Number Representations and their Development: A Connectionist Model of Number Comparison

Mark Rose Lewis (lewis505@umn.edu)
Department of Educational Psychology, 56 East River Road
Minneapolis, MN 55455

Sashank Varma (sashank@umn.edu)
Department of Educational Psychology, 56 East River Road
Minneapolis, MN 55455

Abstract

Building on prior work, the current study evaluated whether connectionist models can account for the distance and size effects in adults and the development of the distance effect in children. A family of models was constructed by orthogonally varying training environment (naturalistic versus non-naturalistic) and number representation (one-to-one versus magnitude). The ability of the models to account for the adult distance and size effects depended critically on a naturalistic training environment but was relatively independent of number representation. With respect to the developmental data, the naturalistic/one-to-one model provided a good account of response times and errors. The relation between the current models and prior models and avenues for future exploration are discussed.

Keywords: number comparison; distance effect; size effect; connectionism; models; development

Introduction

The nature of number representations is an enduring question in cognitive science. One clue to this representation is the *distance effect*: the time it takes to judge the greater (or lesser) of two numbers decreases with the distance between the numbers (Moyer & Landauer, 1967). For example, 1 vs. 9 is judged faster than 1 vs. 3. Another clue is the *size effect*: the time to judge the greater (or lesser) of two numbers that are a fixed distance apart increases with the absolute magnitude of the numbers (Parkman, 1971). For example, 7 vs. 9 is judged more slowly than 1 vs. 3. The distance and size effects conform to psychophysical laws (i.e., $RT = Klog(\frac{larger}{larger - smaller})$) and are therefore commonly interpreted as evidence that numbers are represented as analog representations, perhaps localized to the intra-parietal sulcus (Dehaene, Piazza, Pinel, & Cohen, 2003). Researchers have proposed various implementations of these analog representations. The classic ones are as points on a compressed mental number line (Dehaene & Mehler, 1992; Rule, 1969) and as points on a linear mental number line associated with increasing variability (e.g., Gallistel & Gelman, 2000). More recently, two connectionist models of number representation have appeared. Zorzi and Butterworth (1999) assumed magnitude representations whereby numbers are represented by banks of overlapping units. This model was able to account for the adult distance effect. By contrast, Verguts, Fias, and Steven (2005) assumed a coarse-coded representation, with each number corresponding primarily to one unit, but with graded activation of adjacent units. This model was able to account for the adult distance and size effects.

The purpose of the current study was to evaluate the ability of connectionist models to (1) account for the adult distance and size effects as a function of training environment and number representation and to (2) account for the development of the distance effect. In these regards, the reported simulations are the first of their kind.

With respect to training environment, some connectionist models (Zorzi & Butterworth, 1999) have employed a *non-naturalistic* training environment (i.e., every one-digit number appears with equal likelihood). However, corpus studies indicate that the frequency of a number falls off as a power function of its magnitude (Dehaene & Mehler, 1992), implying that one-digit numbers are non-uniformly distributed in a *naturalistic* environment. Some connectionist models have employed a naturalistic training environment (Verguts et al., 2005). We sampled comparisons (i.e., pairs of one-digit numbers) from these contrasting training environments to evaluate whether the distance and size effects were contingent upon naturalistic input.

With respect to number representation, we considered the magnitude representation implemented by the Zorzi and Butterworth (1999) model and a one-to-one variant of the coarse-coded representation implemented by the Verguts et al. (2005) model[1].

Finally, in the first study to model the development of the distance effect, we evaluated whether improvements in model performance throughout training parallel improvements in children's response times and error rates throughout development.

Method

We developed four connectionist models by orthogonally varying training environment (naturalistic versus non-naturalistic) and number representation (magnitude versus one-to-one). The models were implemented within a common connectionist architecture patterned after Verguts et al. (2005).

[1] Both of these codings represent exact numbers. We use the label "magnitude" to reflect the fact that the number of representation nodes activated in this coding corresponds to the number being compared.

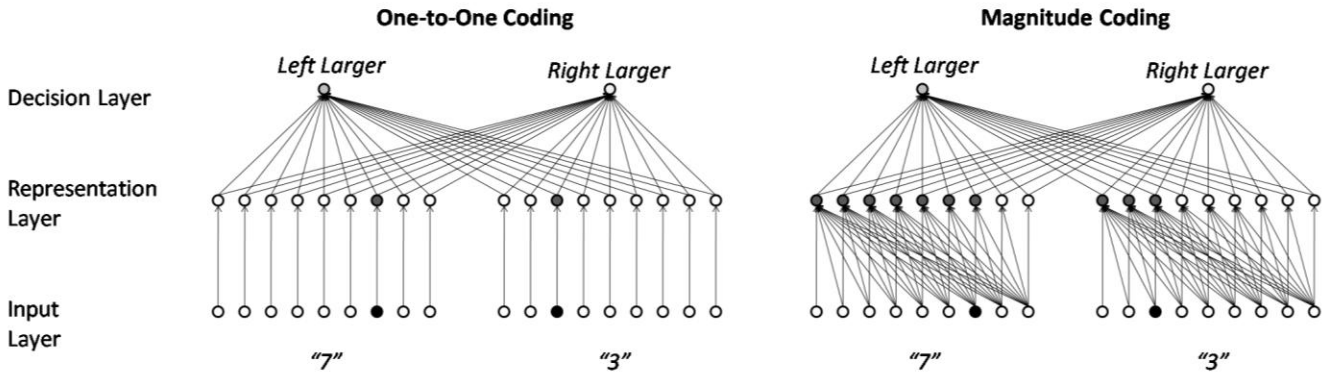

Figure 1: Schematic of models using one-to-one coding (left) and magnitude coding (right).

This architecture consisted of three layers of units (*input, representation,* and *decision* layers) (Figure 1). Each layer contained left and right fields. On each trial, the input units corresponding to the numbers being compared were clamped to an activation level of 1, and activation spread forward throughout the network. When a decision unit (*left larger* or *right larger*) reached an activation of 0.5 or greater, the model was considered to have made a decision

Architecture and Number Representation

Each model consisted of three layers of units. The input layer consisted of two fields of nine units each that corresponded to the numbers 1-9. The left field corresponded to the number presented on the left and the right field to the number on the right. Each number corresponded to one (and only one) unit in the input layer.

The representation layer consisted of two sets of nine units. The left field, *M*, represented the number presented on the left, and the right field, *N*, represented the number presented on the right. The left input field was connected to the left representation field and the right input field to the right representation field by connections with weights 0 or 1. The number representation scheme of the model determined the pattern of connection weights between the input and representation layers. For magnitude representations, the number of units activated in a representation field corresponded to the magnitude of the number presented (*e.g.,* if the number 5 was presented on the left, the 5 leftmost units of the left representation field would be activated). For one-to-one representations[2], one (and only one) unit in a representation corresponded to the number presented. The weights of the connections between the input and representation layers were held constant throughout learning to maintain the type of representation that the model *a priori* employed.

The decision layer consisted of two units representing *left larger* and *right larger* decisions. Units in the representation layer were fully connected with units in the decision layer. The initial weights of these connections were randomly sampled from a uniform distribution (0 to 1) and were adjusted during training by a supervised learning rule.

Model Dynamics

On each trial, the model compared two numbers, judging which was greater. (Following prior work, we did not model both greater and lesser judgments.) The left number was presented to the left input field by clamping the activation of the corresponding unit to 1, and the right number was presented similarly to the right input field. Activation spread from the input layer to the representation layer according to the equation[3]:

$$(1)\ \Delta r_{Mk}(t) = r_{Mi}(t-1) + \sum_{i=1}^{i=9} w_{Mkin_i}\lceil in_i(t) - \theta\rceil^{+}$$

Where $\Delta r_{Mk}(t)$ is the change in the activation of the k^{th} representation unit in the left field (*M*), $in_i(t)$ is the activation of the i^{th} input unit, w_{Mkin_i} is the weight of the connection between these two units, and θ is a firing threshold (set to .08 for these simulations). This equation results in the activation of representation units asymptotically approaching their maximum values.

Activation spread from the representation layer to the left-larger unit of the output layer according to the equation:

$$(2)\ \Delta o_{Left}(t) = o_{Left}(t-1) + \sum_{i=1}^{i=9} w_{Mi,Left}\lceil r_{Mi}(t) - \theta\rceil^{+} + \sum_{i=1}^{i=9} w_{Ni,Left}\lceil r_{Ni}(t) - \theta\rceil^{+}$$

Where $\Delta o_{Left}(t)$ is the change in the activation of the left-larger unit, $r_{Mi}(t)$ is the activation of i^{th} representation unit

[2] We employed one-to-one representations instead of coarse-coded representations (Verguts et al., 2005) to equate the architecture across models. Coarse-coding would have required adding additional units to the representation layer of models that employed magnitude representations, muddying the comparison of the models.

[3] All equations are for left fields. Equivalent equations governed model dynamics in the right fields.

in the left field, $w_{Mi,Left}$ is the weight of the connection between these two units, $r_{Ni}(t)$ is the activation of the i^{th} representation unit in the right field, and θ is the firing threshold. This equation results in the activation of decision units asymptotically approaching their maximum values once the representation units have reached the firing threshold. A decision was considered made once activation in one of the decision units exceeds a threshold of 0.5.

Supervised Learning

During learning, weights between representation and decision units were adjusted according to the delta rule:

$$(3)\ \ \Delta w_{ik} = \epsilon(t_k - o_k)r_i$$

Where Δw_{ik} is the change in the weight between the i^{th} representation unit and the k^{th} decision unit, ϵ is a learning rate parameter, $(t_k - o_k)$ is the difference between the target decision unit activation t_k (1 for larger, 0 for smaller) and the actual decision unit activation o_k, and r_i is the activation of the i^{th} representation unit. The delta rule apportions blame for incorrect decisions and adjusts weights accordingly. For this study, the learning rate parameter ϵ was set to 0.02. During learning, activation was allowed to settle prior to weight adjustment. Each model was trained for 30,000 trials, and weights were adjusted at the end of every trial.

Training Environment

Models were trained on one of two training environments. Naturalistic training environments were constructed by assuming, following Dehaene and Mehler (1992), that the frequency of a number in the environment is a decreasing function of its magnitude. Although Dehaene and Mehler favored a power function, Verguts et al. (2005) adopted a closely related exponential function. To facilitate the comparison of our results, we formed training comparisons by sampling pairs of numbers from an exponentially decreasing distribution (where the frequency of number i is $e^{-0.2i}$). The distribution of individual numbers and of comparisons (as a function of distance) is shown in Figure 2.

Non-naturalistic training environments were constructed by assuming that numbers are distributed uniformly in the environment. Training comparisons were formed by sampling from this distribution. The results are also shown in Figure 2.

It is interesting that naturalistic and non-naturalistic training environments result in strikingly similar distributions of comparisons as a function of distance. However, as we shall see, these environments have important differences as indicated by the ability of the resulting models to account for the adult distance and size effects.

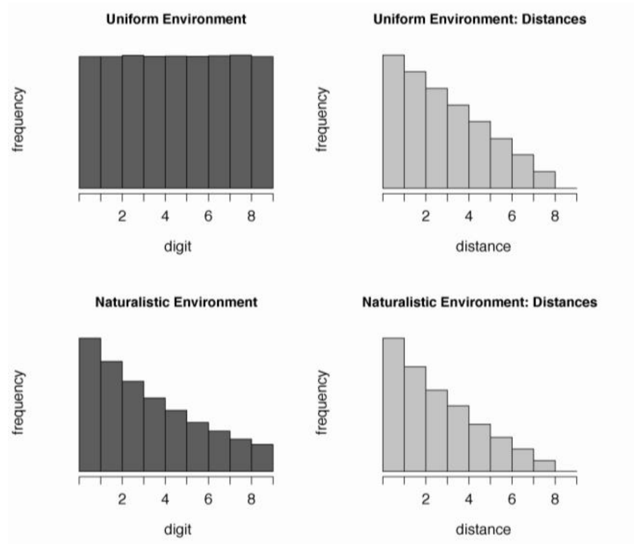

Figure 2: Training environments. Dark gray histograms give the distribution of numbers in the environment, light gray histograms the distribution of distances between the resulting comparisons (i.e., number pairs).

Training and Testing

Ten copies of each model (naturalistic versus non-naturalistic crossed with magnitude versus one-to-one) were created. Each copy was trained for 30,000 trials (following Verguts et al.) and tested with all possible pairs of numbers between 1 and 9 (excluding ties).

Results

Distance Effects

All four models produced distance effects (Figure 3).

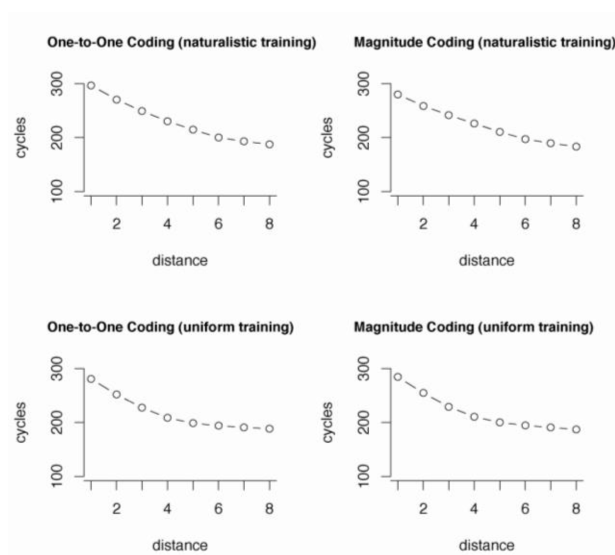

Figure 3: Distance effects for the four models.

Table 1: Model fits for the distance effect.

Representation	Training	R^2	p
Magnitude	Naturalistic	.73	< .001
Magnitude	Uniform	.43	< .001
One-to-One	Naturalistic	.78	< .001
One-to-One	Uniform	.46	< .001

To evaluate the fit of each model to human performance, we followed Zorzi and Butterworth (1999) in regressing human performance as captured by the equation:

$$\widehat{RT} = K\log\left(\frac{larger}{larger-smaller}\right) + C$$

against the number of cycles to make a decision. The results are shown in Table 1.

First, consider the question of training environment. The results indicate that models trained in naturalistic training environments provide better accounts of the distance effect than models trained in non-naturalistic environments. Although Figure 2 suggests that the difference between the training environments is negligible with respect to the amount of experience with different distances, the fit statistics indicate that differences between uniform and naturalistic environments are critical to the distance effect .

Next, consider the question of number representation. The results indicate that a model's ability to account for the distance effect is independent of whether it uses magnitude or one-to-one number representations. Additional work is necessary to determine how fundamentally different types of numerical coding can produce such similar results with respect to the distance effect.

Size Effect

The size effects produced by the four models are shown in Figure 4. There is a striking qualitative difference in the performance of models trained with naturalistic versus non-naturalistic training environments.[4] The former produce a generally positive linear relation between number size and judgment time, with the exception of distances 1-2. By contrast, the latter shows a size effect only for distances 5-8, and diverge considerably from a linear relation for distances 1-4. Additional modeling is necessary to determine what factors contribute to the failure of the uniformly-trained models to produce size effects for distances 1 and 2.

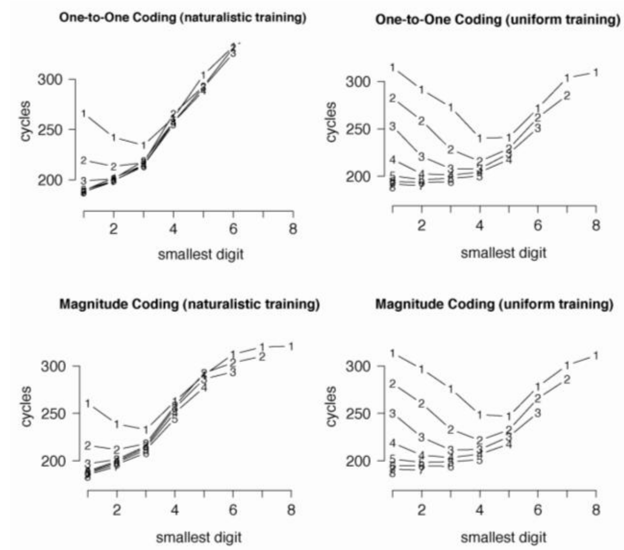

Figure 4: Size effects for the four models. Each line represents comparisons of the same distance.

By contrast, the ability of a model to account for the size effect appears to be relatively independent of whether it uses a magnitude or one-to-one number representation. As with the distance effect, additional work is necessary to determine how fundamentally different types of number representation can produce such similar results with respect to the size effect.

Development of the Distance Effect

We next turn to the development of the distance effect. The results thus far indicate that naturalistic training environments are critical for accounting for adult distance and size effects. Additionally, pilot simulations indicated that models that utilize magnitude number representations do not produce enough errors to account for that dimension of development. For these reasons, we focused our developmental efforts on the naturalistic/one-to-one model.

Sekuler and Mierkiewicz (1977) investigated the distance effect in kindergarten, first grade, fourth grade, seventh grade and adult subjects. Their results are shown in Figure 5. They reported that kindergarteners were significantly slower than other ages, first graders were significantly slower than all age groups except kindergarteners, and that the decision times of fourth graders, seventh graders and adults did not differ significantly. They also reported that the slope of the distance response curves was steeper for kindergarteners than other age groups.

Figure 6 presents the distance effect (averaged across 10 simulations) produced by the naturalistic/one-to-one model after 1200, 1600, 2000, 2400, and 2800 trials[5]. The model provides a nice qualitative account of the developmental data, showing distance effects at all time points as well as a steady decrease in response time.

[4] At the time of submission, we did not have access to empirical data on the size effect to quantify these models fits. We are working on gaining such access.

[5] These time points were chosen to align model-produced error rates with the error rates reported by Sekuler and Mierkiewicz.

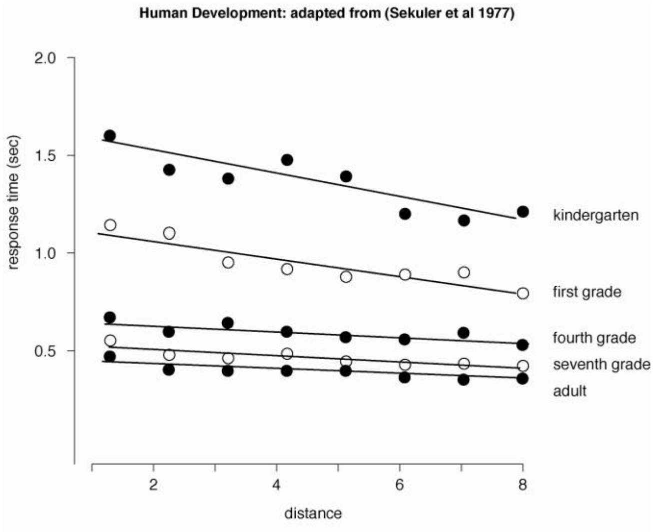

Figure 5: Development of distance effect from kindergarten to adulthood (adapted from Sekuler & Mierkiewicz, 1977).

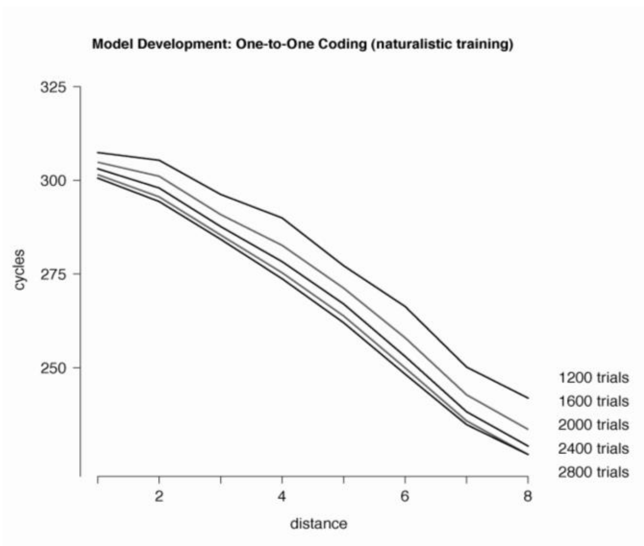

Figure 6: Development of distance effect for the naturalistic/one-to-one model from 1200 to 2800 trials)

However, the model fails to capture the interaction reported by Sekuler and Mierkiewicz : the slope of the 1200 trial line (corresponding to the kindergarten distance effect) is not qualitatively steeper than the slope of the 2800 trial line (corresponding to the adult distance effect).

We were unable to evaluate the quantitative fit of the model to the Sekuler and Mierkiewicz (1977) response time data because it is no longer available. However, Holloway and Ansari (2008) recently performed a similar experiment.[6] They had six, seven, and eight year old children make comparisons at distances 1-6. Their results are shown in Figure 7.

[6] We thank Daniel Ansari and Ian Holloway for sharing their data with us.

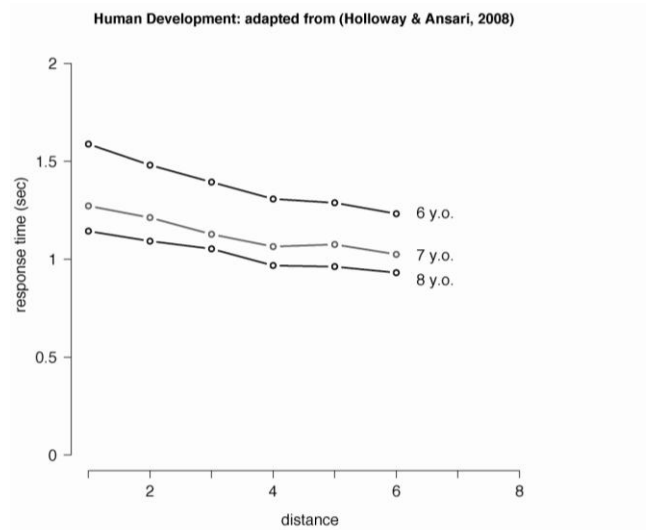

Figure 7: Distance effects at age 6, 7, and 8 years old (data from Holloway &Ansari, 2008).

We linearly regressed the performance of the model at 1200, 1600, and 2000 trials against their six, seven, and eight year old data, respectively. The model accounted for 44% of the variance in the data ($R^2 = .441,\ p = \ .003$).

Although we were unable to evaluate the quantitative fit of the model to Sekuler and Mierkiewicz's response time data because it was not available, we were able to evaluate the fit of the model to their error rate data because it was reported numerically in the original article. Table 2 presents their developmental error rate data and the error rates of our model. The model provides a good quantitative account of error rate as a function of age. The correlation between the model and the human data is 0.97 (*p* = .004).

Table 2: Error rates for the developmental simulations of the naturalistic/one-to-one model.

Human (Age)	Errors (%)	Model (Trials)	Errors (%)
Kindergarten	18.4	1200	18.3
First Grade	16.7	1600	15.1
Fourth Grade	11.8	2000	12.8
Seventh Grade	12.5	2400	12.1
Adult	7.9	2800	8.3

Discussion

The current study extends prior connectionist efforts to understand the distance and size effects. We systematically varied training environment and number representation and examined the effects on the adult distance and size effects. Models trained in naturalistic training environments, where the frequency of numbers falls off as a function of their absolute magnitude, provide better quantitative accounts of the distance effect and better qualitative accounts of the size effect. By contrast, the choice of number representation had little effect on these models' ability to account for the adult distance and size effects.

The current study is the first to address the development of the distance effect. The naturalistic/one-to-one model provided a good qualitative account of distance effects at different ages. It also provided a good account of decreasing error rates with development.

One limitation of the development simulation was that it did not account for the interaction observed by Sekuler and Mierkiewicz (1977), whereby the distance effect is most pronounced for kindergarteners and decreases throughout development. Further research is necessary to understand this limitation of the model.

Another limitation, one shared with the pioneering Zorzi and Butterworth (1999) model, is that the models considered here only perform the comparison task. By contrast, the Verguts et al. (2005) model also performs naming and parity judgment tasks and can thus be evaluated against a broader range of data. Future research is required to extend the range of the models considered here to new tasks.

Although the developmental model produced changes in error rates and comparison speed that parallel human data, further work is necessary to more completely model the development of number comparison. In particular, the model needs to account for the more pronounced distance effect of younger participants reported by Sekuler and Mierkiewicz. One reason our model may have failed to capture this trend is that we trained the model using distributions based on the occurrence of numbers in adult language. One approach to improving the developmental simulations may be to use training data that parallel the distributions of numbers in children's and child-directed speech.

Acknowledgments

Mark Lewis's work was supported by a U.S.D.E. Institute of Education Sciences (IES) training grant (Title: Interdisciplinary Education Sciences Training Program IES Award #R305C050059 University of Minnesota PRF# 473473).

References

Dehaene, S., & Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition*, *43*(1), 1-29.

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

Holloway, I. D., & Ansari, D. (2008). Domain-specific and domain-general changes in children's development of number comparison. *Developmental Science*, *11*, 644-649.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature (215),* 1519-120.

Parkman, J. M. (1971). Temporal aspects of digit and letter inequality judgments. *Journal of Experimental Psychology*, *91*, 191-205.

Rule, S. J. (1969). Equal discriminability scale of number. *Journal of Experimental Psychology*, *79*, 35-38

Sekuler, R., & Mierkiewicz, D. (1977). Children's judgments of numerical inequality. *Child Development*, 630-633.

Verguts, T., Fias, W., & Stevens, M. (2005). A model of exact small-number representation. *Psychonomic Bulletin & Review*, *12*(1), 66-80.

Zorzi, M., & Butterworth, B. (1999). A computational model of number comparison. In *Proceedings of the Twenty-First Annual Conference of the Cognitive Science Society* (pp. 772–777).