Title

Beyond Magnitude:
How Math Expertise Guides Number Representation

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 37(0)

Authors

Murphy, April D
Rogers, Timothy T
Hubbard, Edward Moscoso del Prado
et al.

Publication Date

2015

Peer reviewed

Beyond Magnitude: How Math Expertise Guides Number Representation

April D. Murphy[1], Timothy T. Rogers[1], Edward M. Hubbard[2], Autumn Brower[2]
(admurphy2@wisc.edu, ttrogers@wisc.edu, emhubbard@wisc.edu, ajbrower@gmail.com)
[1] Department of Psychology, 1202 W. Johnson Street Madison, WI 53706 USA
[2] Department of Educational Psychology, 1025 W. Johnson Street Madison, WI 53706 USA

Abstract

Previous studies on numeric cognition have focused primarily on magnitude, based on its role as a core feature of number knowledge. In this paper, we report the results of three experiments investigating adults' sensitivity to properties of number apart from magnitude. In Experiment 1, we use a triadic judgment task to replicate a classic study of number properties. In Experiment 2, we compare these representations among expert and non-expert groups. In Experiment 3, we examine whether instruction can tune representation of number properties. Results indicate that the triadic comparison task is a reliable method of assessing sensitivity to number properties. We found that magnitude is difficult to suppress among non-experts, who are primarily attuned to magnitude and parity. Mathematically sophisticated participants were sensitive to a range of number properties compared with the non-expert group. We discuss implications for theories of number concepts and their relation to special populations.

Keywords: number representation, magnitude, individual differences

Introduction

Psychological approaches to number knowledge often focus on the representation of magnitude, under the assumption that the semantic core of number knowledge is quantity (Booth et al., 1999, Booth & Seigler 2008, Dehaene et al. 1993, Fias et al. 1996). Numbers have many properties beyond quantity, however. The number 3, for example, can be prime, odd, or a factor of 27; it can call to mind a set of triplets, a bronze medal for third place, or a triangle. Though research on the representation of quantity has advanced rapidly in recent years, less is known about whether and to what extent aspects of number knowledge beyond magnitude shape number concepts.

A classic study by Shepard, Kilpatrick, and Cunningham (1975) probed adult representation of number properties using multidimensional scaling. Adults were asked to make pairwise similarity judgments among the single-digit numbers 0-9 presented in various forms (e.g., digits, number words, dot arrays). The results of a multidimensional scaling on these data indicated that, in addition to magnitude, features such as parity and powers of two and three were used to make the judgments. The implications of these findings for theories of number concepts remains unclear, however. Observations were only collected from 4 participants, all colleagues or students of Shepard at Stanford who likely had substantial expertise in mathematics. Do non-magnitude properties of numbers shape number concepts broadly for people in our society, or does sensitivity to such properties only arise as a consequence of specialized learning and expertise?

A few studies have approached this question in cognitive development. Miller & Gelman (1983) used a explored children's and adults' sensitivity to the properties identified in Shepard et al. (1975), using a triadic judgment task whereby subjects rated the most similar of 3 numbers printed on either cardboard wheels (for children) or index cards (for adults). In this study they found sensitivity to both magnitude and parity relations among adults and 6th gade children, while kindergarteners and 3rd grade children were only sensitive to magnitude. Similarly, Berch et al. (1999) found that children from 4th grade onward were reliably sensitive to parity in addition to magnitude.

To our knowledge, however, the original results identified by Shepard et al. (1975) have not been replicated in a sample of adults who are both representative of a college-educated population and naïve to the experimental goals. Yet the questions about the structure of number concepts are important for understanding the origins and nature of numeracy in human cognition. The literature's current focus on number magnitude has been useful in connecting research in animal cognition, human development, and neural bases of number knowledge (e.g., Feigenson et al. 2004, Verguts & Fias 2004, Dehaene & Changeaux 1993, Wynn 1992, Xu & Spelke 2000, Kadosh et al. 2008), and one implication of this work is that numeracy may be grounded in an innate sense of quantity that is conserved across both phylogeny and ontogeny (Brannon & Terrace 2002). It is less easy to see how other number properties, such as parity or primeness, might connect to or emerge from sensitivity to magnitude as observed in non-human animals, young infants, and neural signals. If such properties broadly shape the relationships we discern amongst numbers, this suggests that there are important unanswered questions about the nature and origins of numerical concepts.

The current study uses contemporary multidimensional scaling (MDS) methods to measure the similarities people discern amongst single-digit numbers and assess whether they reflect the properties identified by Shepard et al. (1975). Whereas these authors used overt similarity ratings and classical MDS to generate embeddings, we employ a triadic matching task in which participants must decide which of two items is more similar to a third reference item, and estimate embeddings with non-metric MDS. In

Experiment 1 we assess whether this approach can replicate the original findings, focusing on judgments of colleagues with expert mathematical knowledge. We then compare the structures uncovered by these methods in groups of university undergraduates and in another special population with expert mathematical knowledge, graduate students in computer science and mathematics. Finally, we assess whether the structures revealed in these studies change or remain the same when participants are explicitly instructed to ignore number magnitude when making their decisions. The results allow us to assess whether magnitude is the nucleus of numeric representation in university-educated adults broadly speaking, and whether the saliency of magnitude can be shifted either by expertise or through explicit task instruction.

Experiment 1

The aim of Experiment 1 was to assess whether we could replicate the results of Shepard et al. (1975) using triadic comparisons and non-metric MDS to estimate the similarity structure among single-digit numbers. To this end we imitated Shepard et al.'s strategy of studying number concepts in colleagues with extensive mathematical knowledge. Sixteen participants, all graduate students, research assistants, or faculty at the University of Wisconsin-Madison, participated in the task.

Triadic comparison task

The experiment was conducted on a computer using a Web-based paradigm that allows participants to visit a URL and complete an experiment from a Web browser. After logging in, subjects read an introduction to the experiment which pointed out that numbers can have many properties such as even or odd, large or small, prime, multiples of 3 and so on. Participants were then instructed that, on each trial of the study, they must decide which of two numbers is most similar to a third, taking into account everything they know about the three numbers. The experiment then randomly selected three single-digit numbers (0-9) without replacement and presented them on a computer screen. The reference number was presented at the top of the screen, and the two other numbers were presented below on the right or left side of the screen. Participants made judgments by pressing the left or right arrow key that corresponded to their choice, after which the next triad would automatically appear. Participants were told to complete as many judgments as they could in 10 minutes. At the end of the time limit, the experiment automatically terminated and a debriefing message was displayed.

To investigate the structure underlying participant judgments, we used a form of non-metric multidimensional scaling (non-metric MDS) to generate several low-dimensional representations of the response data. This was accomplished by using the participants' responses to situate the 10 target numbers in a low-dimensional space, which we will refer to as an embedding. In this space, the distance among the targets directly corresponds to their similarity. The non-metric MDS embedding was computed using stochastic gradient descent on a hinge-loss objective function [see NEXT website for implementation details: https://next.discovery.wisc.edu]. The computations are performed across non-aggregated response data in a different random order each time until the embedding reaches a steady state whereby additional iterations have minimal influence on the positions of the targets and the overall error of the embedding. Reliability is evaluated by testing the model results on a hold-out portion of the data that was not used to generate the original embedding.

Again following Shepard et al. (1975), we simply inspected the resulting embedding for evidence of sensitivity to the properties explicitly identified in their study—magnitude, parity, powers of two, powers of three, and the special status of zero—as well as an additional important property, primeness.

Figure 1 shows the resulting embedding, which closely replicates the original findings. The embedding clearly expresses dimensions that capture number magnitude and parity, and there exist linear planes in the 2D space that separate powers of 2 from other numbers, zero from other numbers, and primes from non-primes. The only non-magnitude property reported in the original study that is not clearly reflected in this scaling is powers of 3. The study thus validates the triadic comparison method as capable of revealing non-magnitude properties in number concepts of expert participants—but as in the original study it remains unclear whether these results arise from expertise specifically or reflect aspects of number concepts in a broader population.

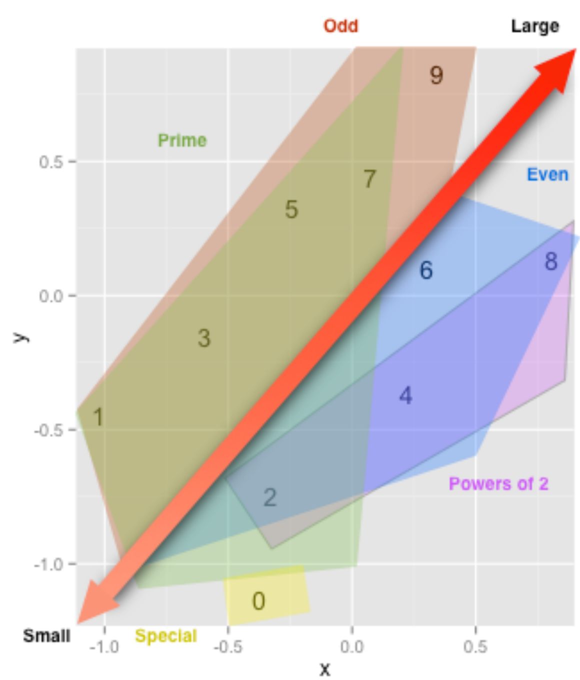

Figure 1
2D solution of Shepard et al. (1975) replication

Experiment 2

The goal of experiment 2 was to compare the structure of number concepts in a sample of university undergraduates not necessarily possessing expert math knowledge to that elicited from a new sample of math experts who were naïve to the experimental goals, all graduate students in math and computer science. The central question was whether the same properties of numbers would be equally well represented in the two groups. We adopted two different approaches to testing this question quantitatively.

Participants and Task

The experiment was conducted in a computer lab in the Psychology department at the University of Wisconsin-Madison. Participants consisted of 23 undergraduate students and 9 Computer Science and Mathematics graduate students at the University of Wisconsin-Madison. Participants were recruited from the Psychology department volunteer subject pool, as well as email solicitations to the Computer Science and Mathematics departments.

Participants completed the same task with the same instructions as in Experiment 1. Each participant made 100 similarity judgments, which took between 15-20 minutes Upon reaching the trial limit, the experiment automatically terminated and a debriefing message was displayed.

Analysis

To compare the richness of the structures underlying exper and novice judgments, we computed separate embeddings i 1-4 dimensions for each group using the same method a Experiment 1. This allowed us to assess whether th underlying dimensionality of representations governin number judgments is similar or different in novices versu experts. If experts possess richer knowledge about numbe attributes, their performance should be best fit by a riche (higher-dimensional) underlying representation.

Figure 2
Error curves for non-metric MDS embeddings in Experiment 2

Figure 2 indicates the relationship between training and testing error across the 1D-4D solutions. The relatively flat error curve for the non-expert group suggests that these data may be best fit to a 1D solution; that is, a single numeric property may be the only reliable dimension to which this group is sensitive. For the expert group, the error is lowest in the 4D solution, indicating that these participants may be sensitive to a broader range of numeric properties.

Second, for each group we used logistic regression as a linear classification model to assess whether particular number properties are present in the 2D embeddings estimated for each group. Because our pilot study, as well as prior research (e.g. Dehaene et al., 1993), indicated that 0 is a special class of number for which discrimination of non-magnitude properties may be difficult, we restricted our classification model to numbers 1-9. The classifier was trained to discriminate numbers possessing or not possessing a numeric property as a function of the coordinate vectors on each dimension of the 2D scaling. We used leave-one-out cross-validation to evaluate the predictive accuracy of the model, doing this for all 9 digits; thus, in each iteration, the model was trained on 8 digits and tested on the 9th. For instance, to assess whether an embedding contains reliable information about parity, the classifier was trained to discriminate even from odd numbers for the digits 1-8, and the resulting model was used to classify the digit 9.

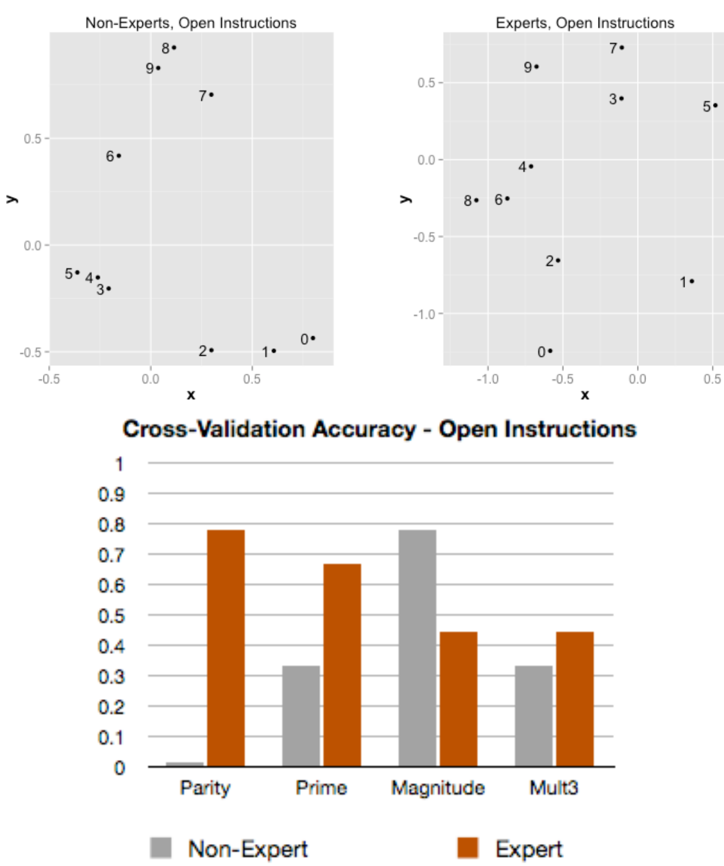

Figure 3
2D solution of Experiment 2 data, and cross-validation model accuracy across categories and expertise groups

This process was repeated, omitting each digit from the training set once, and the total accuracy of the classifier across all 9 runs was used as a measure of the degree to which parity is expressed in the embedding. Using this procedure, we assessed four number properties, which split the 9 digits into two roughly equal-sized classes: magnitude (large vs. small), parity, primeness, and multiples of three.

Figure 3 illustrates the 2D solution from non-metric MDS and the cross-validation classification accuracy for each of the four categories across the expert and non-expert groups.

Results

Our objective in this study was to evaluate whether the four properties of numbers we identified in the pilot study would be well represented across both expert and non-expert groups. Our quantitative approach made use of cross-validation procedures with both the non-metric MDS and logistic regression classifier to support model accuracy and avoid overfitting the data.

In the non-expert group, our leave-one-out cross-validation indicated that the most reliably classified feature was magnitude, with a classification accuracy of 77%, as visual inspection of the 2D embedding suggests. Other features were less well classified, with parity surprisingly being classified with 0% accuracy. Prime and multiples of 3 were both classified with 33% accuracy, indicating that few participants consistently used these dimensions to determine conceptual similarity among the numbers.

In the expert group, classification was most reliable for parity, with an accuracy of 77%. Primeness was classified with an accuracy of 66%, and it is notable that magnitude was not as reliable a predictor with classification of 44%. Visual inspection of the 2D embedding suggests it is plausible that the experts may have also used powers of two in addition to or instead of parity to make similarity judgments. Classification accuracy for the multiples of three category was only slightly better in the expert group, at 44%.

These results bring to light two interesting findings. First, the properties of numbers which were strongly identified in Shepard et al. (1975) and in Experiment 1 were not consistent among the experiment-naïve mathematics experts. Second, only the dimension of magnitude was salient among the non-expert group, while parity was most reliable in the expert group—this supports prior research on aspects of number knowledge (Miller & Gelman 1983, Dehaene et al., 1993). However, there was only weak representation of other properties across both expert and non-expert groups.

This raises the question of whether the saliency of magnitude, particularly among the non-expert group, suppressed other aspects of number besides parity. Additionally, we hypothesized that there may high variability in the experts' sensitivity to various properties of numbers, which led them to develop different strategies for judging similarity.

To address these questions, we conducted a third experiment in which we explicitly instructed participants to ignore magnitude when making similarity assessments.

Experiment 3

Experiment 2 indicated that judgments among the four number categories were represented unequally across the undergraduate sample and the expert sample. We hypothesized that because magnitude is so central to number knowledge, this dimension may need to be explicitly suppressed to reveal alternate aspects of number. In this experiment, we gave participants instructions to avoid using magnitude relations when making similarity judgments of the stimuli. Our objective was to tune the strength of non-magnitude number representations and evaluate the consistency of these across expertise groups.

Participants and Task

The experiment was conducted in a computer lab in the Psychology department at the University of Wisconsin-Madison. Participants consisted of a second group of 23 undergraduate students and 8 Computer Science and Mathematics graduate students at the University of Wisconsin-Madison, who had not previously participated in the experiment. Participants were recruited from the Psychology department volunteer subject pool, as well as email solicitations to the Computer Science and Mathematics departments.

Participants completed the same computer-based task as in Experiment 1. The experiment instructions were modified to indicate that participants were to not think about magnitude when making judgments. Each participant made 100 similarity judgments, which took between 15-20 minutes. Upon reaching the trial limit, the experiment automatically terminated and a debriefing message was displayed.

Analysis

As with Experiment 2, we used non-metric MDS to compute separate embeddings in 1-4 dimensions for each group using the same method as Experiments 1 and 2. Figure 4 indicates the relationship between training and testing error across the dimensions. In contrast to Experiment 2, the error curve for the non-expert group flattened out around 3 dimensions, suggesting that ignoring magnitude may successfully persuade participants to attend to alternative features. For the expert group, the error is again lowest in the 4D solution.

Figure 4
Error curves for non-metric MDS embeddings in Experiment 3

We used logistic regression as a classifier to evaluate whether the four number categories identified in Experiment 2—prime, parity, magnitude, and multiples of 3—would be present in the 2D embeddings estimated for each group in the ignore-magnitude condition.

Figure 5 illustrates the 2D solution from non-metric MDS and cross-validation classification accuracy for each of the four categories across the expert and non-expert groups.

Results

Our aim in Experiment 3 was to constrain the influence of magnitude on participants' judgments of number similarity in the triadic comparison task. By directing participants to avoid using magnitude, and instead asking that they exercise their knowledge of other types of number relations, we expected that the saliency of non-magnitude properties of numbers would be revealed.

Results for the non-expert group were surprising. The leave-one-out cross-validation indicated that, rather than suppressing magnitude's importance, it remained reliably classifiable with an accuracy of 88%. Importantly for our hypothesis however, activating non-magnitude knowledge through instruction allowed parity to be perfectly classified at 100%; this is particularly evident in the 2D visualization.

Our hypothesis was also well supported within the expert group, with all three non-magnitude properties classified reliably better than in the open instructions condition. The multiples of three category was perfectly separable with 100% accuracy, while parity representation was also well classified with an accuracy of 88%. Primes and magnitude were classified at 66% accuracy, changing little from the open instructions condition.

The results of Experiment 3 provide compelling evidence of the role of both instruction and expertise in guiding representation of non-magnitude properties of numbers. We note that, while further representational complexity may be sacrificed in the expert group by compressing information from higher dimensions into a 2D embedding, the results from this potentially more coarse approach remain encouraging with regard to our hypotheses.

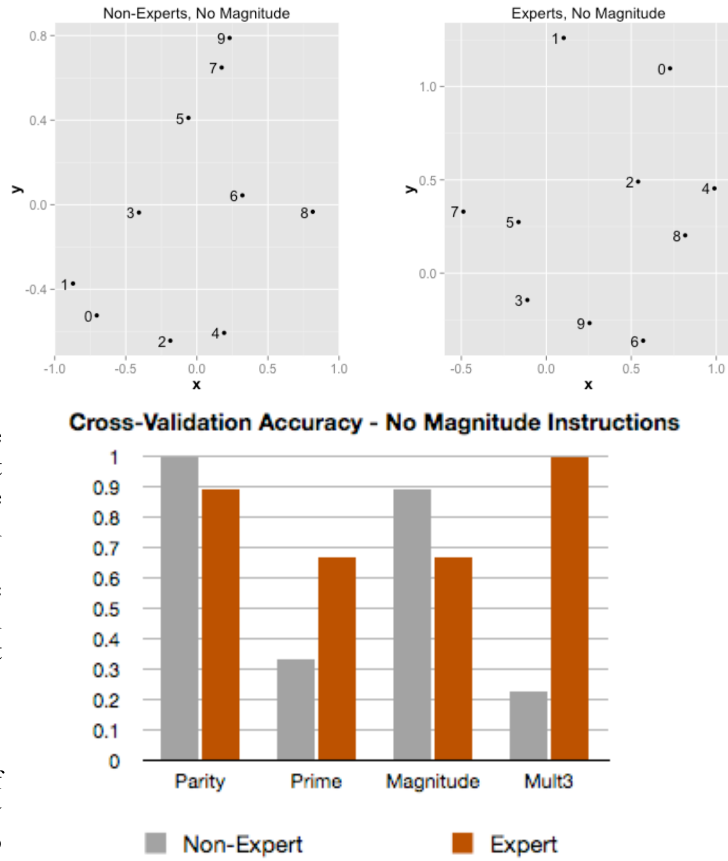

Figure 5
2D solution of Experiment 3 data, and cross-validation model accuracy across categories and expertise groups

General Discussion

These findings provide important implications for the study of number representation and expertise. While Shepard et al. (1975) claimed that adults are broadly sensitive to several non-magnitude properties of number, we found this claim to be supported only among highly trained experts. Furthermore, when non-metric multidimensional scaling and logistic regression are used to evaluate the inherent dimensionality and classification accuracy of similarity judgments, we can address questions of number knowledge more quantitatively compared with visual inspection of relationships in a 2D scaling.

We found that task instructions played a role in the malleability of number knowledge for both experts and non-experts. With open instructions, non-experts were most reliably sensitive to magnitude, supporting prior research in these domains (Miller & Gelman 1983, Dehaene et al. 1993), while experts were most reliably sensitive to parity. When asked to suppress knowledge of magnitude relations, non-experts had difficulty doing so, even while also accurately discriminating parity. In contrast, the ignore-magnitude instructions allowed experts to improve reliability at uncovering the conceptual structure of all three non-magnitude properties.

Although we assumed that features related to primeness, parity, or multiplication should be commonly known and easily accessed by both expertise groups, the college-level mathematics skills possessed by the non-expert group did not necessarily predict an ability to make similarity judgments based on these features. Additionally, while domain expertise may permit greater flexibility in shifting representation away from magnitude, even then magnitude must be explicitly suppressed to allow other features to reliably surface.

Non-magnitude properties of number represent highly abstract conceptual knowledge, and these studies address classical findings while taking steps towards investigating a relatively under-studied domain. While our results indicate that these aspects of number knowledge can be highly variable depending on individual expertise and task demands, future research is needed to fully explore the consequences of these findings for number cognition more broadly.

References

Armstrong, S. L., Gleitman, L. R., & Gleitman, H. (1983). What some concepts might not be. *Cognition*, *13*(3), 263-308.

Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition*, *44*(1), 75-106.

Berch, D. B., Foley, E. J., Hill, R. J., & Ryan, P. M. (1999). Extracting parity and magnitude from Arabic numerals: Developmental changes in number processing and mental representation. *Journal of experimental child psychology*, *74*(4), 286-308.

Booth, J. L., & Siegler, R. S. (2008). Numerical magnitude representations influence arithmetic learning. *Child Development*, *79*(4), 1016–1031.

Brannon, E. M., & Terrace, H. S. (2002). The evolution and ontogeny of ordinal numerical ability. *The cognitive animal: Empirical and theoretical perspectives on animal cognition*, 197-204.

Starkey, P., & Cooper, R. G. (1980). Perception of numbers by human infants. *Science*, *210*(4473), 1033-1035.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*,*122*(3), 371.

Dehaene, S., & Changeux, J. P. (1993). Development of elementary numerical abilities: A neuronal model. *Cognitive Neuroscience, Journal of*, *5*(4), 390-407.

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

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

Fias, W. (1996). The importance of magnitude information in numerical processing: Evidence from the SNARC effect. *Mathematical cognition*, *2*(1), 95-110.

Gallistel, C. R., & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in cognitive sciences*, *4*(2), 59-65.

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, *44*(1), 43-74.

Groen, G. J., & Parkman, J. M. (1972). A chronometric analysis of simple addition. *Psychological review*, *79*(4), 329.

Kadosh, R. C., Muggleton, N., Silvanto, J., & Walsh, V. (2009). Double dissociation of format-dependent and number-specific neurons in human parietal cortex. *Cerebral Cortex*, bhp273.

Kadosh, R. C., Lammertyn, J., & Izard, V. (2008). Are numbers special? An overview of chronometric, neuroimaging, developmental and comparative studies of magnitude representation. *Progress in neurobiology*, *84*(2), 132-147.

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.

Shepard, R. N., Kilpatric, D. W., & Cunningham, J. P. (1975). The internal representation of numbers. *Cognitive Psychology*, *7*(1), 82-138.

Verguts, T., & Fias, W. (2004). Representation of number in animals and humans: a neural model. *Journal of Cognitive Neuroscience*, *16*(9), 1493-1504.

Wynn, K. (1992). Addition and subtraction by human infants.

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*(1), B1-B11.