously included, we included a random intercept by condition.

Results are presented in Table 4. The analysis shows the following: (1) the expected main effect of Sequence Type, indicated in the SequenceType row of the ANOVA table and the Different.vs.Rank and Rank.vs.Ratio coefficients; (2) a main effect of Condition, driven by lower overall ratings in the *Within* conditions that included *Same* stimuli, as indicated by the Condition row in the ANOVA table and the Across.vs.Within(+Same) coefficient; and (3) an interaction between Sequence Type and Condition driven by the smaller differences in ratings for different sequence types in the *Within* conditions that included *Same* trials, as indicated by the significant Rank.vs.Ratio:Across.vs.Within(+Same) and the marginal Different.vs.Rank:Across.vs.Within(+Same) coefficients.

In addition, the model failed to detect an interaction between Sequence Type and condition for the *Within* (not including *Same*) *vs. Across* comparison, as indicated by the non-significant Different.vs.Rank:Across.vs.Within(ØSame) and Rank.vs.Ratio:Across.vs.Within(ØSame) coefficients. This suggests that there might be no inherent difference in the strength of information transfer between within-dimension and cross-dimension comparisons.

Overall the omnibus analysis revealed a global pattern of sensitivity to rank and ratio during sequence comparison across sequence types and modalities.

5. General discussion

We have demonstrated that subjects can compare sequences on the basis of the precision of relative-magnitude information preserved between sequence pairs. Specifically, patterns that preserved inter-item ratio information were rated as more similar than patterns that only preserved inter-item rank information, a distinction not required by the task but spontaneously imposed by subjects. This is consistent with the hypothesis that a dimen-

Table 4
Omnibus analysis results. As with previous tables, the top portion 'vs.' indicates a contrast and ':' indicates a multiplicative interaction. Within(+Same) indicates within-dimension conditions that included *Same* trials, while Within(ØSame) indicates within-dimension conditions that did not. Coefficients in bold are discussed in more detail in the main text.

Omnibus Regression Model					
Coefficient	*B*	*SE*	*df*	*t*	*p*
(Intercept)	0.03122	0.01308	172.95	2.386	0.01811*
Different.vs.Rank	0.2143	0.02458	190.68	8.717	1.33E−15***
Rank.vs.Ratio	0.15879	0.01956	223.4	8.119	3.15E−14***
Across.vs.Within(Øsame)	−0.02686	0.03607	170.61	−0.745	0.45755
Across.vs.Within(+Same)	−0.07341	0.01989	198.44	−3.69	0.00029***
Different.vs.Rank:Across.vs.Within(ØSame)	−0.02687	0.06759	186.09	−0.397	0.69145
Rank.vs.Ratio:Across.vs.Within(ØSame)	−0.0532	0.05355	215.27	−0.993	0.32159
Different.vs.Rank:Across.vs.Within(+Same)	−0.06951	0.03802	229.01	−1.828	0.0688·
Rank.vs.Ratio:Across.vs.Within(+Same)	−0.1022	0.03104	277.95	−3.293	0.00112**
Omnibus ANOVA Summary					
Factor	*SS*	*MS*	*df*	*F*	*p*
SequenceType	7.1494	3.5747	215.45	58.12	2.20E−16***
ConditionType	0.8377	0.4189	181.98	6.81	0.001406**
SequenceType:ConditionType	0.786	0.1965	227.8	3.195	0.014067*

B = slope, *SE* = standard error, *SS* = sum of squares, *MS* = mean square, *df* = degrees of freedom.
* Significance level: $p < 0.05$.
** Significance level: $p < 0.01$.
*** Significance level: $p < 0.001$.
· Significance level: $0.05 < p \leq 0.1$.

sionless representation of a ratio scale supports the ability to reason about abstract magnitudes within and between dimensions. In addition, subjects discriminate between *Same* and *Ratio* sequences, suggesting that absolute-magnitude information is not lost. Finally, subjects discriminate between *Different* and *Rank* sequences, indicating that subjects can make use of more impoverished relative magnitude representations to compare sequences.

Further, consistent with the speculation that the system of musical contour representations may be connected with generalized magnitude representations, subjects behaved similarly in the auditory domain, visual domain, across modalities, and across magnitude dimensions.

5.1. Reframing the debate on generalized magnitude

Past authors have proposed that different kinds of quantitative representations share a common, analog-magnitude code (eg., Gallistel & Gelman, 2000; Holyoak & Glass, 1978; Walsh, 2003), and that dimensions as different as size and brightness could be compared, without specifying how such a code could take a highly general format. Our study demonstrates that a plausible representation that is meaningful across dimensions and sensory modalities is a ratio code. Subjects spontaneously represent changes in the absolute values along each dimension on an internal ratio scale and compare those relative values.

Consistent with previous studies, rank orderings also qualify as an additional candidate for a generalized magnitude representation, though subjects are aware that some relative-magnitude information is lost in this kind of representation. This suggests that ratio scaling may be the primary means of cross-dimension comparisons, while ranks are used when that information is unavailable. Ratios may be especially important to track across dimensions because they are preserved under linear scaling; for example, stimulus dimensions that undergo the same scaling transformation may be causally related. This is true of processes with an approximately constant or constant-on-average rate, such as the relationship between distance traveled and trip duration.

While these dimensionless representations explain transfer of information across the dimensions tested here, future work will need to identify the limits of their scope, as we have not shown in what domains they might fail. For instance, would other ordered symbol sequences such as letters elicit similar behavior? While Arabic numerals and letters share rank-orderings that can be mapped onto spatial locations (Gevers, Reynvoet, & Fias, 2003) and classic behavioral signatures such as the comparison distance effect (Jou & Aldridge, 1999; van Opstal, Gevers, de Moor, & Verguts, 2008), the representations of these sequences dissociate in some crucial task variants (Cheung & Lourenco, 2016; van Opstal & Verguts, 2011). Thus, it is unclear whether the cross-sequence comparison behavior reported here would extend to letters.

In addition, future work will need to identify how these representations relate to other characterizations of a generalized magnitude system in the literature. In particular, future studies need to (1) explore their contribution to bias effects in dual-magnitude-judgment tasks (eg., Casasanto & Boroditsky, 2008; Merritt, Casasanto, & Brannon, 2010; Starr & Brannon, 2016), (2) their contribution to binding together two or more magnitude measurements in memory (eg., Srinivasan & Carey, 2010) and to cross-dimension priming effects (Lourenco, Ayzenburg, & Lyu, 2016), and (3) whether ratios are calculated in a centralized module.

To explain bias effects in dual-tasks, some authors have suggested that a generalized magnitude system could function as a cross-dimension, cue-combination mechanism (Lambrechts, Walsh, & van Wassenhove, 2013), with more reliable dimensions receiving more weight. While this is possible in principle, it is also possible that idiosyncratic mechanisms may contribute to the strength and asymmetry of interactions between different pairs of simultaneously presented dimensions. Srinivasan and Carey (2010) discuss a distinction between the superset of *structurally similar* dimensions, which share descriptive characteristics, and a subset of those called *functionally overlapping* dimensions, which share an additional, privileged relationship in the environment (or in the brain) that could facilitate mapping and binding of representations in two, simultaneously presented dimensions.

A related distinction exists between the problem of transferring relational information across dimensions via analogical reasoning and the problem of information integration across two or more causally related dimensions. Unlike functional overlap, the latter does not necessarily imply the former; for instance, mappings between absolute magnitudes may not be related to analogical relationships between the two dimensions' abstract structure. The problems *can* intersect, but that does not also imply that a single system handles each instance where they intersect in the same way. Inferences about causal structures in the environment that generate mappings among proportions may require different internal models. From this point of view, a single magnitude system may be unsuitable for solving many different types of information-integration problems. Subjects may use all of the above representations and computational strategies across the different tasks they encounter in the environment, including ratio-scaled comparisons, direct mappings between absolute magnitudes, statistical inference in cue integration, and analogical reasoning (Bonn & Cantlon, 2012). Thus, the so-called "generalized magnitude system" may not be one coherent representational system, but rather a set of relational mapping phenomena.

Similarly, a common coding scheme that facilitates cross-dimension mappings need not have a centralized representation. Future research should carefully separate questions posed at the level of implementation—for example, whether generalized magnitude representations converge in a centralized module or not—from questions posed at the level of computation, which include the computation of abstract quantities.

5.2. Conclusion

The idea of a generalized magnitude system has often been invoked as an explanation for interactions between dimensions in adults and increasingly in infants and children (de Hevia, Izard, Coubart, Spelke, & Streri, 2014; Lourenco & Bonny, 2016; Newcombe, Levine, & Mix, 2015), but currently there is little evidence for specific, positive claims about candidate formats. The experiments presented in this paper fill in part of the hypothesis space by specifying and verifying one plausible type of representation, and thus represent an important step in building a more precise characterization of generalized magnitudes.

Acknowledgments

The authors wish to thank Richard Aslin, Steven Piantadosi, and Véronique Izard for their comments. An earlier version of this work was submitted to the University of Rochester as a portion of C.D. B.'s dissertation. The research was partially supported by an NSF-GRFP, NSF-GROW, and STEM-Chateaubriand to C.D.B. and NSF 1459625 and NIH R01HD064636 to J.F.C.

Appendix A. Supplementary material

The dataset and scripts required to run the experiments and analyses reported in this paper are freely available at https://osf.io/vtm8r, hosted by the Open Science Framework, along with instructions for either accessing a full, running version of each condition or for implementing replications. The table of simple effects from Experiment 1a as well as additional methodological details are available in the Supplementary Materials. Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2017.07.012.

References

Balci, F., & Gallistel, C. R. (2006). Cross-domain transfer of quantitative discriminations: Is it all a matter of proportion? *Psychonomic Bulletin & Review, 13*(4), 636–642. http://dx.doi.org/10.3758/BF03193974.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory & Language, 68*(3), 255–278. http://dx.doi.org/10.1016%2Fj.jml.2012.11.001.

Bonn, C. D., & Cantlon, J. F. (2012). The origins and structure of quantitative concepts. *Cognitive Neuropsychology, 29*(1–2), 149–173. http://dx.doi.org/10.1080%2F02643294.2012.707122.

Brosch, M., Selezneva, E., Bucks, C., & Scheich, H. (2004). Macaque monkeys discriminate pitch relationships. *Cognition, 91*, 259–272. http://dx.doi.org/10.1016/j.cognition.2003.09.005.

Bueti, D., & Walsh, V. (2009). The parietal cortex and the representation of time, space, number, and other magnitudes. *Philosophical Transactions of the Royal Society B, 364*, 1831–1840. http://dx.doi.org/10.1098/rstb.2009.0028.

Cantlon, J. F., Platt, M. L., & Brannon, E. M. (2009). Beyond the number domain. *Trends in Cognitive Sciences, 13*(2), 83–91. http://dx.doi.org/10.1016/j.tics.2008.11.007.

Casasanto, D., & Boroditsky, L. (2008). Time in the mind: Using space to think about time. *Cognition, 106*, 579–593. http://dx.doi.org/10.1016/j.cognition.2007.03.004.

Cheung, C.-N., & Lourenco, S. F. (2016). The associations between space and order in non-numerical sequences. *Consciousness and Cognition, 45*, 124–134. http://dx.doi.org/10.1016/j.concog.2016.08.013.

Cohen Kadosh, R., Lammertyn, J., & Izard, V. (2008). Are numbers special? An overview of chronometric, neuroimaging, developmental and comparative studies of magnitude representation. *Progress in Neurobiology, 84*(2), 132–147. http://dx.doi.org/10.1016/j.pneurobio.2007.11.001.

de Hevia, M. D., Izard, V., Coubart, A., Spelke, E. S., & Streri, A. (2014). Representations of space, time and number in neonates. *Proceedings of the National Academy of Sciences, 111*(13), 4809–4813. http://dx.doi.org/10.1073/pnas.1323628111.

de Hevia, M. D., & Spelke, E. S. (2010). Number-space mapping in human infants. *Psychological Science, 21*, 653–660. http://dx.doi.org/10.1177%2F0956797610366091.

Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, & pitch recognition in memory for melodies. *Journal of the Acoustical Society of America, 49*(2), 524. http://dx.doi.org/10.1121/1.1912382.

Gallistel, C. R., & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in Cognitive Sciences, 4*(2), 59–65. http://dx.doi.org/10.1016/S1364-6613(99)01424-2.

Gati, I., & Tversky, A. (1982). Representations of qualitative and quantitative dimensions. *Journal of Experimental Psychology: Human Perception and Performance, 8*(2), 325–340. http://dx.doi.org/10.1037//0096-1523.8.2.325.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York: Cambridge.

Gevers, W., Reynvoet, B., & Fias, W. (2003). The mental representation of sequences is spatially organized. *Cognition, 87*(3), B87–95. http://dx.doi.org/10.1016/S0010-0277(02)00234-2.

Holyoak, K. J., & Glass, A. L. (1978). Recognition confusion among quantifiers. *Journal of Verbal Learning and Verbal Behavior, 17*, 249–264. http://dx.doi.org/10.1016/S0022-5371(78)90165-2.

Jacob, S. N., & Nieder, A. (2009). Tuning to non-symbolic proportions in the human frontoparietal cortex. *European Journal of Neuroscience, 30*(7), 1432–1442. http://dx.doi.org/10.1111/j.1460-9568.2009.06932.x.

Jacob, S. N., Vallentin, D., & Nieder, A. (2012). Relating magnitudes: The brain's code for proportions. *Trends in Cognitive Sciences, 16*(3), 157–166. http://dx.doi.org/10.1016/j.tics.2012.02.002.

Jou, J., & Aldridge, J. W. (1999). Memory representation of alphabetic position and interval information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(3), 680–701. http://dx.doi.org/10.1037/0278-7393.25.3.680.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). *lmerTest: Tests in linear mixed effects models. R package version 2.0-30*. Retrieved from <http://CRAN.R-project.org/package=lmerTest>.

Lambrechts, A., Walsh, V., & van Wassenhove, V. (2013). Evidence accumulation in the magnitude system. *PLoS ONE, 8*(12), e82122. http://dx.doi.org/10.1371/journal.pone.0082122.

Lourenco, S. F., Ayzenburg, V., & Lyu, J. (2016). A general magnitude system in human adults: Evidence from a subliminal priming paradigm. *Cortex, 81*, 93–103. http://dx.doi.org/10.1016/j.cortex.2016.04.013.

Lourenco, S. F., & Bonny, J. W. (2016). Representations of numerical and non-numerical magnitude both contribute to mathematical competence in children. *Developmental Science*. http://dx.doi.org/10.1111/desc.12418.

Lourenco, S. F. (2015). On the relation between numerical and non-numerical magnitudes: Evidence for a general magnitude system. In D. C. Geary, D. B. Berch, & K. M. Koepke (Eds.), *Evolutionary origins and early development of number processing* (pp. 145–174). Amsterdam: Elsevier.

Lourenco, S. F., & Longo, M. R. (2010). General magnitude representation in human infants. *Psychological Science, 21*(6), 873–881. http://dx.doi.org/10.1177/0956797610370158.

Luce, R. D. (1990). "On the possible psychophysical laws" revisited: Remarks on cross-modal matching. *Psychological Review, 97*(1), 66–77. http://dx.doi.org/10.1037/0033-295X.97.1.6.

Luce, R. D. (2002). A psychophysical theory of intensity proportions, joint presentations, and matches. *Psychological Review, 109*(3), 520–532.

Marvin, E. W. (1997). Tonal/atonal: Cognitive strategies for recognizing transposed melodies. In Beach, Baker, & Bernard (Eds.), *Music theory in concept and practice*. Rochester, NY: University of Rochester.

Marvin, E. W., & Laprade, P. A. (1987). Relating musical contours: Extensions of a theory for contour. *Journal of Music Theory, 31*, 225–267. Retrieved from <http://www.jstor.org/stable/843709>.

McDermott, J. H., Lehr, A. J., & Oxenham, A. J. (2008). Is relative pitch specific to pitch? *Psychological Science, 19*(12), 1263–1271. http://dx.doi.org/10.1111/j.1467-9280.2008.02235.x.

Merritt, D. J., Casasanto, D., & Brannon, E. M. (2010). Do monkeys think in metaphors? Representations of space and time in monkeys and humans. *Cognition, 117*, 191–202. http://dx.doi.org/10.1016/j.cognition.2010.08.011.

Newcombe, N. S., Levine, S. C., & Mix, K. S. (2015). Thinking about quantity: The intertwined development of spatial and numerical cognition. *WIREs Cognitive Science, 6*, 491–505. http://dx.doi.org/10.1002/wcs.1369.

Pinel, P., Piazza, M., Le Bihan, D., & Dehaene, S. (2004). Distributed and overlapping cerebral representations of number, size, and luminance during comparative judgments. *Neuron, 41*, 1–20. http://dx.doi.org/10.1016/S0896-6273(04)00107-2.

Prince, J. B., Schmuckler, M. A., & Thompson, W. F. (2009). Cross-modal melodic contour similarity. *Canadian Acoustics, 37*(1), 35–49. Retrieved from <https://jcaa.caa-aca.ca/index.php/jcaa/article/view/2114/1861>.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin, 2*, 110–114. http://dx.doi.org/10.2307/3002019.

Shepard, R. N. (1981). Psychological relations and psychophysical scales: On the status of "direct" psychophysical measurement. *Journal of Mathematical Psychology, 24*, 21–57. http://dx.doi.org/10.1016/0022-2496(81)90034-1.

Srinivasan, M., & Carey, S. (2010). The long and the short of it: On the nature and origin of functional overlap between representations of space and time. *Cognition, 116*(2), 217–241. http://dx.doi.org/10.1016%2Fj.cognition.2010.05.005.

Starr, A., & Brannon, E. M. (2016). Visuospatial working memory influences the interaction between space and time. *Psychonomic Bulletin & Review*. http://dx.doi.org/10.3758/s13423-016-1043-4.

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

Stevens, J. C., Mack, J. D., & Stevens, S. S. (1960). Growth of sensation on seven continua as measured by force of handgrip. *Journal of Experimental Psychology, 59*, 60–67 <http://psycnet.apa.org/doi/10.1037/h0040746>.

Trehub, S. E., Thorpe, L. A., & Morrongiello, B. A. (1987). Organizational processes in infants' perception of auditory patterns. *Child Development, 58*(3), 741–749. http://dx.doi.org/10.2307/1130211.

Vallentin, D., & Nieder, A. (2008). Behavioral and prefrontal representation of spatial proportions in the monkey. *Current Biology, 18*, 1420–1425. http://dx.doi.org/10.1016/j.cub.2008.08.042.

Vallentin, D., & Nieder, A. (2010). Representations of visual proportions in the primate posterior parietal and prefrontal cortices. *European Journal of Neuroscience, 32*, 1380–1387. http://dx.doi.org/10.1111/j.1460-9568.2010.07427.x.

van Opstal, F., Gevers, W., de Moor, W., & Verguts, T. (2008). Dissecting the symbolic distance effect: Comparison and priming effects in numerical and nonnumerical orders. *Psychonomic Bulletin & Review, 15*(2), 419–425. http://dx.doi.org/10.3758/PBR.15.2.419.

van Opstal, F., & Verguts, T. (2011). The origins of the numerical distance effect: The same-different task. *Journal of Cognitive Psychology, 23*(1), 112–120. http://dx.doi.org/10.1080/20445911.2011.466796.

Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space, and quantity. *Trends in Cognitive Sciences, 7*(11), 483–488. http://dx.doi.org/10.1016/j.tics.2003.09.002.