Research Article

Number Line Compression and the Illusory Perception of Random Numbers

Arnaud Viarouge,[1,2,3,4] Edward M. Hubbard,[1,2,3] Stanislas Dehaene,[1,2,3,5] and Jérôme Sackur[1,6,7]

[1]INSERM-CEA Cognitive Neuroimaging Unit, Gif/Yvette, France, [2]CEA, DSV/I2BM, NeuroSpin Center, Gif/Yvette, France, [3]University of Paris-Sud, Gif/Yvette, France, [4]CNRS, Centre de Recherche en Epistémologie Appliquée, Paris, France, [5]Collège de France, Paris, France, [6]Laboratoire de Sciences Cognitives et Psycholinguistique, (CNRS/ENS/EHESS), Paris, France, [7]Département d'Etudes Cognitives, Ecole Normale Supérieure, Paris, France

Abstract. Developmental studies indicate that children initially possess a compressed intuition of numerical distances, in which larger numbers are less discriminable than small ones. Education then "linearizes" this responding until by about age eight, children become able to map symbolic numerals onto a linear spatial scale. However, this illusion of compression of symbolic numerals may still exist in a dormant form in human adults and may be observed in appropriate experimental contexts. To investigate this issue, we asked adult participants to rate whether a random sequence of numbers contained too many small numbers or too many large ones. Participants exhibited a large bias, judging as random a geometric series that actually oversampled small numbers, consistent with a compression of large numbers. This illusion resisted training on a number-space mapping task, even though performance was linear on this task. While the illusion was moderately reduced by explicit exposure to linear sequences, responding was still significantly compressed. Thus, the illusion of compression is robust in this task, but linear and compressed responding can be exhibited in the same participants depending on the experimental context.

Keywords: numerical cognition, mental number line, psychophysical staircase

One of the fundamental issues in numerical cognition is the representational scale of the numerical continuum (Dehaene, 2001; Gallistel & Gelman, 1992). The *linear model* argues that numbers are represented on a linear scale with increasing variability as numerical quantity increases (Gallistel & Gelman, 1992), while the *logarithmic model* suggests that numbers are represented on a compressed scale (notably logarithmic) with constant variability. Studies of numerical capacities in infants (Xu & Spelke, 2000) and adults who lack exact number words (Pica, Lemer, Izard, & Dehaene, 2004) and single-unit neuronal responses (Nieder & Miller, 2003) suggest that numerosity is more likely to be coded on a compressive scale. In fact, both models are equally able to account for behavioral data in human and nonhuman participants (Dehaene, 2001; Gallistel & Gelman, 1992; Jordan & Brannon, 2006), in which judgments of similarity between numbers show a compressed mental number line.

While a compressed mental number line is usually inferred from reaction times obtained in tasks where participants are required to perform speeded judgments on pairs of numbers (Dehaene, Dupoux, & Mehler, 1990; Moyer & Landauer, 1967), some experimental contexts have directly revealed compressed responding when participants' intuitions concerning the distances between numbers are investigated in a more holistic way. For example, Siegler and Opfer (2003) found that when they asked children to map numbers onto positions on a line segment (number to position or N-P task) the children's responses were better fit by a logarithmically compressed curve that exaggerated the distances between smaller numbers, and underrepresented the distances between large numbers. However, this illusion of compression seems to give way to linear responding in participants older than 8 years, probably because the older children were more accustomed to manipulating numbers in the range used for these tasks. Opfer and Siegler (2007) subsequently showed that feedback could also quickly lead to a recalibration of the mental number line, so that after training children's performance was better fit by a linear function than by a compressed function. This recalibration occurred even after providing children with feedback on only a single trial (i.e., by indicating the correct location of one number on the number line) and was greatest for numbers for which there was the greatest difference between the linear and logarithmic predictions. Similarly, a recent study with Mundurucu adults who lack exact number words showed that they performed the N-P task in a way that was consistent with a compressed understanding of numerical distance (Dehaene, Izard, Spelke, & Pica, 2008). However, in

Western adults, this kind of compressed responding is not found using the N-P task (Dehaene et al., 2008; Siegler & Opfer, 2003), perhaps, as Siegler and Opfer assume, because of the spontaneous use of mathematical strategies to evenly divide the numerical interval used in the experiments.

The current experiments explored two questions. First, can the continued presence of compressed responding be demonstrated in adults in the same kind of task investigating the participants' intuitions about numerical distance, despite their linear behavior on the N-P task? Second, can this responding be recalibrated in the same way that visual feedback recalibrates children's performance on the N-P task? To explore these questions, we adapted a method first introduced by Banks and Coleman (1981). Banks and Coleman (1981) asked participants to rate "how randomly and evenly" various prerecorded auditory series of numbers sampled a numerical interval. The series that were rated as most random were those generated from a power function[1] with an exponent greater than 1.0. In their most quantitative experiment, the authors generated seven series of 15 numbers in the interval [1, 2,000] using seven different fixed exponents (0.91, 1.11, 1.43, 2, 2.77, 3.33, and 10; we here report the inverse of the values reported by Banks and Coleman to make them comparable with the values reported in our study). On average, participants rated the series generated with an exponent of approximately 2.8 (oversampling small numbers) as most random. However, due to limited computational power, participants were tested only with a few prespecified exponents and gave ratings on a 1–10 scale for only one example of each exponent. This may limit the precision and reliability of the estimates obtained for each participant as such methods may sample the appropriate range too coarsely and will therefore fail to detect subtle changes in threshold.

Taking advantage of the much greater computational power available with modern computers, we were able to improve on these methods to develop a psychophysical staircase procedure that would allow us to test a whole range of exponents more finely, and obtain a more precise estimate of the subjective rate of compression of numerical distances. Participants listened to a series of randomly ordered numbers, sampled from distributions which were either flat across the range of numbers tested, oversampled small numbers (as if sampling randomly from a compressed continuum), or undersampled small numbers. They judged whether each sequence contained too many small or large numbers (distribution judgment task), and a staircase procedure was used to modify the amount of compression on the next trial. The procedure converged to the point of subjective equality, enabling us to measure each participant's subjective compression. Across five experiments, we found that adults possess a compressed intuition of numerical distance which was remarkably stable, even after participants received feedback about what constitutes a true random sequence.

General Methods

We tested a total of 74 participants (16 per experiment in Experiments 1, 2, 4, and 5; 10 in Experiment 3; 19 males, 55 females). Participants were 18–34 years old (mean 23.1, the ages from four participants were lost) and were naïve to the study hypotheses. All participants gave informed consent and received 7 euros (approximately $10) for their participation. They were all right-handed (Oldfield, 1971) native French speakers. Our participants were a representative sample of university students, who had a basic mathematical background in line with the French high-school curriculum, including some understanding of basic probability theory, algebra, and geometry, up to and including, in most cases, introductory calculus. Preliminary analyses suggested that the compression did not differ between students pursuing scientific or literary majors, so we stopped collecting detailed information on the participant's course of study.

On each trial, participants were presented with a series of randomly ordered spoken number words separated by 500 ms blanks. The sound files were compressed (using Praat; Boersma & Weenink, 2005) so that they were matched on overall length as much as possible while still remaining understandable to naïve participants (see below for details). The numbers x_n in each series were generated using the equation $x_n = \alpha n^E + \beta$, $n = 1$–10, with α and β chosen so that $x_1 = a$ and $x_{10} = b$.[2] This equation generated 10 numbers in the interval [a, b]. Depending on the experiment, the last number b was or was not presented. In order to minimize strategic influences, participants were not told what the interval used in the task was, nor the number of items in the sequences. A random jitter (−2 to 2) was added to each number to prevent the exact same sequence from being heard repeatedly. Each sequence had a mean duration of 13 s (SD = 658 ms), and total task duration was approximately 13 min. By varying the exponent E, we presented sequences with varying degrees of compression, such that some sequences oversampled small numbers ($E > 1$), and others oversampled large numbers ($E < 1$). Participants were asked to judge whether the sequence seemed to contain too many small or large numbers by using the left- or right-arrow keys to choose between two fixed distributions, represented by a graph with a sequence of nine tick marks generated by an exponential function with two different exponents ($E < 1$ and $E > 1$, see Appendix). The meaning of these graphs was explained to the participants, who were then asked to generate an example of each kind of sequence. To be sure that the instructions were correctly understood, they also judged an example of each kind of sequence (with exponents of 4.0 and 0.3) before beginning the task. The exponent for the next sequence was adapted according to the participant's response. Two randomly interleaved staircases were used, one with an initial exponent of 4.0

[1] Of the form $x_n = \alpha n^E$, with x_n being the nth number of the series, α chosen so that the first and last numbers of the series approximately corresponded to the lower and upper bounds of the numerical interval tested, respectively, and E varying according to the series tested.

[2] We followed Banks and Coleman (1981) in adopting this function because it allows for an easy manipulation of the rate of compression and includes linearity as a special case ($E = 1$). Although the internal continuum is believed to be logarithmic, the log function cannot be easily parametrized, and a power function, as used here, provides a close approximation when used on a relatively small numerical interval.

(oversampling small numbers) and the other with an initial exponent of 0.3 (oversampling large numbers). We chose these initial exponents to balance the compression of the initial sequences. An initial exponent of 4.0 (close to the inverse of 0.3) yields a compression of the first sequence which is comparable, in the opposite direction, to the one obtained for the sequence generated from the initial exponent 0.3. For each trial, one of the two staircases was randomly chosen. The exponent was increased or decreased for the next trial in the corresponding staircase depending on whether the participant indicated that the sequence oversampled large or small numbers, respectively. The initial step size was 0.4, and was divided by two at each reversal, until it reached a minimum of 0.1. The program stopped when each staircase contained eight reversals.

To estimate the compression of each participant's mental number line, we identified the last five reversals for each staircase and averaged the exponents obtained for each trial in the relevant staircase. This represents the compression at which participants could no longer determine whether the sequences contained more small or large numbers (the point of subjective equality), or the sequence judged as "most random" by the participants.

Experiments 1 and 2

Experiment 1 tested whether the compression of numerical distance is influenced by the range of numbers tested, while Experiment 2 tested whether the presence of the upper boundary of the interval affected this compression. Banks and Coleman (1981) have previously argued that the presence of the upper boundary leads to the use of a more linear representation of numbers, or at the very least, more linear responding. Crucially, however, Banks and Coleman did not test the exact same range in both the bounded and unbounded conditions. Thus the effect of interval and the presence of the boundaries cannot be disentangled in their experiments. In order to more systematically explore the effects of interval size and boundedness, we independently varied interval size and the presence or absence of the upper bound, yielding two different intervals of the same size: Interval 1 = [3, 750] and Interval 2 = [153, 900], both in an unbounded (Experiment 1) and bounded (Experiment 2) condition. We chose these nonstandard intervals (i.e., avoiding intervals like [1, 1,000]) in order to minimize the possibility that participants would base their responses on simple arithmetic procedures like mentally dividing the range into 10 equal parts, and chose a range that was large enough to decrease the possibility that the tested sequences would be too similar when the exponents were close. Despite the compression of the sounds files (see General Methods, above), the one- and two-digit number words were significantly shorter (mean = 346 ms, *SD* = 130 ms) than the three-digit number words (mean = 566 ms, 139 ms; $t = 15.72$, $p = 2.2e-16$). However, since the interval [153, 900] contains only three-digit numbers, the auditory stimuli are equivalent in length in this condition. Indeed in this interval, we were able to successfully match the length of the words for smaller numbers (i.e., 150–525, mean = 574 ms, *SD* = 145 ms) so that they did not differ significantly from the length for larger numbers (526–900, mean = 585 ms, *SD* = 114 ms; $t = 1.12$, $p = .26$).

Method

Each participant completed one session for each of the two intervals with order counterbalanced between participants. In Experiment 1, the upper boundary was removed from the sequence so that participants heard nine numbers. In Experiment 2, the upper boundary was presented as part of the sequence, so participants heard 10 numbers.

Results

The mean exponent for each interval was significantly greater than 1.0 in both experiments, clearly indicating a compression of the mental number line (Figure 1). Experiment 1: $E = 1.98$, $t(15) = 7.63$, $p < .001$ and $E = 1.44$, $t(15) = 5.01$, $p < .001$, for Intervals 1 and 2, respectively. Experiment 2: $E = 2.23$, $t(15) = 11.85$, $p < .001$ and

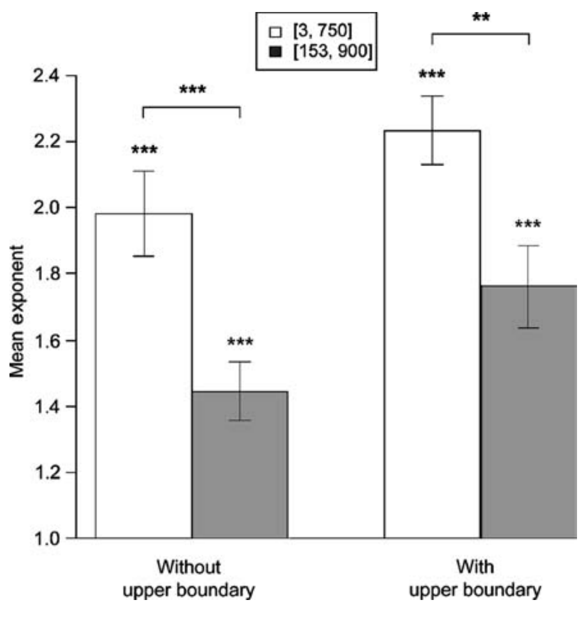


Figure 1. Results of Experiments 1 and 2. The figure shows the mean exponents calculated for both interval [3, 750] and interval [153, 900], in the case of auditory sequences without and with the upper boundary. We see that the mean exponent is greater than 1.0 in all conditions. The mean exponent is lower with interval [153, 900] in both unbounded and bounded conditions, whereas the bounded condition leads to a greater overall mean exponent than the unbounded condition.

Table 1. Functions that best fit participants' mean judgments in Experiments 1 and 2 for the two intervals used

	Experiment 1	Experiment 2
[3, 750]	$x = 7.89 \times n^{1.98} - 4.89$	$x = 4.37 \times n^{2.23} - 1.37$
[153, 900]	$x = 27.86 \times n^{1.44} + 125.13$	$x = 13.16 \times n^{1.76} + 139.84$

$E = 1.76$, $t(15) = 6.14$, $p < .001$. The corresponding best fitting equations are shown in Table 1.

Interestingly, while the exponents for Interval 2 were significantly greater than 1.0, they were significantly smaller than for Interval 1, $F(1, 15) = 22.48$, $p < .001$ (Experiment 1) and $F(1, 15) = 9.02$, $p < .01$ (Experiment 2). This outcome is consistent with participants relying on a logarithmically compressed mental number line. Indeed, one property of the logarithm function is that its curvature decreases as number increases. Thus, when fitted locally by a geometric series, the recovered exponent should be smaller over the second interval [153, 900] than over the first one [3, 750], consistent with our observations. Figure 2 illustrates the shape of the geometric series generated with the participant's mean exponent (*x* axis), plotted against a linear spacing of numbers in the same interval (*y* axis). On this graph, the smaller curvature associated with

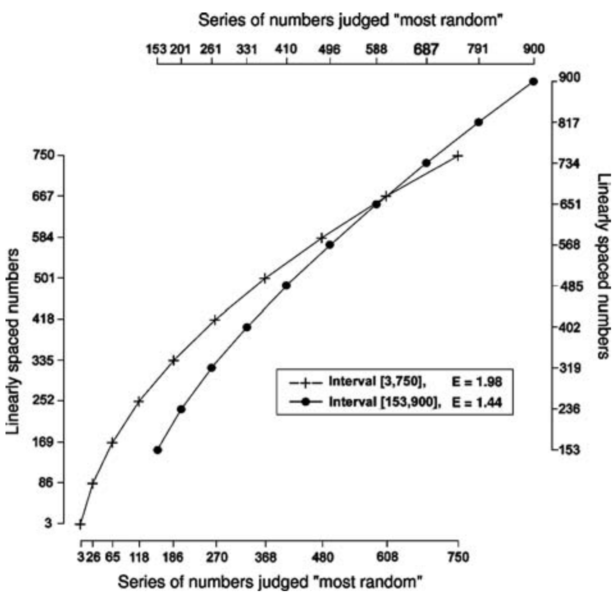


Figure 2. Graphical representation of the numerical sequences judged as "most random" according to Experiment 1 (exponents taken from the unbounded condition). The *x* axis shows the numbers that were presented to participants, while the *y* axis shows a linear scale over the same interval of numbers (either [3, 750] or [153, 900]). In this representation, the compression for large numbers is evident. The smaller exponent for the larger interval [153, 900] than for the smaller interval [3, 750] can readily be interpreted as a smaller curvature of the internal compression function, compatible with this function being a logarithm.

Interval 2's smaller exponent fits with the hypothesis of an overall fixed compression over the whole range of numbers. Indeed, an overall compressed scale predicts that the illusion of compression will be greater when the range tested covers smaller numerosities. The fact that we observed different exponents for different numerical intervals also suggests that participants do not recalibrate their mental number line according to the new range of numbers tested. Instead, the smaller exponent obtained for larger numbers is consistent with a single compressed scale in which 153 is subjectively closer to 900 than 3 is to 750 (see also Shepard, Kilpatrick, & Cunningham, 1975). Additionally, the fact that the compression in the interval [153, 900], where the length of the number words was equalized, was still significantly greater than 1.0 argues against the possibility that the compression observed in the interval [3, 750] was merely an artifact due to a cognitive overrepresentation of the bigger numbers compared to the smaller ones as a result of differences in the sound file length.

More surprisingly, we find a stronger compression of the mental number line (i.e., a greater mean exponent) when participants listen to bounded intervals (where the upper-bound number x_{10} is included in the series) than when they listen to unbounded intervals ($F(1, 30) = 5.11$, $p < .05$). This finding demonstrates that, contrary to the suggestion by Banks and Coleman (1981), the presentation of the upper boundary is not sufficient to elicit the use of a linear scale. Other factors may affect the illusion of compression. For instance, in Experiment 1 with unbounded sequences participants heard 9 numbers, while in Experiment 2 with bounded sequences participants heard 10 numbers. Longer sequences may be more difficult to keep in working memory, leading to a greater reliance on an intuition of number size based on the compressed number line, thus resulting in a larger deviation from linearity. Conversely, Banks and Coleman's stimuli (10 numbers in the range from 1 to 1,000) may have made it particularly easy for participants to mentally divide the range, thus allowing them to rely on calculation strategies which masked the illusion of compression.

However, one possible concern about our results is that participants may be basing their responses on some other simpler strategy such as the magnitude of the last number or the direction of the change between the last two numbers, rather than judging the entire sequence. Because of the compression of the series, it is not possible to test whether the last number predicts the participants' responses simply by testing whether the mean varies between "more larger" and "more smaller" responses. Indeed, the more compressed the series is, the more likely it is that the last number of the corresponding series is large. As predicted, we found that the mean of the last numbers was larger when participants responded "more large numbers" than when they responded "more small numbers" (280.95 vs. 242.67, $t(15) = 2.73$, $p = .015$). The same statistical analysis on the mean of the last three numbers of the series yielded a larger difference between the two response conditions (273.84 vs. 225.97, $t(15) = 6.34$, $p = 1.34\text{e}{-05}$), and an even larger difference was found when all of the numbers of the series were included (270.30 vs. 217.99, $t(15) = 22.94$, $p = 4.28\text{e}{-13}$). These analyses confirm that participants correctly performed the task, since the series for which the participants should respond "more large numbers" did indeed contain more large numbers.

In order to test whether participants based their decision on the last number in the series, rather than overall compression, it was thus necessary to perform an analysis using a parameter that was independent of the exponent of the compression. We tested the correlation between the rank of the last number in the series (independent of compression) and the percentage of "more large numbers" responses in each case. By computing a Spearman rho correlation coefficient we found no significant correlation between the rank of the last number presented and the response ($\rho = .4$, $p = .29$). Finally, we also tested whether the direction of the difference between the last two numbers (increasing or decreasing) could predict participant's responses. We found no significant difference between the percentage of "more large numbers" responses for increasing (53.7%, $SD = 13\%$) and decreasing (46.3%) final pairs ($t(15) = .85$, $p = .40$), indicating that this feature of the stimuli also could not explain the results we see here. These additional analyses, combined with the fact that we find similar compression for the interval [153, 900], suggest that participants are not relying on a simple memory strategy, and our findings of an illusion of compression cannot be explained by a simple cognitive overrepresentation of the stimuli due to difference in length, or other memory limited strategies.

Experiment 3

In Experiments 1 and 2, we chose initial exponents that balanced the compression of the initial sequences (0.3 and 4.0, where the exponents 0.3 and 3.33 (1/0.3) lead to symmetric compression). However, given that we also used an additive staircase to update these values more trials were required to reach an exponent of 1.0 from the upper starting point than from the lower starting point. Thus, participants initially made more "more small numbers" responses, which could conceivably have biased the results. Although randomly interleaved staircases are robust to biases in the initial starting positions of the two staircases (Cornsweet, 1962; Nachmias & Steinman, 1965), to probe whether this asymmetry in the number of steps required to reach an exponent of 1.0 influenced our results, we conducted a control experiment in which the initial exponents required the same number of linear steps to reach 1.0 (0.3 and 1.7). Note that with these equidistant starting values on a linear scale, the compression of the two sequences was asymmetric, with the sequence generated by the upper starting point (1.7) being more similar to a linear sequence than that generated by the lower starting point (0.3).

Method

Ten participants took part in this experiment. They performed the same task as in Experiments 1 and 2. The sequences sampled the range [3, 750], with the upper boundary removed.

Results

Despite the fact that these new starting points yield compressions that are initially biased toward a linear sequence, this experiment still yielded a mean final exponent greater than 1.0 (1.54, $t(9) = 5.89$, $p < .001$), revealing that the illusion of compression of the mental number line is robust and not due to any artifactual influence of the starting points.

Experiment 4

In Experiment 4, we explored whether we could reduce, or even eliminate, the illusion of number compression uncovered in Experiments 1–3 by providing participants with visual feedback, consisting in a conversion task, which has been shown to elicit a shift from a compressed to a linear number-space mapping in children (Opfer & Siegler, 2007). We wanted to directly compare the results of Siegler and Opfer's task with our distribution judgment task. Since adult participants generally show linear behavior on this N-P task, we hypothesized that, if our distribution judgment task and the N-P task both tap into the same representation of number, then asking participants to perform the N-P task may lead to a corresponding shift in participants' behavior on our distribution judgment task. In addition, we hypothesized that several presentations of the N-P task would more powerfully elicit a shift from logarithmic to linear responding.

Method

Sixteen participants took part in this experiment. We used an A-B-A′ design in which A and A′ represent our distribution judgment task, and B represents the feedback task, in this case, Siegler and Opfer's (2003) number to position (N-P) conversion task. In the N-P task, a black line ($L \times H = 492 \times 2$ pixels = 159.5×0.65 mm) appeared in the center of the computer screen, with the endpoints (black vertical bars, $L \times W = 20 \times 2$ pixels) labeled with the boundaries of the interval used in the randomness judgment Blocks A and A′. A red mark ($L \times W = 30 \times 2$ pixels) appeared at a randomly chosen position on the line, and a number was presented approximately 5 cm below the line (in "FreeSansBold," a sans serif font similar to Arial, used by default in Pygame, with a height of 32 pixels). Participants moved the red mark using the left-and right-arrow keys (each key press shifted the marker 6 pixels or about 1.2% of the total length of the line) to indicate the position which subjectively corresponded to the number indicated. After they pressed the enter key on the keyboard, the participant's response remained visible on the screen and the actual (linear) position of the number, labeled with the original number, appeared on the line. This feedback helped participants to calibrate their N-P task performance.

We used the interval [3, 750] without the upper boundary, as in Experiment 1, both before and after the feedback block. In the N-P task, participants were asked to indicate the position of eight numbers each corresponding to one ninth of the interval. We repeated the N-P task five times for each number (40 trials total), in random order. For each number, we plotted the mean response for the *n*th presentation ($n = 1–5$).

Results

In the N-P conversion task, participants' responses were close to linear from the first presentation of each value. On the first presentation, there was a small tendency to underestimate the actual position of the two largest values, 586: $t(15) = -3.17$, $p < .01$, 670: $t(15) = -2.52$, $p < .05$ and to overestimate the position of the smallest value, 86: $t(15) = 3.15$, $p < .01$, consistent with a slight compression of the mental number line. Oddly, 503 was overestimated, $t(15) = 5.20$, $p < .001$. These deviations were corrected for subsequent presentations, and even somewhat overcorrected for the fifth presentation, 170: $t(15) = -2.33$, $p < .05$, 336: $t(15) = -3.16$, $p < .01$ (Figure 3). A regression analysis performed on the median estimates confirms that the participants' responses were close to linear from the first presentation, with the best fitting equation being $y = 0.94 \times x - 20.6$, $r^2 = .97$ for the first and $y = 1.03 \times x - 17.04$, $r^2 = .99$ for the fifth presentation.

As in Experiments 1–3, the mean exponent was significantly greater than 1.0, both before ($E = 2.21$, $t(15) = 8.60$, $p < .001$) and, most importantly, after performing the N-P task ($E = 2.16$, $t(15) = 9.06$, $p < .001$) (Figure 4). There was not even a significant effect of block ($F(1, 15) = 0.25$, $p = .625$). Thus, feedback from the N-P task failed to modify the compression of the mental number line. This null effect was not due to experimental noise, but rather reflects stable differences in individual participants' compression. To demonstrate the stability of the observed compression, we calculated the Pearson correlation between the exponent obtained for each participant in Block A and that obtained in Block A′. We found that the exponents of the second session were highly correlated with those of the first session ($r^2 = .55$, $t = 4.18$, $p < .001$), indicating that our method provides a precise estimate of each participant's preferred compression, and that these did not differ after performing the N-P task.

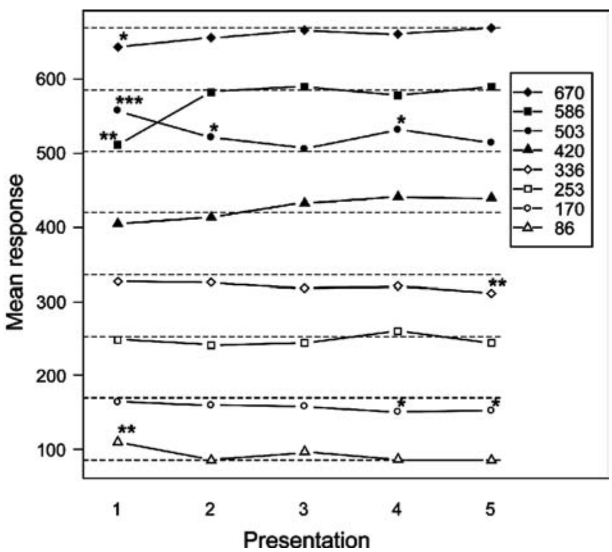


Figure 3. The results of the N-P conversion task. The figure shows the mean response indicated by the participants at the five ranks of presentation and for each of the eight values asked. We see that the responses are quite close to the actual position (dashed lines) from the first rank on.

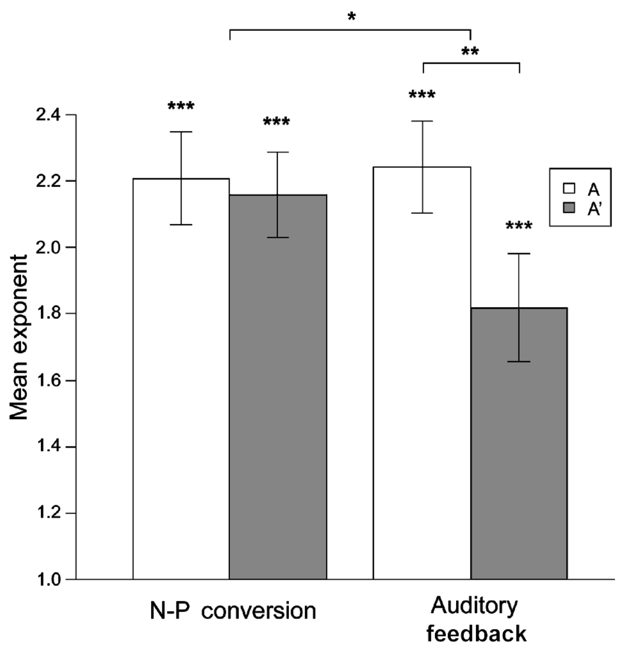


Figure 4. Results of Experiments 4 and 5. The figure shows the mean exponents calculated for Blocks A (prior to intervention) and A′ (after intervention). The intervention consisted either in an N-P conversion task (Experiment 4) or in explicit auditory feedback (Experiment 5). We see that only the auditory feedback led to a significant decrease of the mean exponent, indicating a reduced subjective compression of the mental number line, yet without canceling it (the exponent remains significantly greater than 1).

Experiment 5

In Experiment 5, we explicitly presented participants with linear sequences of auditory numbers to test whether explicit feedback on the task of interest could modify the illusion of compression.

Method

Sixteen participants took part in this experiment. The stimuli and design were the same as in Experiment 4 with the training task now being to listen to sequences of numbers generated with a linear function. Participants were presented with 10 linear auditory sequences, again with added jitter. In order to maximize the potential impact of the presentation of the linear sequences, participants were told that the sequences were linear and were asked to visualize the numbers and to attend to their equal spacing. Additionally, by training and testing on the same numerical range, we maximized the possibility of transfer from the training set to the test set.

Results

As in the four first experiments, the mean exponent was significantly greater than 1.0 both before ($E = 2.24$, $t(15) = 9.04$, $p < .001$) and after the feedback task ($E = 1.82$, $t(15) = 5.03$, $p < .001$). Thus, even with direct training on the task of interest the illusion of compression was not eliminated, demonstrating its robustness. Using the same correlation measure as in Experiment 4, we found that the exponents for Block A′ were correlated with those for Block A, suggesting that individual differences in compression were robust across the two sessions ($r^2 = .34$, $t = 2.70$, $p < .05$). Unlike with the N-P task, we found a significant effect of block, with a lower mean exponent after the auditory feedback task, which indicates that training with linear sequences modified the illusion of compression, $F(1, 15) = 9.47$, $p < .01$, even if it did not eliminate it. Experiments 4 and 5 differed as shown by a significant Block × Feedback task interaction in a joint ANOVA ($F(1, 30) = 5.01$, $p < .05$; see Figure 4 and Table 2). Thus, direct auditory feedback, unlike the N-P conversion task, leads to a change in performance on the distribution judgment task. However, this change could either be due to subtle changes in the underlying representation, or to learning of additional cues or strategies, such as anchoring. Although it is possible that even more intensive training would elicit a shift from a logarithmic to a linear responding in our distribution judgment task, we note that the presentation of a *single trial* of feedback has been shown to modify performance on the N-P task (Opfer & Siegler, 2007), while in our experiment as many as 10 repetitions of the auditory sequences failed to eliminate the compressed responding observed in our auditory distribution judgment task.

Discussion

Our results demonstrate that the sequences judged to sample the numerical continuum most randomly are, in fact, sequences generated from a power function with an exponent significantly greater than 1.0. These results add additional support to the hypothesis that adult human beings possess a compressed sense of numerical distances and that it is necessary to present sequences that oversample small numbers in order to compensate for this compression.

Siegler and Opfer's studies highlighted the coexistence of both compressed and linear representational scales in young children (see also Lourenco & Longo, 2009, where the coexistence is found in adults), but their N-P conversion task suggested a shift to a linear responding around age eight. Our results indicate that, using the same kind of task investigating the number-space mapping in the proper context, an illusion of compression of numbers can be observed even in adults with spoken number words. That is, in the specific experimental context we used, we observe compressed responding on a task where participants were required to report their subjective judgment of numerical distance over the entire range of numbers tested. Previous work mostly focused on the mental representation of concrete numerosities such as sets of dots, in which case Weber's law is ubiquitous and readily lends itself to an explanation in terms of an internal compressive scale (Izard & Dehaene, 2007; Van Oeffelen & Vos, 1982; although see Cordes, Gelman, Gallistel, & Whalen, 2001; Whalen, Gallistel, & Gelman, 1999 for a different explanation). For symbolic numerals, prior evidence is scarce (for review and discussion, see Dehaene, 1992, 2007; Krueger, 1989). It has been shown that intuitions of prices conveyed as Arabic numerals follow Weber's law and can be interpreted as showing Gaussian variability on a logarithmic scale (Dehaene & Marques, 2002). Similarly, multidimensional scaling of number similarity judgments hints at a compressive mental number line (Buckley & Gillman, 1974; Shepard et al., 1975). Finally, and most relevant to the present work, judgments of randomness are biased toward compressed sequences (Banks & Coleman, 1981; Banks & Hill, 1974). Our results go beyond this prior work by showing that (1) a staircase procedure can be used to quickly and precisely quantify the amount of numerical compression within each participant; (2) the intuition of compression is a strong illusion that resists training (our N-P training task lasted 5–10 min, while our auditory feedback lasted approximately

Table 2. Functions that best fit participants' mean judgments in Experiments 4 and 5 before and after training

	Experiment 4	Experiment 5
Before training	$x = 4.66 \times n^{2.21} - 1.66$	$x = 4.29 \times n^{2.24} - 1.29$
After training	$x = 5.21 \times n^{2.16} - 2.21$	$x = 11.54 \times n^{1.82} - 8.54$

3 min); and (3) the illusion can coexist with linear responding on Siegler and Opfer's N-P task.

Why did adults behave nonlinearly in our task and linearly in the N-P task? Although linear responding may be observed in the N-P task with speeded responses (see Pinhas & Fischer, 2008), it is likely that the untimed version of the N-P task typically used allows for the application of arithmetic strategies (e.g., roughly dividing the target number by the upper bound of the interval to estimate its position on a linear scale). Education and training with number lines, rulers, graphs, and other mathematical devices probably strengthens and automatizes this strategic response mode to the point where it may be observed in adults even in speeded tasks. Siegler and Opfer themselves suggest that these factors could be influential and that their finding that the logarithmic function was a better fit to their data in younger children could be due to the fact that the task involved large numbers in the hundreds and thousands that the children were not used to manipulating on a physical line.

Several experimental factors may influence the manner in which participants respond on this task. The presence of the upper bound may make the participants more aware of the numerical interval and allow for the use of a division strategy, especially if the upper bound is a round number such as 1,000. However, our results do not confirm Banks' hypothesis that bounded intervals always lead participants to rely on a linear mental number line, since the mean exponent we obtained for bounded intervals was still significantly greater than 1.0. The fact that many factors that we have considered (the interval tested, the number of items in each series) were not held constant in Banks' studies may be responsible for the differences between their experiments and ours. In addition, the standard interval [1, 1,000] used by Banks when including the upper boundary (and the fact that the series in this experiment contained 10 numbers) may explain their results, since these experimental conditions tend to turn the task into a simple division problem. On the other hand, larger ranges may make it more difficult for the participants to judge the series and explain why Banks and colleagues seem to obtain slightly greater exponents (with intervals [1, 100,000] and [1, 2,000]) than we do here. However, bearing in mind the coarse sampling of possible exponents in the Banks' studies, these slight differences may reflect the fact that the fixed exponent chosen was not the exact value participants would have preferred, had they been presented with other similar exponents. Crucially, the use of a staircase procedure allowed us to detect differences (e.g., the exponent changes by 0.42 before and after feedback in Experiment 5) that were smaller than the steps between tested exponents used in Banks and Coleman's study (the smallest step is .56 in the range of exponents from 1.43 to 3.33).

Our task encourages estimation without calculation – a situation that may increase reliance on intuition of a compressed continuum. We suggest that education and experience with manipulating linear scales in mathematics encourages the strategic use of a linear scale, and that it is therefore necessary to use a difficult task that prevents explicit calculation in order to reveal an intuitive, logarithmic understanding of numerical magnitudes and distances. In our experiments, the speed with which the sequences were presented, the random order, and the nonstandard range of numbers, all made the application of mathematical rules difficult. The use of an auditory task, with all the numbers presented prior to the participants judging the sequences, is likely to force the participants to rely on their intuitions of numerical magnitude by taxing working memory. Furthermore, we explicitly required participants to try to respond spontaneously and to rely on their spatial intuition.

The results of Experiments 4 and 5 clearly indicate that training, even with explicit presentation of linear sequences, does not suffice to eliminate the illusion of compression. In previous studies of the N-P task (Opfer & Siegler, 2007), linear responses were obtained once participants were explicitly taught the spatial location of a target number. Perhaps training provided participants with precise anchor points that allowed them to adjust their subsequent evaluations to a linear scale (similar recalibration occurs in a dot estimation task; Izard & Dehaene, 2007). In our case, listening to linear sequences may have provided the same sort of anchor points, but only allowed participants to slightly correct their nonlinear responses. Furthermore, this strategy was ineffective when visual feedback was provided using the N-P task in Experiment 4, suggesting that there may be distinct modality- or task-specific representational scales for the mental number line.

Taken together, these experiments add further support to the existence, in adult human beings, of a nonlinearly compressed understanding of numerical magnitudes, which has a strong influence on the intuitive perception of numbers, and which is only partially overcome by education and training. These results are thus consistent with one proposed explanation for the observed shift from a compressed responding to a linear one, namely that with education, children learn to use arithmetic strategies to "linearize" their behavior on the N-P task. These findings thus suggest that there may be multiple understandings of numbers, both linear and compressed, and that such representations are accessed in a task- and/or modality-specific manner (see Cohen Kadosh & Walsh, in press and commentaries therein).

In Dehaene et al.'s study (2008), the Mundurucu adults with the most experience using Portuguese number words showed a striking dissociation in their pattern of responding. Whereas they continued to demonstrate a compressive mapping on the N-P task using Mundurucu number words, they performed linearly with Portuguese number words. Indeed, it was found that greater education significantly changed the responses to Portuguese number words from logarithmic to linear, yet left the responses to Mundurucu numerals and dot patterns unchanged. These results with the Mundurucu are consistent with our suggestion, based on results with educated French-speaking participants, that both linear and compressed ways of responding may coexist, and that they may be observed in different experimental contexts, or in the case of the Mundurucu, when using different languages. Thus, the results of any one paradigm should not be considered as giving a unique characterization of numerical representations, but rather as showing how a certain experimental context leads participants to rely on one out of the several representations that coexist in adult, literate participants.

References

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

Banks, W. P., & Hill, D. K. (1974). The apparent magnitude of number scaled by random production. *Journal of Experimental Psychology (Monograph series), 102*, 353–376.

Boersma, P., & Weenink, D. (2005). *Praat: Doing phonetics by computer (Version 4.3.20)* [Computer program]. Retrieved from http://www.praat.org/.

Buckley, P. B., & Gillman, C. B. (1974). Comparison of digits and dot patterns. *Journal of Experimental Psychology, 103*, 1131–1136.

Cohen Kadosh, R., & Walsh, V. (2009). Numerical representation in the parietal lobes: abstract or not abstract? *Behavioral and Brain Sciences, 32*, 313–328.

Cordes, S., Gelman, R., Gallistel, C. R., & Whalen, J. (2001). Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. *Psychonomic Bulletin & Review, 8*, 698–707.

Cornsweet, T. M. (1962). The staircase-method in psychophysics. *The American Journal of Psychology, 75*, 485–491.

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition, 44*, 1–42.

Dehaene, S. (2001). Subtracting pigeons: Logarithmic or linear? *Psychological Science, 12*, 244–246.

Dehaene, S. (2007). Symbols and quantities in parietal cortex: Elements of a mathematical theory of number representation and manipulation. In P. Haggard, Y. Rossetti, & M. Kawato (Eds.), *Attention & performance XXII*. London: Oxford University Press.

Dehaene, S., Dupoux, E., & Mehler, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 626–641.

Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures. *Science, 320*, 1217–1220.

Dehaene, S., & Marques, J. F. (2002). Cognitive euroscience: Scalar variability in price estimation and the cognitive consequences of switching to the Euro. *Quarterly Journal of Experimental Psychology, A, 55*(3), 705–731.

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition, 44*, 43–74.

Izard, V., & Dehaene, S. (2007). Calibrating the mental number line. *Cognition*, doi:10.1016/j.cognition.2007.06.004.

Jordan, K. E., & Brannon, E. M. (2006). A common representational system governed by Weber's law: Nonverbal numerical similarity judgments in 6-year-olds and rhesus macaques. *Journal of Experimental Child Psychology, 95*, 215–229.

Krueger, L. E. (1989). Reconciling Fechner and Stevens: Toward a unified psychological law. *Behavioral and Brain Sciences, 12*, 251–276.

Lourenco, S. F., & Longo, M. R. (2009). Multiple spatial representations of number: Evidence for co-existing compressive and linear scales. *Experimental Brain Research, 193*, 151–156.

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

Nachmias, J., & Steinman, R. M. (1965). An experimental comparison of the method of limits and the double staircase-method. *The American Journal of Psychology, 78*, 112–115.

Nieder, A., & Miller, E. K. (2003). Coding of cognitive magnitude: Compressed scaling of numerical information in the primate prefrontal cortex. *Neuron, 37*, 149–157.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*, 97–113.

Opfer, J. E., & Siegler, R. S. (2007). Representational change and children's numerical estimation. *Cognitive Psychology, 55*, 169–195.

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

Pinhas, M., & Fischer, M. H. (2008). Mental movements without magnitude? A study of spatial biases in symbolic arithmetic. *Cognition, 109*, 408–415.

Shepard, R. N., Kilpatrick, D. W., & Cunningham, J. P. (1975). The internal representation of numbers. *Cognitive Psychology, 7*, 82–138.

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

Van Oeffelen, M. P., & Vos, P. G. (1982). Configurational effects on the enumeration of dots: Counting by groups. *Memory & Cognition, 10*, 396–404.

Whalen, J., Gallistel, R., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science, 10*, 130–137.

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition, 74*, B1–B11.

Received March 12, 2009
Revision received July 22, 2009
Accepted August 18, 2009
Published online April 13, 2010

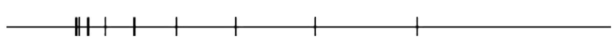
Arnaud Viarouge

Department of Psychology and Human Development
Vanderbilt University
Peabody #552
230 Appleton Place
Nashville
TN 37203-5721
USA
E-mail arnaud.viarouge@vanderbilt.edu

Appendix

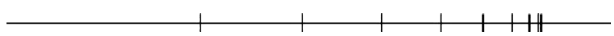
Verbatim Instructions (Translated From French)

You are about to hear several sequences of random numbers. For each of these sequences, you will have to judge the way the numbers are distributed. Thus, if you think that there are more numbers on the side of smaller numerosities (see Example 1), press the left-arrow key on the keyboard.



Example 1 (left-arrow key)

If, on the contrary, there are more numbers on the side of larger numerosities (Example 2), press the right-arrow key.



Example 2 (right-arrow key)

