Acta Psychologica 144 (2013) 308–315

Contents lists available at ScienceDirect

Acta Psychologica

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

Comparative judgments of symbolic and non-symbolic stimuli yield different patterns of reaction times

Tali Leibovich [a,b,*], Sarit Ashkenazi [b], Orly Rubinsten [c], Avishai Henik [b]

[a] Department of Cognitive Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel
[b] Department of Psychology and the Zlotowski Center for Neuroscience, Ben-Gurion University of the Negev, Beer-Sheva, Israel
[c] Department of Learning Disabilities, University of Haifa, Israel

ARTICLE INFO

Article history:
Received 28 January 2013
Received in revised form 8 July 2013
Accepted 10 July 2013
Available online 7 August 2013

PsycINFO classification:
2300
2340

Keywords:
Ratio
Non-symbolic comparison
Numerical cognition

ABSTRACT

Are different magnitudes, such as Arabic numerals, length and area, processed by the same system? Answering this question can shed light on the building blocks of our mathematical abilities. A shared representation theory suggested that discriminability of all magnitudes complies with Weber's law. The current work examined this suggestion. We employed comparative judgment tasks to investigate different types of comparisons — conceptual comparison of numbers, physical comparison of numbers and physical comparison of different shapes. We used 8 different size ratios and plotted reaction time as a function of these ratios. Our findings suggest that the relationship between discriminability and size ratio is not always linear, as previously suggested; rather, it is modulated by the type of comparison and the type of stimuli. Hence, we suggest that the representation of magnitude is not as rigid as previously suggested; it changes as a function of task demands and familiarity with the compared stimuli.

1. Introduction

What is the relationship between the ability to distinguish between two numbers and the distance between them? This question attracted much attention in the literature of numerical cognition. Many suggested that this relationship obeys Weber's law and that numerical values are compared similarly to other dimensions (e.g., size). The current study examines these issues, reports deviations from Weber's law and suggests that different dimensions give rise to different comparative functions.

Moyer and Landauer (1967) asked adult participants to compare two Arabic numerals (to choose the numerically larger number). The authors plotted reaction time (RT) as a function of the numerical distance between the to-be-compared numbers and reported that the best fit to describe their data was the equation RT = K * log (larger/larger-smaller) (i.e., the Welford function), which accounted for 75% of the variance. Accordingly, the authors suggested that comparisons of numbers are made "… in much the same way that comparisons are made between physical stimuli such as loudness and length of lines" (p. 1520). The same methodology (i.e., comparative judgments, and plotting RT as a function of the distance or ratio between the compared stimuli) was employed by others and led to the conclusion that the pattern of results was compatible with Weber's law.

Weber's law states that $\Delta I/I = K$. That is, the amount necessary to detect a difference between two stimuli (e.g., ΔI) depends on the initial intensity of the stimulus (e.g., I). The ratio of the just noticeable difference (JND) to intensity is constant (e.g., K). To examine changes in the ability to discriminate between magnitudes, researchers use comparative judgment tasks and examine changes in performance — accuracy and speed of responding (RT) — as a function of the ratio between two magnitudes. The underlying assumption is that RT measures the ability to discriminate between two magnitudes. As such, RT should increase with increase in stimuli ratio because increase in stimuli ratio means increase in the similarity between the to-be-compared stimuli or difficulty to discriminate between them. When RT as a function of ratio was linear it was taken as an indication for compatibility with Weber's law. For example, comparisons of the conceptual size of pictures of objects (Paivio, 1975), words representing different animals (Moyer, 1973), and comparisons of dot arrays by monkeys and humans (Brannon, 2006). Note that all these studies used RT and discussed Weber's law. Hence, in the numerical cognition literature, RT is an acceptable measure of discriminability; Moyer (1973) cites Johnson's (1939) results revealing Weber's law performance in the comparison of two line lengths and using RT as the dependent measure, and Verguts, Fias, and Stevens (2005) cite a work by Festinger (1943) that discusses Weber's law in the context of RT experiments.

On the basis of this common ground, Cantlon, Platt, and Brannon (2009) suggested that all magnitudes are processed by the approximate number system (ANS), the hallmark of which is Weber's law. This shared representation and the compliance with Weber's law are highly

* Corresponding author at: Dept. of Cognitive Sciences, Ben-Gurion University of the Negev, P.O.B. 653, Beer-Sheva 84105, Israel. Tel.: +972 8 6477209; fax: +972 8 6472072.
E-mail address: labovich@post.bgu.ac.il (T. Leibovich).

0001-6918/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.actpsy.2013.07.010

acceptable principles in the numerical cognition literature and a large number of studies use these assumptions as their point of departure (Beran, Decker, Schwartz, & Schultz, 2011; Buhusi & Cordes, 2011; Droit-Volet, 2010; Möhring, Libertus, & Bertin, 2012; Piazza, 2010; Piazza, Izard, Pinel, Le Bihan, & Dehaene, 2004; Piazza et al., 2010; Tokita & Ishiguchi, 2011; Walsh, 2003).

However, this line of evidence has several shortcomings. First, some studies reported the distance effect, but examined only 2–3 different distances (Cohen Kadosh, Henik, & Rubinsten, 2008; Rubinsten & Henik, 2002; Vigliocco, Vinson, Damian, & Levelt, 2002). Under those conditions it is hard to find subtle differences between different magnitudes. Second, in many studies (e.g., Cantlon & Brannon, 2006; Cohen Kadosh et al., 2005; Fias, Lammertyn, Reynvoet, Dupont, & Orban, 2003; Piazza, 2010) the existence of the distance effect is taken as evidence for compliance with Weber's law. This is an inaccurate statement because Weber's law suggests not only that the discriminability depends on the ratio between the to-be-compared magnitudes, but also that this dependency is *linear*. In studies that tried to fit their data to a linear trend, the value of the fit to Weber's law was around 75–79% (Moyer, 1973; Moyer & Landauer, 1967 — fit to the Welford function; Paivio, 1969), while the fit to Weber's law in comparative judgment of line length was 99% (Moyer, 1973, citing Johnson, 1939). In all of those studies there are no reports of attempting to fit the results to functions other than linear. Thus, deviation from Weber's law is possible. The third shortcoming lies in the fact that most studies in the numerical cognition literature focused on *comparative judgments* of two numerosites or the conceptual size of symbolic stimuli, and compared their results to findings regarding physical sizes such as loudness, brightness, etc., found in *estimation* tasks. Given that different magnitudes were studied using different methodologies, it is problematic to suggest, for example, that comparisons of numbers are made similarly to comparisons between physical stimuli (Moyer & Landauer, 1967, p. 1520; see also Brannon, 2006, for a very similar suggestion).

The current study employed the same method — comparative judgments — with the aim to more accurately describe participants' performance while comparing different stimuli. Specifically, participants were asked to decide which of two stimuli was physically or conceptually larger. The stimuli were single-digit numbers that were compared according to their numerical value (conceptual comparison, e.g., 2 7), or their physical size (physical comparison, e.g., 2 2), or the stimuli were two identical punctuation marks (e.g., #, @, &, etc.) or identical Gibson figures (Gibson, Gibson, Pick, & Osser, 1962) — meaningless shapes that have the same visual complexity as numbers — that were compared according to their physical size. Every participant performed only one condition.

In line with studies mentioned above, we expected performance to comply with Weber's law. Namely, we expected that a constant increase in ratio between two numbers would result in a constant increase in RT. For example, if RT to the pair 2 4 (ratio of 0.5) is 300 ms, and RT to the pair 5 3 (ratio of 0.6) is 350 ms, then RT to the pair 7 5 (ratio of 0.7) is predictable — 400 ms — since for every 10% increment in the numerical ratio, RT increases by 50 ms (a constant). Hence, in the current study we plotted RT as a function of magnitude ratio and fitted it to the function $RT = ax^b + c$, where RT is a measure for discriminability, x is the ratio between the magnitudes (smaller divided by larger) and c is the minimal RT. If the relationship between RT and magnitude ratio is linear, as suggested by previous studies (e.g., Cantlon et al., 2009), and referred to as Weber's law, the exponent b should be 1. Exponent values other than 1 would indicate deviation from Weber's law. Larger exponents mean that the change in RT is not constant and cannot be predicted by a linear function. Plotting RT as a function of magnitude ratio (smaller/larger) has been done in several works. For example, Cantlon and Brannon (2006) had participants compare numerosity of dots (select the array with less dots) and plotted RT as a function of numerosity ratio. They concluded from that linear relationship that their pattern suggested Weber's law.

By finding the exponent (b) for each participant and using its value as a dependent variable, we were able to more thoroughly investigate whether type of stimuli (symbolic or non-symbolic) and type of comparison (physical or conceptual) modulated performance in a comparative judgment task.

The expected results according to the current literature are: (1) there would be no significant difference among exponents of different types of stimuli and comparisons, and (2) these exponents would not be significantly different from 1, suggesting a linear trend and compliance with Weber's law. However, if performance in comparative judgment tasks is modulated by the type of comparison and type of stimuli, we expect the exponents to be different from each other.

2. Experiment 1: conceptual comparison of Arabic numerals

2.1. Method

2.1.1. Participants

Fourteen volunteers (10 females, 4 males), first year students at Ben-Gurion University of the Negev, participated in the experiment for class credit. All participants were native Hebrew speakers and had intact or corrected vision.

2.1.2. Stimuli

Arabic numerals in black Ariel font were presented on a white background, in the same physical size. We manipulated the numerical ratio between the two numbers from 0.1 to 0.8. For example, the ratio of 0.5 was composed of the pairs (2 4), (3 6), etc. For every ratio, we used all the possible pairs (see Table 1). There were 6 pairs of numbers for every ratio. If the number of possible pairs per ratio was smaller than 6, some of the pairs were used more than once. Overall, across all ratios, all the numbers appeared a similar number of times.

2.1.3. Procedure

Participants were asked to decide, as quickly as possible while avoiding errors, which of the two Arabic numerals was numerically

Table 1
Pairs of stimuli by numerical ratio.

Category ratio	Ratio	Large number	Small number
0.1	0.11	9	1
	0.13	8	1
	0.14	7	1
0.2	0.2	5	1
	0.22	9	2
	0.25	4	1
	0.25	8	2
0.3	0.33	3	1
	0.33	6	2
	0.33	9	3
0.4	0.4	5	2
	0.43	7	3
	0.44	9	4
0.5	0.5	2	1
	0.5	4	2
	0.5	6	3
	0.5	8	4
0.6	0.6	5	3
	0.63	8	5
	0.67	3	2
	0.67	6	4
	0.67	9	6
0.7	0.71	7	5
	0.75	4	3
	0.75	8	6
0.8	0.8	5	4
	0.83	6	5
	0.86	7	6

Note. Ratio = (small number/large number) with an accuracy of 2 decimal places.

larger. They were asked to indicate their decision by pressing a key corresponding to the side of the larger number (q key on the left or p key on the right). Each trial began with a central fixation point presented for 300 ms. Five hundred milliseconds after the elimination of the fixation point, a pair of numerals appeared and remained in view until the participant pressed a key. The next trial started 500 ms after response onset (see Fig. 1A). A block of 18 practice trials was presented first, followed by three experimental blocks of 96 trials each (8 ratios × 6 pairs × 2 sides (larger number on the left vs. on the right)). The stimuli within the blocks appeared in a random order. The dependent measures were RT and error rates.

2.2. Results

For every participant we computed mean RT for the correct responses, and accuracy. RTs and accuracy were subjected to a one-way ANOVA (analysis of variance) with ratio as the dependent variable. The main effect of size ratio was significant, $F\ (7, 91) = 36.51$, $MSE = 462$, $p < .001$, $\eta^2 p = .74$; namely, RT increased with ratio. A similar analysis of error rates revealed that accuracy decreased with an increase in size ratio, $F\ (7, 91) = 8.27$, $MSE = .0006$, $p < .001$, $\eta^2 p = .39$. As can be seen in Fig. 1B, plotting RT as a function of conceptual size ratio and fitting the graph to the function $\mathrm{RT} = \mathrm{ax}^{\mathrm{b}} + \mathrm{c}$ yields a power function trend (b-value = 1.71, $r^2 = .97$).

Because RT was plotted as a function of ratio for every participant and fitted to a power function, each participant had an exponent value (b) for a given condition (e.g., conceptual comparisons). These b-values were subjected to a t-test to examine deviations from 1. One sample t-test revealed the exponent values were significantly different from 1, $t\ (13) = 3.45$, $p < .005$. A non-parametric sign test confirmed the significant difference, $z\ (14) = 3.47$ $p < .001$. As mentioned in the introduction, many researchers suggested that the relationship between RT and the ratio between magnitudes best fits a linear function. A way to examine this issue is to compare the variance generated by the two functions (i.e., linear function vs. power function). To investigate this, we fitted the results of every participant once to a linear function and once to a power function, and extracted the r^2 value. On average, the linear function explained only 79% of the variance in the data, whereas power functions accounted for 97% of the variance (Fig. 1C). This difference was found to be significant in a t-test for dependent samples, $t\ (13) = 3.6$, $p < .001$, $\eta^2 p = .5$, and in a non-parametric sign-test, $Z = 3.47$, $p < .001$, strengthening the suggestion regarding deviation from Weber's law. Note that the explained variance of the linear trend found here (0.79) is very close to the r^2 values reported by Moyer and Landauer (1967) and Moyer (1973) (0.75 and 0.793, respectively).

The plots of Experiment 1 were fitted once to a power function and once to a linear function. Although the plot better fitted a power function, fitting the plot to a linear function replicated former suggestions that numbers are compared similar to other physical continua. If this is indeed the case, we expect a similar trend when comparing the physical size of two identical numbers or shapes. Accordingly, in Experiments 2 and 3, participants compared the physical size of two identical numbers or Gibson figures, and the plots were fitted to a power function. The exponents of the functions were then compared to those of Experiment 1.

3. Experiments 2 and 3: physical comparisons of symbolic and non-symbolic stimuli

3.1. Method

3.1.1. Participants

Fourteen participants (5 males, 9 females) compared the physical size of symbolic stimuli (Arabic numerals), and fourteen (9 males, 5 females) compared the physical size of non-symbolic stimuli (Gibson figures). We used the same selection criteria as in Experiment 1, using students who had not participated in the previous experiments.

3.1.2. Stimuli

For the physical comparisons of symbolic stimuli, we used the numerals 1–9. In every trial, the same number appeared in different physical sizes (e.g., 2 2) — see examples in Figs. 2A and 3A. For the non-symbolic stimuli, we used 9 different Gibson figures (Gibson et al., 1962). These stimuli have only one dimensions — physical size. Each image (Arabic numeral or Gibson figure) was presented in 9 different physical sizes (heights of 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4 and 4.5 cm) to create 8 different size ratios, from 0.1 to 0.8. For example, to create the physical ratio of 0.5 we used the sizes (0.5 cm, 1 cm), (2 cm, 4 cm), etc. There were 6 pairs of physical sizes for every ratio. If the number of possible pairs per ratio was smaller than 6, some of the pairs were used more than once (see Table 2). Overall, across all ratios, all the physical sizes appeared a similar number of times.

3.1.3. Procedure

The procedure was similar to that of Experiment 1. Participants were asked to decide, as quickly as possible while avoiding errors, which stimulus was physically larger. A block of 18 practice trials was presented first, followed by three experimental blocks of 96 trials each [8 ratios × 6 pairs of different sized stimuli × 2 sides (larger image on the left vs. on the right)]. The stimuli within the block appeared in a random order. The dependent measures were RT and error rates.

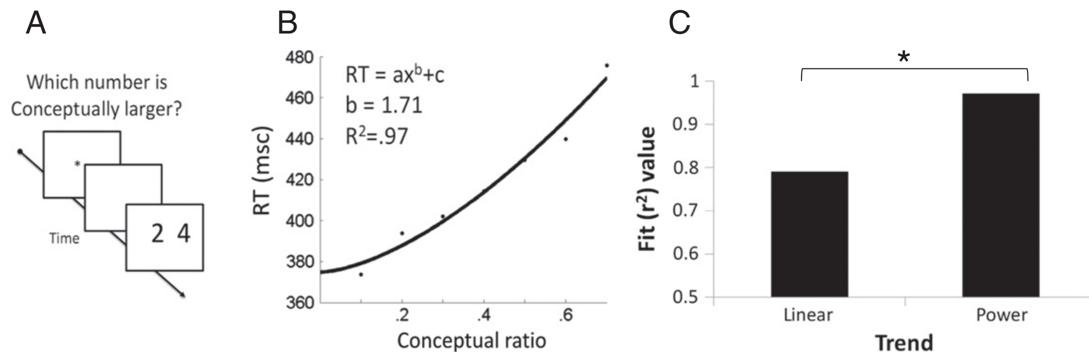

Fig. 1. Comparative judgments of conceptual comparisons of Arabic numerals. A) Procedure. Participants indicated the conceptually larger numeral by pressing a key. B) RT as a function of the conceptual ratio. This plot was fitted to a power function. C) Fit comparisons. The same RTs were fitted once to a linear trend (similar to previous studies) and once to a power function trend. r^2 values were significantly higher for the power function trend.

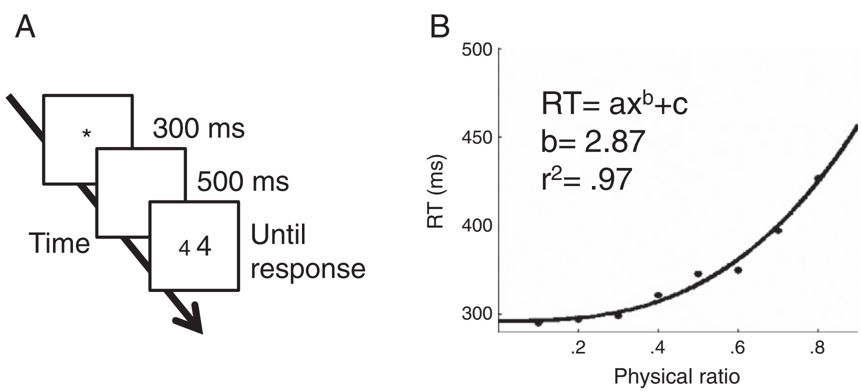

Fig. 2. Comparative judgments of physical size of numbers. A) Procedure. Participants indicated the physically larger numeral by pressing a key. B) RT as a function of the physical ratio of the numerals. This plot was fitted to a power function.

3.2. Results

For every participant we computed mean RT for the correct responses, and accuracy. RTs and accuracy were subjected to a one-way ANOVA with ratio as the dependent variable. RT analyses revealed significant main effects of size ratio, $F\ (7,\ 91) = 25.6$, $MSE = 441$, $p < .001$, $\eta^2p = .66$, and $F\ (7,\ 91) = 30.63$, $MSE = 620$, $p < .001$, $\eta^2p = .7$, for symbolic (numbers) and non-symbolic (Gibson figures) stimuli, respectively, suggesting that RT increased with an increase in size ratio. A similar analysis of error rates revealed that accuracy decreased with an increase in size ratio, $F\,(7, 91) = 5.17$, $MSE = .0008$, $p < .001$, $\eta^2p = .28$ and $F\ (7,\ 91) = 7.48$, $MSE = .0008$, $p < .001$, $\eta^2p = .37$, for symbolic and non-symbolic stimuli, respectively. Similar to Experiment 1, we computed b-values for each participant in each condition. As can be seen in Figs. 2B and 3B, plotting RT as a function of ratio yields power function trends; for symbolic comparisons $b = 2.87$, $r^2 = .97$ and for non-symbolic comparisons $b = 5.38$, $r^2 = .96$. One sample t-tests revealed that the exponent values were significantly different from 1, $t\,(13) = 4.01$, $p < .005$ and $t\,(13) = 7.45$, $p < .001$, for symbolic and non-symbolic comparisons, respectively.

4. Experiment 4: physical comparison of punctuation marks

Gibson figures were chosen as non-symbolic stimuli, due to their visual similarity to numbers and their lack of conceptual size. However, there is one important difference between numbers and Gibson figures — numbers are very familiar shapes whereas Gibson figures are novel and unfamiliar stimuli. Therefore, it is possible that the large exponent of the power function for Gibson figures was due to the lack of familiarity. To examine this possibility, another group of participants compared the physical sizes of two identical punctuation marks (e.g., !, @, *, ?). Those shapes are very familiar, are visually similar to numbers, but have no conceptual size. We repeated the procedure of Experiment 3, but instead of nine Gibson figures, we used nine punctuation marks.

4.1. Method

4.1.1. Participants

Fourteen participants (4 males, 10 females) compared the physical size of punctuation marks. We used the same selection criteria as in Experiment 3, using students who had not participated in the previous experiments.

4.1.2. Stimuli

The stimuli were nine punctuation marks: !, @, #, $, %, ^, &, * and ?. The stimuli were presented as described in Experiment 3.

4.1.3. Procedure

We used the same procedure as in Experiment 3, with two changes. First, the fixation point was a cross (+) instead of an asterisk (*) (Fig. 4A), since an asterisk was one of the stimuli. Second, at the end of the task, participants were asked if the stimuli reminded them of numbers. None of the participants connected a specific punctuation mark to a number.

4.2. Results

For every participant we computed mean RT for the correct responses, and accuracy. RTs and accuracy were subjected to a one-way

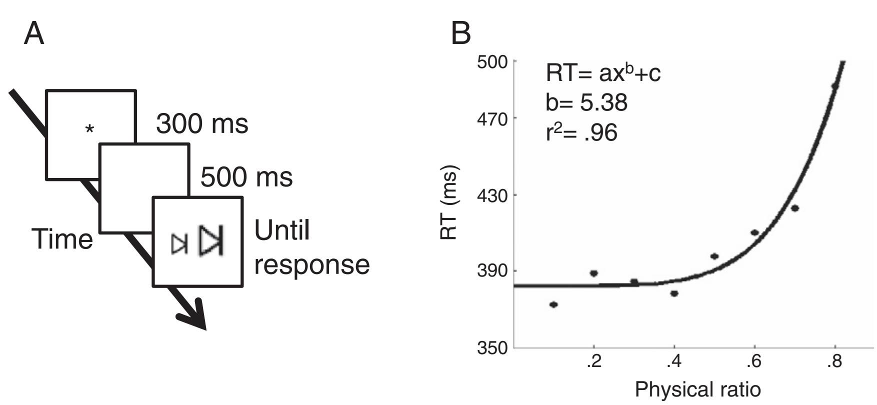

Fig. 3. Comparative judgments of physical size of Gibson figures. A) Procedure. Participants indicated the physically larger image by pressing a key. B) RT as a function of the perceptual (physical) size of the figures. This plot was fitted to a power function.

Table 2
Pairs of stimuli by physical ratio.

Category ratio	Ratio	Large number	Small number
0.1	0.11	4.5	0.5
	0.13	4	0.5
	0.14	3.5	0.5
0.2	0.2	2.5	0.5
	0.22	4.5	1
	0.25	2	0.5
	0.25	4	1
0.3	0.33	1.5	0.5
	0.33	3	1
	0.33	4.5	1.5
0.4	0.4	2.5	1
	0.43	3.5	1.5
	0.44	4.5	2
0.5	0.5	1	0.5
	0.5	2	1
	0.5	3	1.5
	0.5	4	2
0.6	0.6	2.5	1.5
	0.63	4	2.5
	0.67	1.5	1
	0.67	3	2
	0.67	4.5	3
0.7	0.71	3.5	2.5
	0.75	2	1.5
	0.75	4	3
0.8	0.8	2.5	2
	0.83	3	2.5
	0.86	3.5	3

Note. Ratio = (small size/large size) with an accuracy of 2 decimal places.

ANOVA with ratio as the dependent variable. RT analyses revealed significant main effects of size ratio, F (7, 91) = 68.47, MSE = 223, $p < .001$, $\eta^2p = .84$, suggesting that RT increased with an increase in size ratio. A similar analysis of error rates revealed that accuracy decreased with an increase in size ratio, F (7, 91) = 11.69, MSE = .0003, $p < .001$, $\eta^2p = .47$. Similar to Experiment 1, we computed b-values for each participant in each condition. As can be seen in Fig. 4B, plotting RT as a function of ratio yields a power function trend; b = 3.26, $r^2 = .97$.

5. Comparisons across experiments

5.1. Reaction time and accuracy rates analyses

In order to investigate whether the pattern of results differs among the four experimental conditions, we performed a two-way ANOVA with condition (numbers — conceptual comparison, numbers — physical comparison, punctuation marks, and Gibson figures) as an independent between-subjects variable, ratio (0.1–0.8) as an independent within-subjects variable and RT as a dependent variable. This analysis revealed a significant main effect of ratio, F (7, 364) = 136.17, MSE = 364, $p < .001$, $\eta^2p = .72$; namely, RT increased with an increase in size ratio. There was also a marginally significant main effect for condition, F (3, 52) = 2.57, MSE = 52, p = .07, $\eta^2p = .12$. The interaction between ratio and task was significant, F (21, 364) = 2.28, MSE = 996, $p < .001$, $\eta^2p = .12$. Further analyses of this interaction suggested that RTs for conceptual comparisons of numbers were significantly slower than for physical comparisons; F (3, 50) = 281034, *Wilks value* = .00059, $p < .001$. A similar analysis without the conceptual comparison condition revealed a main effect of ratio, F (7, 273) = 103.5, MSE = 428, $p < .001$, $\eta^2p = .73$, but no main effect for task, F (2, 39) < 1, *ns*, $\eta^2p = .04$). Thus, only the RT of the conceptual comparison task was significantly different, and the RTs for all the other comparison tasks were similar.

A similar analysis (i.e., all 4 conditions with ratio as an independent variable) was conducted with accuracy rates as a dependent variable. This analysis revealed only a main effect for ratio, F (7, 273) = 103.4, MSE = 428, $p < .001$, $\eta^2p = .73$. Therefore accuracy rates were not modulated by the different conditions. Average values of RT and accuracy can be seen in Table 3.

5.2. Trend analysis

Exponent (i.e., b) values for the various conditions were subjected to a one-way ANOVA with condition as an independent between-subjects variable and exponent value as a dependent factor. Exponents were modulated by experimental conditions, F (3, 52) = 10.66, MSE = 2.39, $p < .001$, $\eta^2p = .38$. Exponent values for conceptual comparisons of numbers were significantly lower than for physical comparisons of numbers, F (1, 39) = 4.25, MSE = 2.9, $p < .005$. Exponent values for physical comparisons of numbers were significantly lower than for physical comparisons of Gibson figures, F (1, 39) = 9.31, MSE = 2.9, $p < .005$. However, b-values for physical comparisons of numbers did not differ from those of physical comparisons of punctuation marks, $F < 1$, *ns*. The results are summarized in Fig. 5.

6. General discussion

Examination of changes in RT as a function of magnitude ratio revealed that the type of comparison modulated the exponent. Specifically, physical comparisons of non-symbolic figures (e.g., Gibson figures) produced the largest exponents, physical comparisons of numbers and punctuation marks produced smaller exponents and both were greater than exponents produced by comparisons of numerical values.

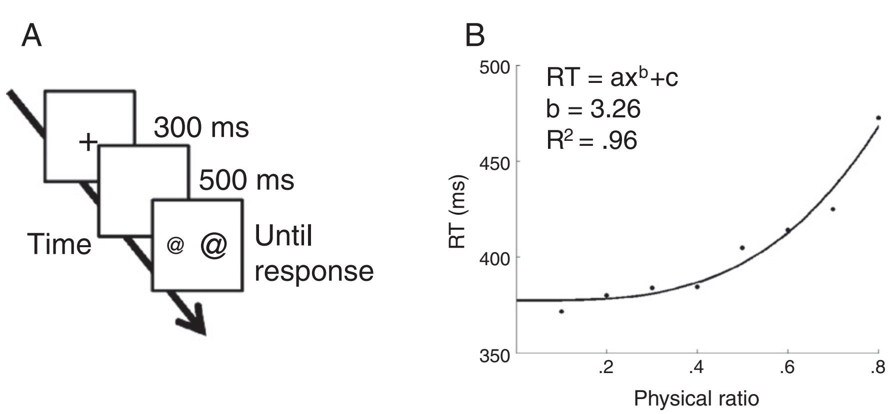

Fig. 4. Comparative judgments of the physical size of punctuation marks. A) Procedure. Participants indicated the physically larger image by pressing a key. B) RT as a function of the perceptual (physical) size of the figures. This plot was fitted to a power function.

Table 3
Average reaction times and accuracy rates in experiment 4.

Task	Accuracy (%)	Reaction times (ms)
Conceptual comparison of numbers	98	450
Physical comparison of numbers	96	371
Physical comparison of punctuation marks	99	404
Physical comparison of Gibson figures	98	405

Note. ms = milliseconds.

In comparative judgment tasks like those introduced here, participants are asked to indicate the larger of two stimuli as fast as possible. RT serves as an acceptable indirect measure of discriminability; the harder it is to discriminate, the longer the RT (Cantlon & Brannon, 2006; Cantlon et al., 2009; Feigenson, Dehaene, & Spelke, 2004). The ANS theory suggests that all magnitudes are compared using a common algorithm and comply with Weber's law. Thus, according to the ANS, we would expect a linear trend to be the best fit for all stimuli and tasks. Fitting the plot to a power function allows us to detect possible deviations from linearity and provides us with the opportunity to reveal different levels of discriminability. According to the numerical cognition literature, the change in our ability to discriminate different magnitudes increases linearly (or monotonically) with the ratio of the compared property (e.g., Brannon, 2006; Cantlon et al., 2009; Feigenson et al., 2004; Moyer, 1973; Piazza, 2010), obeying Weber's law. This means that for a fixed size ratio increment of X, RT will increase in a *constant* amount. Thus, the difference between responses to ratios 0.2 and 0.3 is *identical* to the difference between responses to ratios 0.7 and 0.8; namely, discriminability increases monotonically. We found that the exponent (i.e., b) of the power function ($Y = ax^b + c$) was greater than 1, suggesting that RT does not change by a constant amount, violating Weber's law. Rather, *RT increase* for a fixed increment in size ratio *grows* with the similarity between the stimuli (although not monotonically; e.g., RT increment from ratio 0.2 to 0.3 is much smaller than the increment from ratio 0.7 to 0.8, see Fig. 3B). This implies that discriminability becomes more difficult with increase in similarity. Do these different trends mean that magnitudes are not represented by the same algorithm? Not necessarily. The differences in performances can result from interactions with systems outside the ANS. Different tasks require different resources of attention, memory and language, or even low-level processes. These different requirements can interact with the ANS and influence different tasks in different ways. This, however, does not contradict shared representation. For example, Anobile, Cicchini, and Burr (2011) found that under attentionally-demanding conditions, an otherwise linear mapping becomes compressed and nonlinear. In a number-line bisection task, Ashkenazi and Henik (2010) found that participants with developmental dyscalculia (DD) presented a compressed representation of the mental number line (MNL). The authors attributed this representation to a deficit either in spatial attention or in the orienting network. Formal schooling was also found to modulate the representation of the MNL. Siegler and Opfer (2003) demonstrated that the representation of the MNL is compressed in childhood and becomes linear with age. Cross-cultural studies attributed this change in representation to formal education (Pica, Lemer, Izard, & Dehaene, 2004).

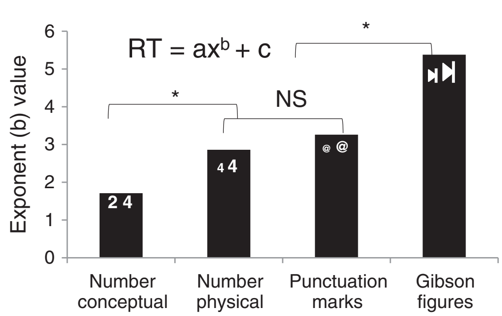

Fig. 5. Type of comparison and type of stimuli modulate psychophysical functions. The exponent (b) values for every condition. Accept for the "number conceptual" task, all the comparisons were physical. NS = non-significant. * = $p < .005$.

There is a debate in the literature whether the ratio dependency performance usually found in a comparative judgment task is due to the nature of the task, or it reflects the organization of the MNL. Verguts et al. (2005) suggested that representation of small numbers is linear (as demonstrated by the performance pattern in a priming task, for example). However, in a comparative judgment task performance will still comply with Weber's law due to noise. The current study employed only comparison tasks. Our results might represent the organization of the MNL or its noise. Nonetheless, the fact that different stimuli produced different "noise" levels (expressed as different exponents) is important since it reveals different factors that are involved with comparative judgments of magnitudes.

In the current work, two factors were found to modulate the exponent. The first was the type of comparison: exponents were smaller for conceptual comparisons than for physical comparisons. The second was familiarity; in physical comparisons, the exponents were smaller for familiar stimuli (numbers and punctuation marks) than for unfamiliar stimuli (Gibson figures). We will now discuss each factor separately.

6.1. Conceptual vs. physical comparisons

RTs in the conceptual task were longer than in the physical tasks. However, the accuracy rates were equally high in all tasks (see Table 3). Thus, the longer RT could be attributed to the process of retrieval of the symbolic meaning of the stimuli from long-term memory before they could be compared, and not to the difficulty of the task. The exponent in the conceptual comparison task was the smallest; namely, the increment in the difficulty of discrimination was the smallest (an exponent close to 1 means very minimal increment beyond the fixed increment), suggesting that it is easier to discriminate the conceptual size of numbers (in comparison to discrimination of physical size). These results are in line with previous findings in the literature (as mentioned earlier, fitting the results to linear trends yields similar fits to previous works). However, our interpretation is more cautious; we suggest a "Weber-like" trend due to the relatively low fit to a linear trend, and the significant improvement when fitting the plots to a power function.

This "Weber-like" trend, in contrast to the trend of physical comparisons, has several possible explanations. First, some claim that processing of numerical magnitudes is different from any other comparison. For example, it has been shown in several studies that the horizontal intraparietal sulcus (HIPS) is more active during comparisons of numbers than during any other comparisons (for a review see Dehaene, Piazza, Pinel, & Cohen, 2003). Thus, the fact that numerical comparisons might be processed by a designated system (see also Mussolin, Mejias, & Noël, 2010) may be the reason for the Weber-like trend.

Second, in Western culture, we are highly trained in comparisons of numerical values as we need to compare the exact difference between numbers — whether in math classes, e.g., a math problem in the classroom, or in the real world, e.g., the amount of money to pay in the store. On the other hand, we are also constantly evaluating physical sizes (e.g., the distance between two cars, etc.). However, in contrast to the conceptual comparison of numbers, physical comparisons are much more approximate. Numbers, unlike physical sizes, have verbal labels. Those verbal labels might be responsible for the linear trend of conceptual comparisons. A suggestion regarding the importance of verbal labels comes from a cross-cultural work by Pica et al. (2004). This study compared performance in a numerosity mapping task of Western adults with adults from a culture without formal education in mathematics and little or no number words in their lexicon (the Munduruku tribe). Participants were asked to place quantities of dots on a line. The authors found that while Western adults mapped numerosities linearly, adults from the Munduruku tribe mapped numerosities logarithmically, namely, small numbers (that have verbal labels) were widely spaced whereas larger numbers (that do not have verbal labels) were almost at the same location on the line, suggesting problems discriminating them.

Third, it is possible that the 'mental number line' that is postulated to follow Weber's law, is not rigid but flexible and changes with task demands. Izard and Dehaene (2008) suggested that the mental number line can be "stretched" or "compressed". When no verbal label is given, participants use a compressed number line. This can explain the elevated RTs in the large ratios in the physical comparison tasks. In contrast, when provided with verbal labels that describe specific quantities, participants transformed the spontaneous number line into a "stretched" number line, making responding more accurate. When comparing the conceptual size of two numbers, the difference has a verbal label. For example, the difference between 6 and 4 has a verbal label—2—but the difference between 2 and 2, or @ and @, does not.

In sum, we suggest to broaden the term "mental number line" to "mental magnitude line" (MML) as was suggested previously (e.g., Holmes & Lourenco, 2011). The MML exists for all magnitudes and might deviate from Weber's law in comparison tasks for some magnitudes. The differences in translation from a given dimension to the MML and differences in noise affect discriminability. In other words, the same system might be flexible enough to represent different magnitudes in different ways (i.e., linear or deviations from it). Symbolic numbers are a special case since we are trained to represent them verbally and with high accuracy. The different exponent values in the current study, combined with the results of previous works (Izard & Dehaene, 2008; Pica et al., 2004; Siegler & Opfer, 2003), suggests that the natural representation of magnitudes is more noisy (deviates from Weber's law), and the representation of numbers depends on cultural and educational factors such as familiarity with the concept of numbers, the frequency of the number, etc.; thus, it is linear for small numbers and compressed for large and less frequent numbers (Anobile et al., 2011; Verguts et al., 2005). Note, however, that the concept of frequency holds for symbolic numbers but not for physical sizes. Because different physical sizes are all around us and we need to make size comparisons every day starting at a very early age, it is hard to evaluate frequency of one physical size in comparison to another. This is why we discuss the term "familiarity" and not "frequency".

6.2. Familiarity

Unlike Gibson figures, which were novel stimuli for our participants, numbers are highly familiar stimuli that carry a conceptual size. Thus, the differences in the exponent between physical comparisons of Gibson figures and numbers can stem from the difference in familiarity or the lack of existence of conceptual size. To distinguish between these two options, we conducted Experiment 4 with punctuation marks — familiar shapes with no conceptual size. The exponent of physical comparisons of punctuation marks was not significantly different from that of physical comparisons of numbers. Hence, we concluded that the familiarity, and not concept of size, affected the exponent. More research is needed to understand through which mechanism familiarity affects performance in a comparative judgment task.

6.3. Implications

Although RT and accuracy rates did not differ for the various physical comparisons, the exponents were different. Accordingly, we suggest that the trend of RT as a function of ratio is a more sensitive measure of performance than the average RT. Accuracy rates are similarly being used to extract a Weber fraction value (Halberda, Mazzocco, & Feigenson, 2008).

Different exponents can also suggest that the mental magnitude line is more flexible than assumed before. Even factors that do not relate to the meaning of a number – such as the familiarity of its physical shape – can make it easier or harder to discriminate its physical size. This finding has some interesting implications, especially for developmental studies. It looks like early in life, the ability to discriminate magnitudes does not depend only on the differences between sizes, but on the familiarity with the to-be-compared objects also. Later in life, and with formal education in mathematics, the ability to discriminate is affected by verbal labels.

6.4. Limitations

Our results are currently confined to the visual modality. Additional experiments are needed in order to extend our conclusions to other modalities.

The symbolic stimuli compared here were one-digit numbers. The representation of these numbers is highly automatic. Furthermore, culturally, we are highly trained with exact representation and comparison of numbers from an early age in our everyday lives (e.g., math problems in school, paying in the store, etc). This is not the case for all size representations. While we are able to tell if a cat is larger than a mouse, we are not trained in *accurately* representing that difference. This can lead to differences in performance when comparing the conceptual size of numbers or two objects.

Our results are not automatically applicable to discrete non-symbolic stimuli such as arrays of dots. The function of these kinds of stimuli was more thoroughly investigated. However, numerosity was always confounded by continuous properties such as density, total surface area, etc. (Gebuis & Gevers, 2011). For this reason, we cannot directly compare our results to those studies. An ongoing work in our lab has revealed a linear trend for comparisons of dot numerosity (Leibovich & Henik, 2013).

7. Conclusions

To summarize, to overcome some shortcomings in previous studies supporting shared representation of magnitudes, we employed an identical task (comparative judgment) to investigate different types of magnitude comparisons. The use of 8 ratios and the fit of the results to a power function make our trend analysis a reliable and sensitive tool to detect differences in discriminability of different magnitudes. Our findings suggest that the relationship between discriminability and size ratio is not always linear as previously suggested; rather, it is modulated by the type of comparison and the type of stimuli.

Authors note

Dr. Sarit Ashkenazi is now at the School of Education. The Hebrew University of Jerusalem, Mt. Scopus. Jerusalem.

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC Grant agreement no. 295644.

References

Anobile, G., Cicchini, G. M., & Burr, D. C. (2011). Linear mapping of numbers onto space requires attention. *Cognition*, *122*, 454–459.

Ashkenazi, S., & Henik, A. (2010). A disassociation between physical and mental number bisection in developmental dyscalculia. *Neuropsychologia*, *48*, 2861–2868.

Beran, M. J., Decker, S., Schwartz, A., & Schultz, N. (2011). Monkeys (*Macaca mulatta* and *Cebus apella*) and human adults and children (Homo sapiens) compare subsets of moving stimuli based on numerosity. *Frontiers in Psychology*, *2*.

Brannon, E. M. (2006). The representation of numerical magnitude. *Current Opinion in Neurobiology*, *16*, 222–229.

Buhusi, C. V., & Cordes, S. (2011). Time and number: The privileged status of small values in the brain. *Frontiers in Integrative Neuroscience*, *5*.

Cantlon, J. F., & Brannon, E. M. (2006). Shared system for ordering small and large numbers in monkeys and humans. *Psychological Science*, *17*, 401–406.

Cantlon, J. F., Platt, M. L., & Brannon, E. M. (2009). Beyond the number domain. *Trends in Cognitive Sciences*, *13*, 83–91.

Cohen Kadosh, R., Cole, N.B., Henik, A., Rubinsten, O., Mohr, H., Dori, H., et al. (2005). Are numbers special? The comparison systems of the human brain investigated by fMRI. *Neuropsychologia*, *43*, 1238–1248.

Cohen Kadosh, R., Henik, A., & Rubinsten, O. (2008). Are Arabic and verbal numbers processed in different ways? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 1377–1391.

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology*, *20*, 487–506.

Droit-Volet, S. (2010). Speeding up a master clock common to time, number and length? *Behavioural Processes*, *85*, 126–134.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*, 307–314.

Festinger, L. (1943). Studies in decision: I. Decision-time, relative frequency of judgment and subjective confidence as related to physical stimulus difference. *Journal of Experimental Psychology*, *32*, 291–306.

Fias, W., Lammertyn, J., Reynvoet, B., Dupont, P., & Orban, G. A. (2003). Parietal representation of symbolic and nonsymbolic magnitude. *Journal of Cognitive Neuroscience*, *15*, 47–56.

Gebuis, T., & Gevers, W. (2011). Numerosities and space; indeed a cognitive illusion! A reply to de Hevia and Spelke (2009). *Cognition*, *121*, 248–252.

Gibson, E. P., Gibson, J. J., Pick, A.D., & Osser, H. (1962). A developmental study of the discrimination of letter-like forms. *Journal of Comparative and Physiological Psychology*, *55*, 897–906.

Halberda, J., Mazzocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*, 665–668.

Holmes, K. J., & Lourenco, S. F. (2011). Common spatial organization of number and emotional expression: A mental magnitude line. *Brain and Cognition*, *77*, 315–323.

Izard, V., & Dehaene, S. (2008). Calibrating the mental number line. *Cognition*, *106*, 1221–1247.

Johnson, D.M. E. (1939). *Confidence and speed in the two-category judgment.* Columbia University.

Leibovich, T., Henik, A., & Bertin, E. (2013). *Magnitude processing in non-symbolic stimuli, Frontiers in Psychology*, *4* (375). http://dx.doi.org/10.3389/fpsyg.2013.00375.

Möhring, W., Libertus, M. E., & Bertin, E. (2012). Speed discrimination in 6-and 10-month-old infants follows Weber's law. *Journal of Experimental Child Psychology*, *111*(3), 405–418.

Moyer, R. S. (1973). Comparing objects in memory: Evidence suggesting an internal psychophysics. *Perception & Psychophysics*, *13*, 180–184.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, *215*, 1519–1520.

Mussolin, C., Mejias, S., & Noël, M. P. (2010). Symbolic and nonsymbolic number comparison in children with and without dyscalculia. *Cognition*, *115*, 10–25.

Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review*, *76*, 241–263.

Paivio, A. (1975). Perceptual comparisons through the mind's eye. *Memory & Cognition*, *3*, 635–647.

Piazza, M. (2010). Neurocognitive start-up tools for symbolic number representations. *Trends in Cognitive Sciences*, *14*, 542–551.

Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., et al. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition*, *16*(1), 33–41.

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron*, *44*, 547–555.

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science*, *306*, 499–503.

Rubinsten, O., & Henik, A. (2002). Is an ant larger than a lion? *Acta Psychologica*, *111*, 141–154.

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation. *Psychological Science*, *14*, 237–243.

Tokita, M., & Ishiguchi, A. (2011). Temporal information affects the performance of numerosity discrimination: Behavioral evidence for a shared system for numerosity and temporal processing. *Psychonomic Bulletin and Review*, *18*, 550–556.

Verguts, T., Fias, W., & Stevens, M. (2005). A model of exact small-number representation. *Psychonomic Bulletin & Review*, *12*, 66–80.

Vigliocco, G., Vinson, D. P., Damian, M. F., & Levelt, W. (2002). Semantic distance effects on object and action naming. *Cognition*, *85*, 61–69.

Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, *7*, 483–488.