

Cutting In Line: Discontinuities in the Use of Large Numbers by Adults

David Landy (dlandy@indiana.edu)

Department of Psychological & Brain Sciences
Indiana University, Bloomington IN

Arthur Charlesworth (acharles@richmond.edu)

Department of Mathematics and Computer Science
University of Richmond, Richmond VA

Erin Ottmar (erin.ottmar@richmond.edu)

Department of Psychology
University of Richmond, Richmond VA

Abstract

Perceptual tools such as telescopes allow the application of robust internal perceptual systems to apply beyond the range of their unadorned capacity. This paper explores how reasoning over culturally provided representations enables the perception of conceptually distant structures. In particular, this paper examines the behavior of typical adults estimating the position of large numbers (1 thousand to 1 billion) on a number line. Participants—even those who closely match linear placement—show discontinuities in placement in the immediate vicinity of 1 million. This pattern was predicted by a theoretical account in which linear behavior across many orders of magnitude is achieved through highly linear patterns of placement on smaller lines that are recycled and scaled to larger numerosities. Just as the telescope allows perception of the imperceptibly distant, reasoning processes over the natural numbers appear to allow intrinsically limited magnitude-perception systems to apply (with distortion) to much larger scales.

Keywords: mathematical cognition, concepts, high-level reasoning, numerical cognition

Introduction

Much reasoning in mathematics involves taking structures well-defined in particular concrete domains and extending them to new less accessible contexts. For example, exponents are often presented initially to learners as repeated multiplications: $x^2=x*x$. This idea is then extended to zero exponents, fractional and real exponents, and even complex-valued exponents. This paper focuses on a very elementary instance of the extension from the concrete to the abstract: the extension of the natural numbers beyond the feasibly countable range. Mathematics often deals with numbers far outside any normal experience. For instance, currently the largest known prime number (the 48th Mersenne prime) would contain 17,425,170 digits if written as an Arabic numeral.

It is easy enough to order these numbers from smallest to largest; it is much, much harder to have any sense of their actual size. Nor is their actual size of any importance, except for practical reasons of computation. Almost exclusively, when magnitude is important for very large numbers, it is in comparison to other numbers related by some thread of reasoning in the same problem.

In this paper, we examine much more prosaic “big numbers”: those numbers just beyond the limit of practical countability—in the range of about 10^6 - 10^{12} . These numbers fall into an important boundary: so large that

experiences of this many individual items are vanishingly rare, but small enough to play important roles in sciences such as geology, astronomy, and macroeconomics, and in political contexts such as budget discussions. There is little reason to think that evolution would have specially prepared us to deal with quantities of this magnitude. How do we come to access them, and take advantage of their properties?

Representations that Contribute to Quantity

For small numbers, it appears that matching numbers to magnitudes requires the coordination of several systems. For instance, small exact set sizes may result from the active coordination of memorized count lists, pointing procedures, and a perceptual system that responds selectively to numerosity (Feigenson, Dehaene, & Spelke, 2004; Carey, 2009). Although exact numbers over 20 are typically not well linked to perceived set sizes (Izard & Dehaene, 2008; Sullivan & Barner, 2012) the linearity of the metric scale of the numbers extends out much further (Siegler & Opfer, 2003). Still, at some point even a log-based neural scale must run out. We simply cannot have lognormal response nodes that span the natural numbers, since our brains are finite in size, nor can we have created the entirety of the natural numbers through a completed infinity of recursions. Furthermore, typical environments do not require the individuation of sets as large as 10^9 (1 billion). How do we deal with these quantities when we do encounter them?

One possibility is that we induce new tokens as needed throughout life, extending a process essentially identical to that used to induce numbers small enough to encounter frequently (Leslie, Gelman, and Gallistel, 2009; Gelman, 2011). We might call this the *domain continuity hypothesis*. Numbers in the range of 10^9 might be recursively generated through successorship. On this account, our cognitive systems do not represent a completed infinity, but represent the natural numbers through unbounded extensibility. A plausible but (to our knowledge) novel hypothesis is that a lognormal representation system could be extended to include large numbers, as needed, by the creation of new, log-normally responsive tokens with some spacing. Representing the entirety of numbers up to 10^{12} this way requires less than 100 times the resources required to get from 2-3. While we cannot hope to construct a representation of a large Mersenne prime this way, we might well deal with large government deficits using the same log-normally distributed quantity representations that seem to be shared across many other species. Alternatively,

to the degree that linear representation schemes and logarithmic internal resources are combined to form linear number representations for smaller numbers (Carey, 2009; Thompson & Opfer, 2010), a plausible domain continuity hypothesis might extend this process to larger numbers, extending linear behavior out to any desired range.

Alternatively, it might be that when numbers exceed some endogenous or exogenous boundary, the manner in which neural resources are coopted to capture magnitude qualitatively shifts from a more-or-less direct one-step mapping from the metric structure of quantities or number words into spatial layout, to a more complex strategy. That is, people may reason about how to line up number words and number lines. Landy, Goldin, and Silbert (2013) found evidence for such mediated reasoning processes in the behavior of college-aged adults and adults recruited online. On a number-line placement task with boundaries of 1 thousand and 1 billion, about 40% of participants placed marks in a ‘piecewise linear’ pattern: the position of 1 million was very wrong (about 37% from the left edge of the page), but numbers between 1 million and 1 billion were placed extremely linearly, as were numbers between 1 thousand and 1 million. Note that on this task, also used in the current article, the correct linear location of 1 million is quite far to the left. Because there are 1 thousand millions in 1 billion, the location of 1 million lies about one thousandth of the way from one thousand to one billion.

One interpretation is that the “piecewise linear” participants were successfully applying their understanding of linearity over the two sub-ranges and simply adjoining the two line segments to yield a combined mapping¹, with the millions range slightly larger than the thousands. We will call this a *reuse hypothesis* (Anderson, 2010), since the resources normally used to process small numbers are rearranged and recycled (rather than extended) to handle numbers outside their typical domain.

Behavioral Predictions

The central theoretical question here concerns the proportion of participants who respond nearly linearly on larger number lines. How do they achieve this accuracy? There are two clear possibilities: one is that people continue the process of constructing linear magnitude representations that is used to construct representations of smaller range. On the other hand, it may be that people shift strategies, and that linearity is achieved through the deliberate cooption and reuse of pre-existing processes.

If linearity is achieved through strategic deployment of small-number resources, then linear-like behavior is, like piecewise behavior, achieved through the use of *two* lines:

¹ Several minor points are worth noting: Nearly all participants in these populations can correctly model the relevant number words as numerals, and vice versa. In Landy et al 2013, results were similar when *all* stimulus numbers were over 1 million, suggesting that these patterns are not a result of particular stimulus distributions. Analogous results obtained when the endpoints were 1 and 1 billion instead of 1 thousand and 1 billion.

Table 1. Stimuli used in the Experiment

Number Range	Stimuli Used
Thousands	10, 60, 100, 150, 230, 250, 310, 380, 420, 480, 500, 580, 640, 680, 720, 780, 840, 890, 940, 950
Millions	1, 2, 3, 4, 60, 100, 150, 230, 250, 310, 380, 420, 480, 500, 580, 640, 680, 720, 780, 840, 890, 940, 950

one for numbers under, and one for numbers over 1 million (more generally, the theory posits one line for each number range involved in the task). This strategy raises a coordination problem not present in other versions of this task: the right end of the ‘thousands’ line must be aligned with the ‘left’ end of the millions line. The four panels of Figure 1 indicate possible outcomes. The left panel indicates truly linear behavior. Note however that the x-axis has been scaled quite unusually. We have highlighted behavior around 1 million by placing 1 million at the middle of the x-axis, and scaling the rest of the axis linearly. This allows the examination of the theoretical predictions more easily than log-scaling the axes, since this way the predictions involve straight-line behaviors. In the leftmost panel of Figure 1, the location of 1 million has been correctly placed by the hypothetical participant very close to the left hand edge; judgments are linear for smaller and larger numbers.

The next three panels show varying kinds of discontinuity. The second panel indicates the kind of behavior reported by Landy et al (2013): a single fixed ‘million’ location, with linear behavior left and right of it. The right two panels illustrate the coordination problem mentioned above: if two separate lines are adjoined to accomplish the task, then there is no reason the left edge of one should align with the right edge of the other. There might be a gap (third panel) or overlap (rightmost panel).

These two theoretical accounts can be discriminated by examining closely the boundary around 1 million. On the *domain continuity hypothesis*, performance should be very close to linear, and any strong deviations are likely to be symmetric and continuous, roughly fitting power-law or linear performance (Barth & Paladino, 2011; Opfer, Siegler, & Young, 2011). On the *reuse hypothesis*, although performance in the aggregate may be linear for some participants, both the linear and non-linear responders should show evidence of ‘joining’ their lines: there should be discontinuities in placement behavior in the vicinity of 1 million. On this account, participants must first make a judgment about which line a particular item belongs on—whether the element is smaller or larger than 1 million. After this, they place the mark appropriately relative to the endpoints of the selected line segment. If so, linear responders might still have a ‘cut’ in the line at one million, but know where the cut goes. One very simple version of this would be to simply use two lines with the left endpoint of both simply treated as 0. This would cause an overlap, but on a line from 1,000 to 1 billion the deviation from linearity for the stimuli used would be less than one percent.

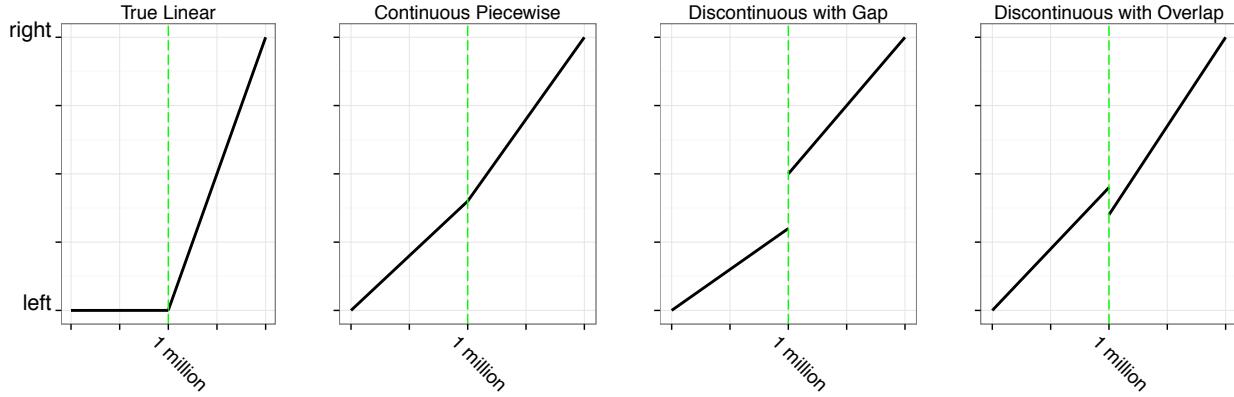


Figure 1: Possible response patterns on the number line placement task. Unlike the number line shown to participants, which showed 1 thousand and 1 billion without showing 1 million, the x-axis here is scaled to place 1 million at the center, and linearly scale on each side, which facilitates detection of discontinuity at 1 million. The left panel indicates truly linear behavior. The next panel indicates continuous placement with 1 million shifted to the right of its normative position. The two right hand panels illustrate two possible discontinuities. The rightmost panel indicates a non-monotonicity, in which some numbers in the thousands are placed to the right of some numbers over 1 million.

Experiment

Method

Participants. 200 participants were recruited from Amazon’s Mechanical Turk (*MTurk*). *MTurk* is an online marketplace in which participants volunteer to complete typically short online tasks in exchange for typically slight compensation. This task took about 20 minutes to complete. In general, participants recruited from *MTurk* have been found to behave similarly to other participants on a range of cognitive tasks when experiments are carefully conducted (Crump, McDonnell, & Gureckis, 2013), and have been extensively used as subject populations in prior work.

Design. Instructions showed an image of a small line (labeled from 0-8), and indicated the placement of ‘6’ as a sample. Participants were informed that the endpoints would be larger than the sample.

Participants were then shown a number line with “1 thousand” under the left end, and “1 billion” under the other. Because the study took place on Mechanical Turk, the physical length of the stimulus line cannot be determined. Participants were sequentially presented numbers in a random order, and selected with the mouse their chosen location for each number. Participants made 182 number line placements. Stimulus numbers were selected to sample the ranges under 1 million and over 1 million roughly evenly. Twenty numbers under 1 million, and twenty numbers above it, were chosen to be integers with one or two significant non-zero digits, and to be close to uniformly spaced within the two subranges. Because the numbers in the vicinity of 1 million were of particular interest, the range just over 1 million was over-sampled: the exact numbers 1 million, 2 million, 3 million, and 4 million were also included. Landy et al, (2013) found little shift in participant behavior in response to adjustments of the range of stimuli; the same was expected here.

The experiment started with 10 warm-up trials, with distinct stimuli in the same range as the test stimuli. Each stimulus was estimated 4 times by each participant; judgments were untimed and separated into blocks of 43 unique stimuli. Because of variations in screen size, a line with a fixed small number of pixels was used. The stimuli presented in the test phase are presented in Table 1.

Because the effect of number representation is of interest here, format was manipulated between participants: 100 participants received *numerals*, such as “54,000,000”; the other 100 received *hybrid notation* stimuli, as in “54 million”. The two stimulus types essentially serve as independent samples to validate conclusions. Results indicate that there were no noticeable differences between formats, so they are collapsed here.

Analysis

Data were analyzed in several steps. First, large “order of magnitude” errors were culled. Second, people were individually classified into linear and piecewise groups; these groups were further subdivided to isolate groups of highly linear responders. Finally, each individual’s responses were separately fitted to behavioral models using a maximum-likelihood procedure, and the models were compared using a likelihood ratio test to find the best-fitting model.

Culling of Order Errors. Several responses were highly compatible with the idea that the participant mis-encoded the order of magnitude, e.g., by reading “thousand” for “million” and vice versa. A two-step process was used to prune these data: first a piecewise linear model was fitted to the data for each subject (see below). Then, if a data point fit the predicted position for an order of magnitude error better than it fit the predicted position for the actual stimulus, it was removed from analysis. Then, the models were refit to the pruned data. This cleaning process made

the results more precise, but affected none of the conclusions reached here.

Categorization of Participants. Participants were divided into four groups. The first partition was based on whether the participant responses best fit into the linear or the piecewise cluster (a threshold of 0.3 for the estimated million point was used as a rough partition). Participant responses were very well fit by either the linear or piecewise linear patterns; however, these responses were distributed bimodally, with one cluster of participants behaving relatively linearly, and a second broader cluster centered around 0.4 (40% of the way from the left-hand endpoint) (Landy et al, 2013). Because we are here interested in the behavior of especially highly linear people, we further divided each of these groups by a median split, leaving one cluster with “million points” of less than 0.05, another with million points between 0.05 and 0.3, a third between 0.3 and 0.48, and a final group with million points above 0.48. Other partitions resulted in identical patterns.

Individual Model Fitting. To detect whether participant responses were continuous and smooth at the location of 1 million, we initially fit three models to each participant. The first was a simple linear regression (the *true linear* model): the endpoints were allowed to deviate from the extreme left and right, so this model had two free parameters. The data were also fitted by a piecewise linear model with a point discontinuity in its slope at 1 million (*continuous piecewise*), and by a model with a discontinuity in both slope and value (*discontinuous*): in this model, the location of ‘1 million’ depended on whether it was treated as the upper bound of the thousands or the lower bound of the millions. In each case, a normal response model was used for simplicity. Response models were fit by a maximum likelihood method, using the R function *optim* (R Development Core Team, 2008), and compared using a likelihood ratio test.

Results

Figure 2 presents mean participant responses, as well as deviations away from the best fitting piecewise linear model. A clear pattern of slope discontinuity can be observed in the figure, starting in the vicinity of 1 million. Indeed, for 75% of participants, the fully discontinuous model improved the fit of the true linear and piecewise linear models ($\alpha = 0.05$, using a χ^2 likelihood ratio test for nested models); an additional 13% were better fit by the piecewise than the true linear model. These patterns held for the most linear participants: of those in the first quartile, the fully discontinuous model provided the best fit for 73%, while 6% of fits were improved by the piecewise model).

Given that systematic discontinuities in placement occurred around 1 million, it is interesting to explore how participants located the million point. For each participant,

the discontinuous model was used to generate two locations for 1 million: one generated from numbers under one million, and one from numbers over 1 million. The results are shown in Figure 3. As in Landy et al (2013), two clusters of participants can be seen: one group places 1 million far to the left; the other exhibits a broader distribution, but places 1 million roughly 40% of the way across the line. Here, however, we can further see strong systematicity in the discontinuity pattern: non-linear participants systematically leave a large “gap” between the thousand and million scales; very linear participants show slight but meaningful overlap. A simple test applied to the four groups (binned, recall, on the mean 1 million location) finds that all four significantly deviate from point-continuity (Most Linear 95% CI=[-0.025, -0.012]; More Linear CI = [-0.018, -0.002], More Segmented CI = [0.017, 0.078]; Most Linear CI = [0.085, 0.15]); the more linear two quartiles show a significant overlap, while the less linear groups show a significant gap. The data are consistent with the idea that the most linear participants treat the left hand edge of the line as both “1 million” and as “1 thousand”—a simple strategy that would lead to overlapping lines, but also high accuracy.

Discussion

We often speak as though *natural number* is a singular concept, and as though the processing of aligning number names with implicit magnitude and individuation representations—gives us access to the entire structure. Here we have argued that not only does such an alignment come over long developmental stretches, it is never fully completed. Larger numbers whose magnitude can be successfully mapped onto a line are not mapped through a process of systematically integrating into a common linear scale. Instead, it seems that both linear and non-linear responders on the task share a common approach consisting of dividing the scale into culturally given multiplicative regions, and applying linear responses over those subscales. These subscales must then be coordinated with each other to approximate a single line.

Empirically, two novel observations support this interpretation: (1) discontinuities in the derivative of the response, located at 1 million, for all groups of participants, and (2) systematic patterns of location discontinuity, shifting from a positive discontinuity or ‘gap’ for non-linear responders, to a small but significant overlap for highly linear responders. These patterns replicated with both hybrid and numeric stimuli, suggesting that they result from participants’ numerical reasoning and their construal of the task.

Although the multiple overlapping lines account does predict point and slope discontinuities near 1 million, it does not predict the very salient pattern in those discontinuities:

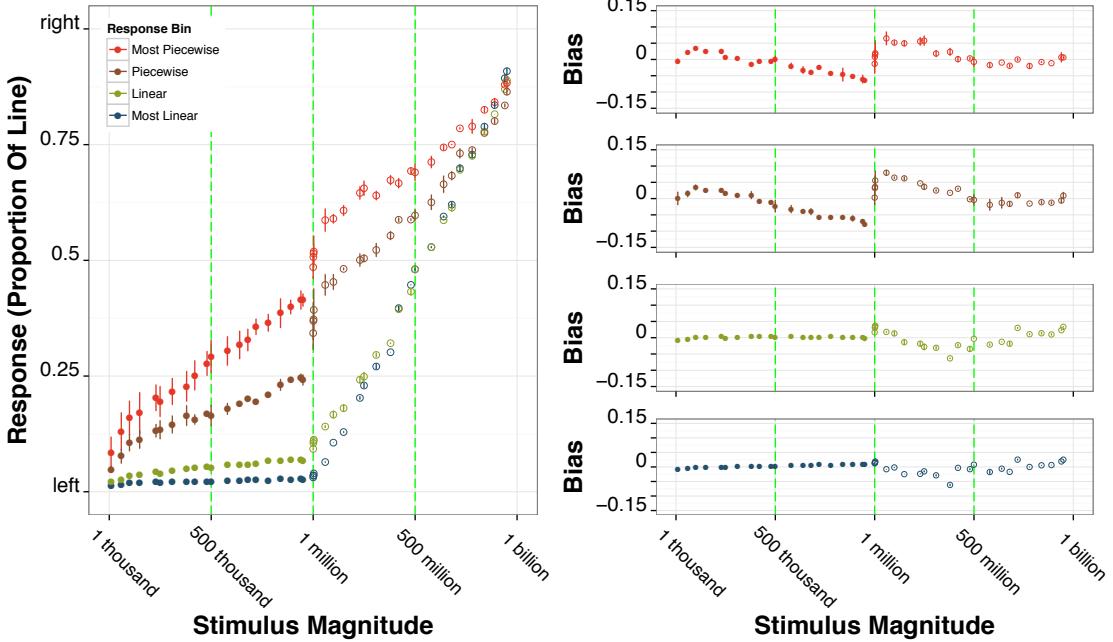


Figure 2: (Left) Mean responses by stimulus condition, binned into groups. Error bars are standard errors around the within-group mean. (Right) Mean residual bias (response-prediction) for the piecewise linear model with a single slope discontinuity at 1 million. For both panels, the x-axis is piecewise linear (see text description).

overlap for the most linear participants, and large gaps—about ten to twenty percent of the total line—for the least linear. Moreover, these results contradict Landy et al., 2013, who found a singular ‘million point’ with a mean of around 0.35-0.4 for non-linear participants. The current modeling approach—which unlike previous approaches allows for a discontinuity at one million—finds two locations for 1 million, one of which is near 0.5 for the segmented groups. The gap between the end of the thousands and the beginning of the millions identified for the segmented groups may be inferred from the fact that the millions range typically starts very close to the midpoint (see Figure 3): participants may integrate a tendency to align the millions scale with a visually salient location (the midpoint) with a realization that the millions cover a ‘larger’ range of numbers than the thousands do—leading to a compressed thousand scale.

It may be tempting to note that the task participants were asked to perform was unreasonable—putting half the marks within a pixel of the left-hand end of the line, and thus to dismiss the observed patterns as ‘task demands’. Such an explanation would overlook the nature of the experimental situation. Participants are always asked to engage in particular, usually unusual behaviors. The ways people grapple with task requirements are informative about the resources available to them (Stenning & Van Lambalgen, 2008). In this case, it appears people can construct “small” linear ranges of around 3 orders of magnitude; beyond that, people make use of culturally available and visually salient reference points. Furthermore, while the pattern of discontinuity was quite similar between very linear and very non-linear responders, the perceived task demands shift considerably; for the non-linear responders, it is not

necessary to “pack” a large number of items near the edge. Finally, if as proposed here number representations in the near large range are constructed through processes of reasoning, it makes sense that they would be task-specific in character. Although on number line estimation, the multiple-overlapping-lines system seems to dominate when numbers have very different magnitudes, it may well be that on other tasks, other approaches are used.

Even for natural numbers just barely beyond the range of common experience, rather than directly extending core conceptual tools, people engage in processes of constructive perception (Landy & Goldstone, 2005; Goldstone & Landy 2010): they coopt existing perceptual analyzers (Carey, 2009) that work well to form linear mappings of smaller number ranges (not accurate numerosity counts), and compose and iterate them to create new number ranges, much in the same manner as external notation systems such as power towers or Knuth up-arrow notation do. We have found that 1 million is a location for a discontinuity (of course, it may not be the only or even the smallest such boundary)—it might have been the case that familiarity or psychophysical factors created a boundary in strategy at any arbitrary number. The observed pattern suggests that people use the culturally provided numeral system to select appropriate magnitudes at which to begin recycling cognitive resources.

Telescopes provide an apt metaphor for these cognitive tools: they extend the natural bounds of perception by connecting them to new contents while also distorting those contents. For example, understanding the magnitude of 2 billion might be less like perceiving its quantity than like believing a system of facts that involve magnitude systems.

The natural numbers have amazing properties that derive entirely from the successorship function. It appears, however, that human representations of natural numbers, at least beyond a paltry few hundred thousand iterations, rely on resources quite distinct from successorship or even a metric “number line”. A fundamental mistake made by classical empiricism was to assume that the inner representations were iconic—that they were *like* the outer represented. When reasoning about large numbers, we appear to rely on representations that are fundamentally unlike the numbers themselves.

Acknowledgments

This research was partially funded by Department of Education, Institute of Education Sciences grant R305A110060, as well as an undergraduate research grant from the University of Richmond. Zach Davis, Megan DeLaunay, and Brian Guay made valuable suggestions.

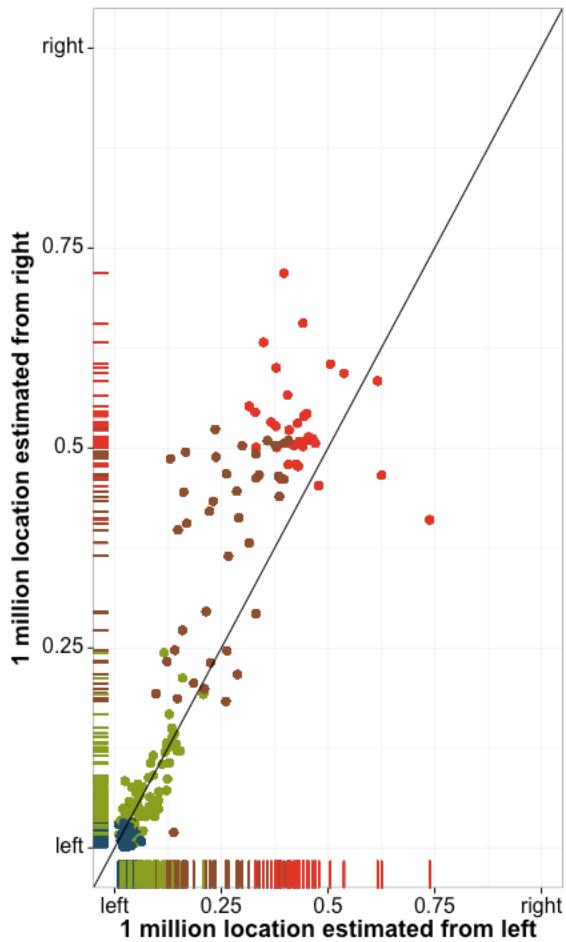


Figure 3: Estimated locations of 1 million in the discontinuous model. The line indicates continuous behavior; points above the line indicate gaplike behavior, while points below the line indicate an overlap in the best-fitting lines.

References

Anderson, M. L (2010). Neural reuse: A fundamental organizational principle of the brain. *Behavioral and Brain Sciences*, 33(4), 245-313.

Barth, H., & Paladino, A.M. (2011). The development of numerical estimation: evidence against a representational shift. *Developmental Science*, 14, 125-135.

Carey, S. (2009). The origin of concepts. New York: Oxford University Press.

Crump MJC, McDonnell JV, Gureckis TM (2013) Evaluating Amazon's Mechanical Turk as a Tool for Experimental Behavioral Research. *PLoS ONE* 8(3): e57410. doi:10.1371/journal.pone.0057410

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Science*, 8(7), 307-314.

Gelman, R. (2011) The case of continuity. *Behavioral and Brain Sciences*, 34(3), 127-128.

Goldstone, R. L. & Landy, D. (2010). Domain creating constraints. *Cognitive Science*, 34(7), 1357-1377.

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

Landy, D. & Goldstone, R. L. (2005). How we learn about things we don't already understand. *Journal of Experimental and Theoretical Artificial Intelligence*, 17, 343-369.

Landy, D., Silbert, N. & Goldin, A. (2013). Estimating large numbers. *Cognitive Science*, 37(5), 775-799. doi: 10.1111/cogs.1202.

Leslie, A. M., Gelman, R., & Gallistel, C. R. (2008). The generative basis of natural number concepts. *Trends in cognitive sciences*, 12(6), 213-218.

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

Opfer, J., Siegler, R., & Young, C. (2011). The powers of noise-fitting: reply to Barth and Paladino. *Developmental Science*, 14, 1194-1204.

R Development Core Team (2008). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.

Siegler, R. S., & Opfer, J. E. (2003). The Development of Numerical Estimation Evidence for Multiple Representations of Numerical Quantity. *Psychological Science*, 14(3), 237-250.

Stenning, K., & Van Lambalgen, M. (2008). *Human reasoning and cognitive science*. MIT Press.

Sullivan, J., & Barner, D. (2013). How are number words mapped to approximate magnitudes?. *The Quarterly Journal of Experimental Psychology*, 66(2), 389-402.

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