Slow drift of individuals' magnitude-to-number mapping.

Edward Vul, David Barner, & Jess Sullivan (evul, barner, jsulliva@ucsd.edu)
Department of Psychology, 9500 Gilman Dr. # 109
La Jolla, CA 92093-109 USA

Abstract

When estimating the number of dots in a set, adults show bias and variability that scale with numerosity. Increasing variance in estimation is thought to reflect constant Weber noise on perceptual magnitude representations, while the increasing bias reflects miscalibrated mappings of number words onto magnitudes. Here we argue that response variability in numerical estimation increases with numerosity in part due to uncertainty and slow drift in the mapping of numbers onto magnitudes. We show that individuals' number-to-magnitude mapping functions drift slowly over the course of the experiment, with a shared-variance half-life of over 100 trials (∼ 10 min). We thus propose a model that treats the word-to-magnitude mapping function as a major source of estimation variability, and that accounts for cross-subject differences in estimation bias and variability, as well as changes to estimation performance within a given subject over time. In doing so, we reconcile the existing literature on the sources of estimation variability, and provide evidence that uncertainty in the word-to-magnitude mapping function is a key limiting factor in estimation performance.

Keywords: Approximate number, number words, numerical estimation

Introduction

Human adults have access to at least two systems for representing numerical quantity. The first is a noisy and evolutionarily ancient nonverbal number system, called the Approximate Number System (ANS; see Dehaene, 1997; Feigenson, Dehaene, & Spelke, 2004 for review). The second is the verbal number word system (e.g., one, two, three, etc.). This system unique to humans allows for the precise representation and manipulation of numerical content. When making estimates, adults draw on both of these systems: they use the ANS to represent the magnitude of the stimulus being estimated, and they use the verbal number system to attach a linguistic label to this magnitude.

The interface between these two systems has been an area of recent interest because it is crucial to characterizing how people can provide explicit verbal estimates of numerosity, and more generally how language relates to perception (e.g., Carey, 2009; Izard & Dehaene, 2008; Thompson & Opfer, 2011; Sullivan & Barner, 2012; Sullivan, Juhasz, Slattery, & Barth, 2011). The question is important for at least two reasons. First, estimation performance has been shown via intervention studies to be causally related to academic success, raising the question of why, and which aspects of training are most important to educational outcomes (Ramani & Siegler, 2011; Siegler & Ramani, 2009). Second, estimation tasks are often argued to elucidate properties of the ANS or their comprehension of number word meanings. In the present study, we tested what contribution – if any – this mapping function makes to estimation error. Specifically, we asked two questions about the nature of the number-to-magnitude mapping function. First, are the mappings between verbal and nonverbal numerical representations stable across individuals? Second, within individuals, are these mapping stable across time? As argued below, the answer to these questions suggests a novel model of numerical estimation, which explains significant aspects of error by appealing to a dynamically changing mapping function, rather than to the ANS.

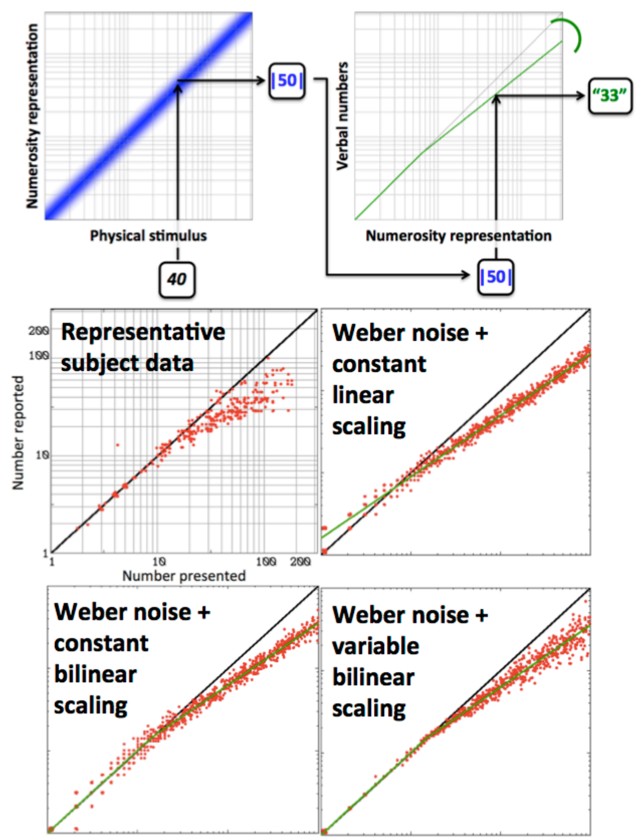

Figure 1: (Top) Our account of the bias and variability in human numerical estimation assumes two transformations between the physical stimulus and a verbal numerical response (following Izard & Dehaene, 2008). First, the approximate number system maps the physical stimulus onto a logarithmic magnitude estimate with constant Weber noise. Second, the magnitude estimates are mapped onto the verbal number; we approximate this mapping as bilinear in log-log space with a variable slope. (Bottom) Two novel features of this account are necessary to capture the patterns of errors in human estimation data (one representative subject shown): (1) the mapping function must be non-linear in log-log space, otherwise the pattern of veridical calibration for small numbers, and systematic mis-estimation for large numbers will not hold, (2) the slope of the high end of the mapping function must be variable to capture the increasing variability of estimation for larger numbers.

In the absence of training or feedback, adults are notoriously inaccurate estimators (Kaufman, Lord, Reese, & Volkmann, 1949; Izard & Dehaene, 2008; Minturn & Reese, 1951). For the purposes of the present project, we are interested in two attributes of this inaccuracy: variability and bias.

First, consider the variability of estimates. The degree of variability in estimates typically increases in proportion to the magnitude of the stimulus being estimated, such that the coefficient of variation (the ratio of the standard deviation of estimates for a given magnitude to the mean estimate of that magnitude) remains constant as magnitude increases (e.g., Whalen et al., 1999). Because variability in estimation behavior scales up with magnitude, estimates are typically said to demonstrate the property of scalar variability.

Scalar variability is thought to arise from Weber noise in the ANS: ANS representations are ratio-dependent, and therefore error in its representation of number also scales with number. For example, it is as easy to tell the difference between 5 dots and 10 dots as it is to tell the difference between 500 dots and 1000 dots using the ANS. Because both nonverbal ANS tasks and verbal estimation tasks exhibit scalar variability, many have concluded that scalar variability in estimation arises because the underlying (ANS) perceptual representations of the magnitudes being estimated exhibits weber noise (Dehaene & Marques, 2002; Izard & Dehaene, 2008; Le Corre & Carey, 2007; Negen & Sarnecka, 2010; Siegler & Opfer, 2003) and relatively uncontroversial.

It is clear from the past literature that much of the variability found in estimation arises from variability in the ANS representations that support estimates. However, in the present paper, we ask whether all of the variability in estimation performance is explained by Weber noise, or, alternatively, whether the mapping function that connects the verbal and nonverbal number system also contributes variability to estimation performance. One reason to believe that variability and bias may arise in part from the mapping function between number language and the ANS is that feedback (e.g., showing a participant an example) reduces estimation variability in both children and adults (Barth, Starr, & Sullivan, 2009; Krueger, 1984; Izard & Dehaene, 2008; Lipton & Spelke, 2005) a finding that one might not expect if variability arose entirely from the ANS. A second reason to believe that estimation variability stems entirely from the ANS is that sometimes – as in the data set we present in this paper – the coefficient of variation (CoV) does not remain constant across all estimates. However, the degree to which estimation variability stems from the word-to-number mapping function remains untested.

Next, consider estimation bias. One frequent finding in the estimation literature is that estimates tend to be biased (e.g., systematically too high or too low). Also, bias in estimation performance tends to increase over the course of an experiment. For example, adults often underestimate magnitudes from the very first trial of an estimation experiment. When they do, this underestimation bias persists and is amplified over the course of the experiment (Krueger, 1982). In fact, even when the degree and direction of estimation error made early in an experiment is experimentally manipulated, bias introduced in the first few trials endures throughout the duration of the entire estimation experiment (Barth et al., 2009; Izard & Dehaene, 2008; Krueger, 1984; Lipton & Spelke, 2005; Shuman, unpublished thesis; Sullivan & Barner, 2012; Sullivan, Juhasz, Slattery, & Barth, 2011). This influence of miscalibration is often described as stemming from changes to the number-to-magnitude mapping function. However, the nature of this change in the mapping function is still poorly understood. Specifically, it remains unknown how and why the degree of estimation bias increases as more estimates are made.

In the present study, we probed the factors that influence errors in estimation performance, with a special focus on the variability and bias found in individual participants estimates.

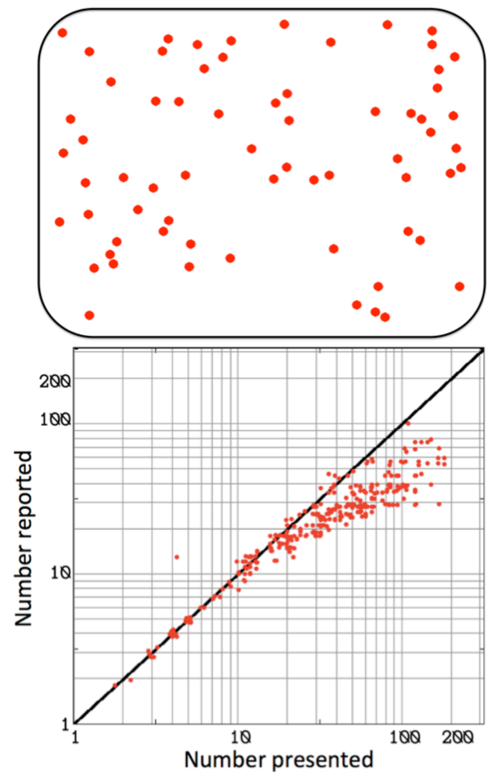

Figure 2: (Top) Participants saw 300 trials in which an array of n dots were briefly presented with the number of dots chosen according to a geometric distribution. Then participants made a guess as to the number of dots presented. (Bottom) A representative subject's data over all 300 trials with number presented (log scale) on the x-axis and number reported (log scale) on the y-axis. We investigate the sources of bias and variability evident in these patterns of mis-estimation.

How do people estimate a quantity of dots?

As already noted, patterns of bias and variability in estimation arise from both numerical perception and the mapping of these magnitude representations to language (see Figure 1). Our proposal is an amendment of Izard & Dehaene's (2008) int that we argue that (1) the slope of the higher portion of this mapping function is variable over time, and (2) uncertainty in this mapping causes it to vary across time, thus introducing additional variability in estimation tasks that increases with numerosity.

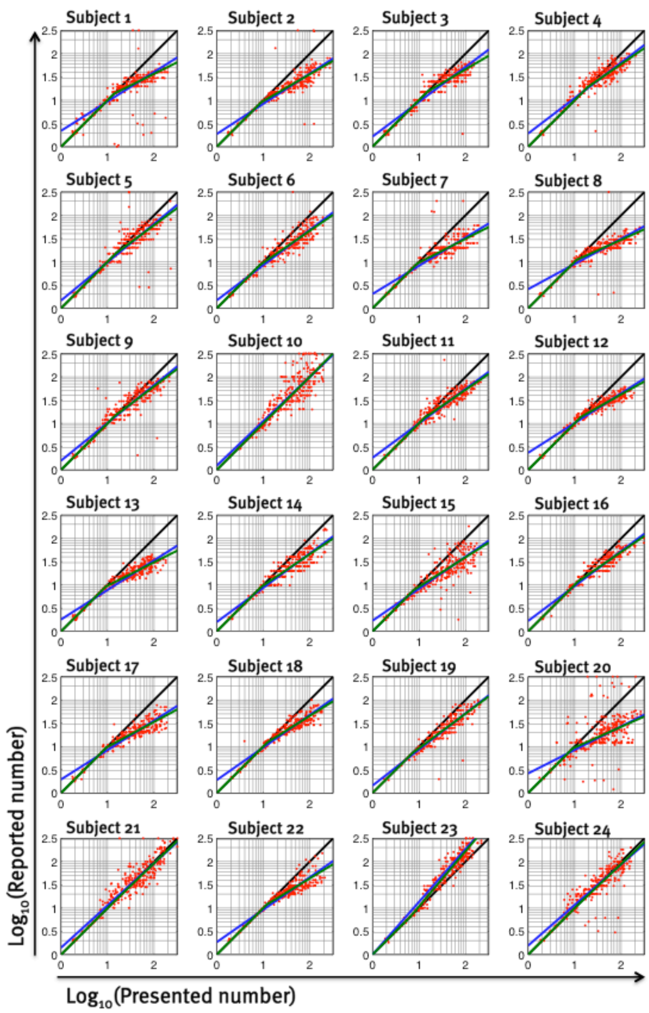

Figure 3: Individual subject estimation data (red points) along with best fitting linear (blue) and bilinear (green) mapping functions. Some of our conclusions may be seen in the raw data alone: (1) Variability in log-log space increases with numerosity. (2) Systematic mis-estimation occurs for larger, but not smaller, numbers. (3) Individual subjects have relatively stable idiosyncratic mis-estimation biases.

For our analyses, we consider this mapping to be bilinear in log-log space: it is veridical (falling on the identity line) for relatively small quantities (e.g., those smaller than 10; Sullivan & Barner, 2012), but then tends towards underestimation for higher numbers. It is not central to our proposal that individuals actually use a strictly bilinear mapping function rather than a more complex mapping – however, our data do not offer the resolution necessary to assess whether a more sophisticated mapping function is used. For our purposes, approximating this function as bilinear makes our results and analyses simpler to describe.

This account offers a means to reconcile previous disagreements in the literature on the approximate number system and numerical estimation. First, while the coefficient of variation (CoV, Weber fraction) is constant for the approximate number system, it is not known to be constant in verbal estimation tasks. For example, in previous work on word-to-magnitude estimation tasks, CoV has typically been analyzed only for a subset of the number-line (Izard & Dehaehe, 2007; Le Corre & Carey, 2007), or is found to increase with numerical magnitude (Siegler & Opfer, 2003), or is not reported at all. In the present paper, we present a dataset in which the coefficient of variation in estimates increases with numerosity, suggesting a non-constant Weber fraction. Our account helps to reconcile a constant CoV in the approximate number system with an apparent increase in CoV in estimation, by showing that the coefficient of variation in estimation tasks is driven by variability in the mapping function over time (note that because discrimination tasks don't require verbal responses, they circumvent this mapping and its associated variability). Second, there is some disagreement as to whether there is a stable (consistent across the numerosity scale) mapping of magnitudes on to verbal numbers (Izard & Dehaene, 2008; Lipton & Spelke, 2005; Sullivan & Barner, 2012). Our proposal suggests that, while the mapping of magnitudes to numbers may be consistent across a range of magnitudes at any one point in time, this mapping is not stable over time, yielding inconsistent behavior over trials.

Figure 1 provides an illustration of our model, and shows predictions from reduced classes of this model as compared to one (representative) subject's data. With only constant Weber noise and a stable linear mapping of magnitudes to numbers, we would erroneously predict overestimation and excessive variability for small numbers. Even with a bilinear mapping, constant Weber noise would predict less variability for large numbers and more for small numbers. Only with a variable slope in a bilinear mapping function can we account for the pattern of miscalibration and increasing variability for large, but not small, numbers. In the subsequent section we describe analyses that explicitly test these claims.

Experiment Methods

Twenty-four subjects recruited from the UC San Diego psychology department pool participated in an hour-long experiment in which they had to guess the number of dots presented onscreen on each trial (see Figure 2). The number of dots shown was sampled on each trial from a geometric distribution with a mean of 50, truncated at the low end so that displays had at least two dots. All the dots in an array were the same size (radius of 10 pixels), presented in red on a white background. The configuration of dots was randomly generated by drawing locations from a uniform distribution over the full display area (1024x768 pixels) with the constraint that the dots did not overlap. On each trial the array of dots was presented for 250 msec, and then subjects were prompted to type in their guess as to how many dots were in the array. Subjects were then asked to type in a second guess about the number of dots in the array. (Our analyses throughout the paper focus on the first of the two guesses, but our conclusions hold if we consider the second guess alone, or the average of the two). Figure 2 also shows one representative subject's data from all 300 trials.

Results

The responses of all 24 participants for all 300 trials in our experiment are shown in Figure 3 on log-log coordinates. Several features of the data immediately jump out. First, estimate variability goes up as a function of number. Since this is an increase in variability in log-log space, it is not consistent with a constant coefficient of variation (a constant Weber fraction) which predicts that variability would be constant in log-log space. Second, while subjects are well calibrated for small numbers, there is a tendency to underestimate for large numbers: most subjects underestimated, but some subjects showed fairly reliable overestimation or veridical average calibration. Third, individual differences in under- or over- estimation appear to be quite reliable. These features are consistent with our account of a variable mapping function, which we elucidate in further analyses below.

Is magnitude-to-number mapping bilinear?

We propose that the mapping function is not linear in log-log space, but bends such that small magnitudes are mapped more or less veridically onto number words, but large numbers show a systematic deviation from the identity line. While we do not believe that the true mapping function that people entertain is strictly bilinear, we do believe that it is not simply linear in log-log space. We can show that a bilinear function that is veridical (falls on the identity line) up to some critical number (c), and then deviates from the identity line with some log-log slope of s can account for data of individual subjects much better than a simple line with an intercept (a) and a slope (b). Since these models both have two parameters, we can simply compare the R^2 values of individual subjects. Although the average R^2 values are similar (0.79, vs 0.81), bilinear fits better describe the data for 20 of 24 subjects (binomial test: $p = 0.0015$), see Figure 3.

This piecewise-linear mapping function could indicate a number of possible processes. Perhaps small numbers (less than about 10 – the average point of departure from veridical mapping across subjects) are not part of the mapping between the approximate number system and words, and instead gain their content from estimates made via subitization (e.g., in the company of chunking). Another possibility is that the mapping function is constrained by previous data which clearly disambiguate the numerosity/numbers correspondence, and that lower numbers have more data, and thus fall on the identity line, while higher numbers are constrained only by a requirement for smoothness and monotonicity. A final possibility is that the nature of the mapping function between the ANS and the verbal number system is qualitatively different for relatively small numbers and relatively larger numbers – for example, participants might rely more strongly on item-based associative mappings for numbers smaller than 10, but more on a structural mapping between magnitudes and the count list for larger numbers (Sullivan & Barner, 2012). We slightly disfavor the first alternative, because the cut-off point between accurate and miscalibrated mapping does not seem to correspond to other cut-offs previously postulated to distinguish between qualitatively different numerosity processes (such as subtilizing and approximate magnitude – Feigenson et al., 2004).

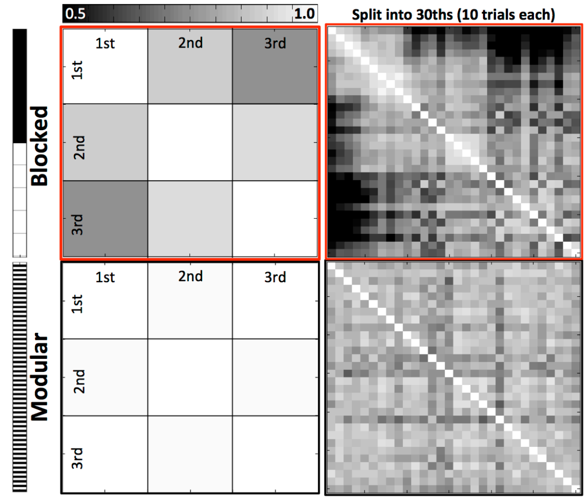

Figure 4: We assess whether variability arises from a slow drift in the mapping function over time by estimating the slope of a bilinear mapping function for different subsets of the 300 trials for each subject. For instance, if we split the 300 trials into thirds (left half), then each third contains 100 trials, if we split into thirtieths (right half) then each thirtieth contains 10 trials. A blocked split (top half) corresponds to taking consecutive portions of the 300 trials: e.g., the first 3rd contains trials 1 through 100, the second 3rd contains trials 101 through 200, the third 3rd contains trials 201 through 300. A modular split (bottom half), corresponds to taking every nth trial, such that the full range of the experimental session is represented in each subset: e.g., the first 3rd contains trials 1, 4, 7, 10, ... 298, the second 3rd contains trials 2, 5, 8, 11, ..., 299, and the third 3rd contains trials 3, 6, 9, 12, ..., 300. We compute the across-subject correlation of slope estimates taken from each subset of trials. Darker colors indicate lower correlations, brighter colors indicate larger correlations. Several observations in these heat maps are indicative of a gradual drift in slopes over time within a given subject. (1) Modular splits yield higher across-subject correlations than blocked splits, suggesting that the blocked splits are subject to additional variability due to a gradual change that modular splits avoid. (2) Blocked splits show decreasing correlations as a function of distance: correlations further from the diagonal are lower – the correlation between the first and second third is higher than the correlation between the 1st and 3rd third – indicating that the correlations are slopes are changing slowly over time.

How reliable is the across-subject variation in the shape of the mapping function?

We assess the across-subject reliability in shape of the bilinear mapping function via a split-half analysis: we estimate the mapping function (particularly the slope of the higher bilinear portion) in individual subjects in 50% of the trials and assess the across-subject correlation across those pairs of estimates, to see whether variation of mapping functions is reliable.

First we assess *Blocked* split-half reliability: we divide the 300 trials into the first half (trials 1-150) and the second half (151-300). The Blocked split-half across-subject reliability of the estimate of the slope of the erroneous part of the mapping function that does not fall on the identity line was highly significant ($r = 0.83$; $t(22) = 6.98$, $p < 0.001$), indicating that people are very consistent in their idiosyncratic magnitude-to-number mapping errors.

We next assessed *Modular* split-half reliability: Modular split-half divides the 300 trials into odd trials (1, 3, 5, ..., 299), and even trials (2, 4, 6, ..., 300). In contrast to Blocked splits, in which trials in a given half of the data are are contiguous and arise from from different portions of the experiment (separated by an average of 150 trials), in Modular splits, the trials in a given half are taken from the full range of the experiment. Modular splits-half across-subject reliabilities were much higher than those for Blocked split-half analyses ($r = 0.97$; $t(22) = 18.7$, $p < 0.001$ – the difference is highly significant using the Fisher r-to-z transform: $z = -2.74$, $p = 0.0061$).

The difference between Modular and Blocked split-half reliability is our first indication that the slope of the magnitude-number mapping function is not stable within individuals over the experimental session: If the slopes we estimate for the mapping function drift over time, we expect that Blocked splits should yield a lower split-half across-subject reliability than Modular splits, because the Blocked splits are taken from different points in time, and would reflect different states of drift of the mapping function, while Modular splits would not.

Does the mapping function vary over time?

We argue that some of the increase in variability of estimates with increasing number arises from variability of the *mapping function* over trials, rather than simple misperception of the approximate magnitude of an individual array. In this section we argue for this view because the internal mapping function drifts slowly over the course of many trials, and we can measure its variation over the course of an experimental session.

To more precisely measure the drift of the mapping function over time, we generalize the Blocked vs. Modular split-half analysis to Blocked vs. Modular split-*n*ths for $n = \{3, 5, 10, 15, 20, 30\}$ (e.g., for split-30th we divide our 300 trials into 30 subsets, each one comprising 10 trials, for instance, the 5th Blocked split-30th subset will contain trials 41-50, while the 5th Modular split-30th subset will contain the 10 trials: 5, 35, 65, 95, 125, 155, 185, 215, 245, 275). By obtaining split-*n*th reliability for Blocked subsets taken from different portions of the experimental session, we can assess how the reliability of the number-mapping slope decreases as a function of time.

We calculate the across-subject slope reliability across different subsets (represented as a matrix in Figure 4), the Blocked split-*n*th reliability between subset 1 and subset 2 measures the across-subject correlation of slopes estimated from two adjacent periods of time in the session which are on average separated by $300/n$ trials. In general if we calculate the correlation between subset i and subset $i+k$ from a Blocked split-*n*th analysis, those subsets are separated by $300 * k/n$ trials. Thus, if slopes are gradually drifting over the course of the experimental session, we would expect across-subject reliability of slope estimates to decrease with k – the separation between Blocked subsets. Nothing of this sort should happen for Modular subsets which contain overlapping trials intermixed over the whole session.

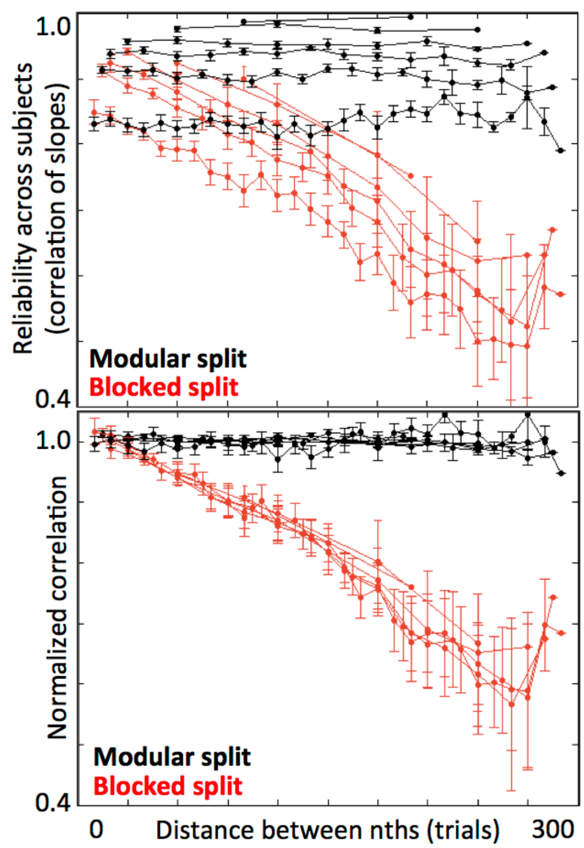

Figure 5: (Top) We can assess the average across-subject correlation in slope estimates as a function of distance between blocks for different splits of the data. A blocked split of the data into thirds yields a two measures of the slope-correlation at an average distance of 100 trials (1st to 2nd block and 2nd to 3rd block) and one measure of the correlation at a distance of 200 trials (1st to 3rd block). If we split the data more finely (here going up to 30ths, as seen in figure , we can more finely measure the drop-off of average correlation as a function of distance. Distance is meaningless for Modular (points in black) splits, but the same analysis can be carried out to measure how much the correlation drops merely as a function of using fewer trials for each slope estimate. (Bottom) we can normalize both blocked (red) and modular (black) correlations to the average of the modular split correlations to make the decrease in correlation with distance as seen in the red lines comparable across splits with different numbers of trials per subset. Correlations drop off with distance very slowly, but even when separated by 290 trials, 10-trial estimates of subjects' mapping function slopes show reliabilities well above 0.

Figure 5(top) shows the split-*n*th reliability for Blocked (red) and Modular (black) subsets as a function of their separation (k). For instance, we estimate the average Blocked split-10th correlation at a separation of $k = 2$ as the average of the across-subject correlations taken between the 8 subset comparisons separated by 2: subset 1 and subset 3, subset 2 and subset 4, ... subset 8 and subset 10. This average would appear at $x = 300/n * k = 300/10 * 2 = 60$. Several features are apparent from the changes in slope reliability across subsets separated by more time: (1) when we split into more subsets both Modular and Blocked correlations drop, since each subset necessarily contains fewer trials to estimate the slope; (2) as expected, only Blocked correlations decrease as a function of distance between subsets. To more clearly display the decrease in reliabilities as a function of subset distance, while factoring out the reduced reliability due to smaller trial-counts within each subset, by normalizing reliabilities by dividing them by the average reliability seen across Modular splits. This yields Figure 5(bottom), which shows the slow decrease in reliabilities over the 300 trials.

A linear regression on the raw correlations in the Blocked split-*n*ths as a function of separation (measured in trials) is significantly negative (95% confidence interval on the slope: $(-0.0015, -0.0012)$ change in correlation per trial, $F(1,22) = 358$, $p < 0.001$). Despite this highly significant decrease, it is very slow over the course of the session, and even mapping function slope estimates based on 10 trials separated by 290 trials show significant across-subject reliability ($r = 0.57$; $t(22) = 3.254$, $p = 0.002$).

Together, these results indicate that subjects' mappings of magnitudes onto verbal numbers drift slowly over time.

Conclusions

We have shown that subjects map numbers onto the verbal number line via a piecewise-linear function in logarithmic representations (piecewise consistent with Stevens' power law). Our results are consistent with two-transformations mapping physical numbers onto number estimates: first physical numbers are represented logarithmically in the approximate number system, and then the approximate number system is mapped onto the verbal number line through an unstable mapping. For small numbers, the mapping appears to be fairly stable and veridical, perhaps due to the considerable amount of evidence people have previously seen for small number estimates. For higher numbers, the mapping is not veridical, and tends to drift slowly over the course of many trials; the variability of the mapping function over time causes increasing estimation variance for large numbers, and may thus resolve theoretical disagreements about the constancy of variability in the approximate number system.

Acknowledgments: EV was supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of the Interior (DOI) contract D10PC20023. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DOI, or the U.S. Government. DB was supported by a grant from the James S. McDonnell Foundation. JS was supported by an NSF GRFP fellowship.

References

Barth, H., Starr, A., & Sullivan, J. (2009). Childrens mappings of large number words to numerosities. Cognitive Development, 24, 248-264.

Carey, S. (2009). The Origin of Concepts. Oxford University Press: New York.

Dehaene, S. (1997). The Number Sense: How the Mind Creates Mathematics. Oxford University Press: New York.

Dehaene, S., & Marques, J. (2002). Cognitive euroscience: Scalar variability in price estimation and the cognitive consequences of switching to the euro. The Quarterly Journal of Experimental Psychology, 55, 705-731.

Ebersbach, M., Luwel, K., Frick, A., Onghena, P., Verschaffel, L. (2008). The relationship between the shape of the mental number-line and familiarity with numbers in 5- to 9-year old children: Evidence for a segmented linear model. Journal of Experimental Child Psychology, 99, 1-17.

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

Kaufman, E., Lord, M., Reese, T., & Volkmann, R. (1949). The discrimination of visual number. The American Journal of Psychology, 62, 498-525.

Krueger, L. (1982). Single judgments of numerosity. Perception and Psychophysics, 31, 175-182.

Krueger, L. (1984). Perceived numerosity: A comparison of magnitude production, magnitude estimation, and discrimination judgments. Perception and Psychophysics, 35, 536-542.

Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. Cognition, 105, 395-438. 9, 159-172.

Lipton, J., & Spelke, E. (2005). Preschool childrens mapping of number words to nonsymbolic numerosities. Child Development, 76, 978-988.

Minturn, A., & Reese, T. (1951). The effect of differential reinforcement on the discrimination of visual number. Journal of Psychology, 31, 201-231.

Negen, J., & Sarnecka, B. (2010). Analogue magnitudes and knower-levels: Re-visting the variability argument. Proceedings of the 32nd Annual Conference of the Cognitive Science Society, Austin, TX: Cognitive Science Society.

Ramani, G., & Sieger, R. (2011). Reducing the gap in numerical knowledge between low- and middle-income preschoolers. Journal of Applied Developmental Psychology, 32, 146-159.

Shuman, M. (unpublished thesis). Computational characterization of numerosity perception and encoding.

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

Siegler, R., & Ramani, G. (2009). Playing linear number board gamesbut not circular onesimproves low-income preschoolers numerical understanding. Journal of Educational Psychology, 101, 545-560.

Sullivan, J., & Barner, D. (2012). How are number words mapped to approximate magnitudes? Quarterly Journal of Experimental Psychology, 66, 389-402.

Sullivan, J., Juhasz, B., Slattery, T., & Barth, H. (2011). Adults number-line estimation strategies: evidence from eye movements. Psychonomic Bulletin and Review, 18, 557-563.

Thompson, C., & Opfer, J. (2010). How 15 hundred is like 15 cherries: Effect of progressive alignment on representational changes in numerical cognition. Child Development, 81, 1768-1786.

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