Published in final edited form as:

Atten Percept Psychophys. 2013 April ; 75(3): 614–629. doi:10.3758/s13414-012-0409-6.

On the relativity of relative frequencies

Erin Warren and

Department of Psychology, University of North Carolina Wilmington, 601 South College Road, Wilmington, NC 28403-5612

Dale J. Cohen

Department of Psychology, University of North Carolina Wilmington, 601 South College Road, Wilmington, NC 28403-5612

Abstract

The most prominent models of numerical representation posit that numerical symbols are converted into a single internal abstract representation prior to estimation and comparison processing. Here, we (1) provide a mathematical analysis of the predictions of the abstract representation hypothesis assuming the validity of the analogue representation hypothesis, (2) run a simulation to assess the pattern of data that results from our mathematical analysis, and (3) conduct two experiments to test the predictions of that model using relative frequencies as inputs. We assess relative frequencies in a typical numerical distance task, whereby the participants are presented two relative frequencies and asked to identify the one that represents the larger quantity. Our data reveal that relative frequencies' numerical representations (1) are analogue and (2) are scale-specific (i.e., non-abstract).

Keywords

Numerical Cognition; Numerical Representation; Proportion; Relative Frequency; Distance effect

Current research suggests that numerical symbols are associated with a psychological representation of quantity that is independent of the format of the numerical symbol (i.e., abstract) and has features similar to physical quantities (i.e., analogue; e.g., Dehaene, 1992; Carey, 2001). The former is often referred to as the “abstract representation hypothesis” and the later is referred to as the “analogue representation hypothesis.” Here, we (1) provide a mathematical analysis of the predictions of the abstract representation hypothesis assuming the validity of the analogue representation hypothesis, (2) run a simulation to assess the pattern of data that results from our mathematical analysis, and (3) conduct two experiments to test the predictions of that model using relative frequencies as inputs.

In a classic experiment, Moyer and Landauer (1967) presented participants with two numbers side-by-side and asked them to determine which number was larger. The authors found that participants' reaction times (RTs) were an inverse function of the numerical distance between the two numbers presented. Moyer and Landauer determined that this effect, now known as the *distance effect*, is well described by the Welford (1960) function:

$$RT = a + k \cdot \log [L / (L - S)]$$

where RT is the response time, a and k are constants, L is the larger number being compared, and S is the smaller number being compared. Because the time it takes observers to distinguish between two stimuli in which quantity is inherent, like line length, also exhibits the Welford function (Welford, 1960), the authors concluded that numerical stimuli are represented analogous to physical quantities. The distance effect is robust and has been found when comparing dots (Buckley & Gillman, 1974), single Arabic digits (Buckley & Gillman, 1974; Dehaene, 1995; Moyer & Landauer, 1967), verbal notation of numbers (Dehaene & Akhavein, 1995), multi-digit comparisons (Dehaene, 1990; Hinrichs, Yurko, & Hu, 1981; Zhang & Wang, 2005), Japanese number words (Takahashi & Green, 1983), among other stimuli.

Analogue Representation

The proposition that the distance effect is a marker for analogue representation is widely accepted in the numerical cognition literature (e.g., Dehaene, 1990; Dehaene, 1995; Hinrichs et al., 1981; Zhang & Wang, 2005). This proposition is based on the premise that a numerical symbol activates a psychological representation of quantity. It is generally assumed that every time one views a symbol, its associated quantity fluctuates a bit. So, when we view a 15, sometimes we understand it as slightly greater than 15, and sometimes slightly less. The psychological distribution of quantities that we associate with the numerical symbol 15 describes our understanding of the numerical symbol 15. This distribution can be summarized by its placement on a psychological number line and its variance. The time it takes one to distinguish between two numbers is hypothesized to be related to the amount of overlap between the two associated psychological distributions. If there is a large amount of overlap, then the RT will be long. If there is a small amount of overlap, then the RT should be short. The numerical distance effect is hypothesized to emerge because numerically close numbers have large overlaps of their distributions and numerically distant numbers have small overlap of their distributions. This interpretation has been the foundation of much of the numerical cognition literature to date.

One can express this description more precisely mathematically. Let Θ represent the symbol for of a given number and let f_T represent the transcoding operator that maps the numerical symbol onto a psychological representation (ψ). Here, the transformation from a numerical symbol to a psychological representation can be expressed mathematically with the following equation,

$$\psi = f_T(\Theta_i) + e_T, \quad (1)$$

where the subscript “ i ” identifies the specific numerical symbol represented, and e_T represents perceptual variability associated with the transcoding process. When comparing two numerical symbols of the same stimulus class (e.g. two integers, Θ_i & Θ_{i+n}), both symbols are first transcribed into psychological representations (ψ_i & ψ_{i+n}) before a comparison occurs on those representations. Let f_C represent the comparison function that operates on the psychological representations. The comparison process can be described with the following equation,

$$\psi_n = f_C(f_T(\Theta_i) + e_T, f_T(\Theta_{i+n}) + e_T) + e_C, \quad (2)$$

where e_C represents perceptual variability associated with the transcoding process. Finally, the participant must convert the psychological representation of the difference (ψ_n) into a response. The response is some function (r) of the output. Here, the response is a physical reaction that is timed (RT),

$$RT = r(\psi_n). \quad (3)$$

RT will be monotonically related to that overlap of the perceptual distributions of Θ_i & Θ_{i+n} if (as is often assumed) r is a monotone transformation function. Therefore, the numerical distance effect is a predicted result of an analogue psychological representation of quantity. Simulations based on these assumptions have successfully modeled the numerical distance effect in a number comparison task (e.g., Birnbaum & Jou, 1990; Link, 1990).

It should be noted, that although the distance effect is predicted by an analogue representation, its presence is not conclusive evidence of an analogue representation. Verguts, Fias, and Steven's (2005) present a network model in which the distance effect manifests from the response selection stage, rather than the comparison stage described above (but see, Cohen Kadosh, Tzelgov, & Henik, 2008). To make conclusions about the nature of the underlying representations of quantity as we attempt here, we must accept one or the other model from which to base our assumptions. In the present article, we work with the premise that the numerical distance effect arises from the perceptual distributions of quantity as described above. If these assumptions are falsified, the conclusions of the present work will have to be re-evaluated. We return to address the Vergut et al. (2005) model in the discussion.

Abstract Representation

The most widely accepted architectures of numerical cognition posit that all numerical symbols link to a single, abstract representation of quantity (see Cohen Kadosh & Walsh, 2009 for review). The majority of experiments examining the abstract representation hypothesis have used the various symbolic expressions of integers as the test stimuli. Integers, however, are unique because all symbolic expressions of the same quantity share a number name (e.g., "five" and 5 share the name [faiv]) and are highly familiar. Here, we test the abstract representation hypothesis using relative frequencies. Relative frequencies (i.e., X in Y , where X represents a position on a scale and Y represents the scale) are also a unique numerical format because a single quantity can be expressed on multiple scales. For example, the same quantity could be presented as 2 in 4 or 1 in 2. The first example is presented as 2 parts on a scale of 4 and the second example is presented as 1 part on a scale of 2. By assessing relative frequencies, one can determine whether numerical representations are scale specific.

We note here that although much of the data supporting the abstract representation and analogue representation hypotheses have been conducted on integers, the foundation of the abstract representation hypothesis is that the source of the distance effect is identical regardless of the numerical symbol used to express a quantity. As such, the abstract representation hypothesis assumes that there should be no difference between the source of the distance effect for real numbers and integers (i.e., the comparison of an analogue representation of quantity). Below, we describe the predictions of the abstract representation hypothesis in the number comparison task.

One can assess the abstract representation hypothesis by asking an observer to compare the quantities of two numerical symbols whereby each symbol belongs to a different stimulus class. For example, an observer may be asked to compare the two following relative frequencies:

$$\begin{array}{l} 4 \text{ in } 10 \\ 2 \text{ in } 8 \end{array}$$

Here, we define a stimulus class as a relative frequency with a specific scale. Therefore, the two presented relative frequencies belong to two different stimulus classes (A and B). Those positing an abstract representation system assume that, when comparison occurs, both physical stimuli (Θ_A & Θ_B) are transcoded into an abstract psychological representations (ψ_A & ψ_B), a comparison process occurs on the abstract representations (f_c), and then there is some output (ψ_{AB}). This process may be described with the following equation, which is a generalization of Equation 2:

$$\psi_{AB} = f_c(f_{TA}(\Theta_A) + e_{TA}, f_{TB}(\Theta_B) + e_{TB}) + e_C \quad (4)$$

Here, the transcoding process (f_T) may be different for every stimulus class, but the comparison process (f_c) is constant because it operates on an abstract representation. So, although the transcoding processes are linked to the stimulus class, and therefore can vary with stimulus class, the comparison process is independent of stimulus class because it has no direct access to the stimulus class. The comparison process only has access to the abstract representations on which it operates. Assuming Equation 4, one can predict the output of the comparison of two different stimuli classes from the outputs of each individual stimulus class. That is, one can assess the specific outputs in a relative quantity task for Stimulus Class A [$\psi_{A12} = f_c(f_{TA}(\Theta_{A1}) + e_{TA1}, f_{TA}(\Theta_{A2}) + e_{TA2}) + e_C$] and Stimulus Class B [$\psi_{B12} = f_c(f_{TB}(\Theta_{B1}) + e_{TB1}, f_{TB}(\Theta_{B2}) + e_{TB2}) + e_C$]. If we assume f_c is a subtraction (the specific form of the function is not important here), we get: $\psi_{A12} = \beta_-(f_{TA}(\Theta_{A1}) - f_{TA}(\Theta_{A2})) + e_{TA1} + e_{TA2} + e_-$, where β_- is the bias associated with the subtraction comparison process and e_- is the error associated with the comparison process. Stimulus Class B would be similar, $\psi_{B12} = \beta_-(f_{TB}(\Theta_{B1}) - f_{TB}(\Theta_{B2})) + e_{TB1} + e_{TB2} + e_-$. When one compares an element in Stimulus Class A to an element in Stimulus Class B, we get, $\psi_{A1B2} = \beta_-(f_{TA}(\Theta_{A1}) - f_{TB}(\Theta_{B2})) + e_{TA1} + e_{TB2} + e_-$ and $\psi_{A2B1} = \beta_-(f_{TB}(\Theta_{B1}) - f_{TA}(\Theta_{A2})) + e_{TB1} + e_{TA2} + e_-$. It is straightforward to demonstrate¹ that the sum of the error variance for $e_{A1B2} + e_{A2B1}$ (i.e., $e_{TA1} + e_{TB2} + e_{TA2} + e_{TB1} + 2e_-$) should equal the sum of the error variance for $e_{A12} + e_{B12}$ (i.e., $e_{TA1} + e_{TA2} + e_{TB1} + e_{TB2} + 2e_-$). That is, no extra error should arise in the system. The presence of extra error suggests that the abstract representation hypothesis may not be valid.

Relative Frequencies

Here, we assess relative frequencies in a typical numerical distance task, whereby the participants are presented two relative frequencies, one above the other, and asked to identify whether the bottom quantity is larger or smaller than the top quantity. Although several researchers have assessed whether the numerical distance effect is exhibited when numerical symbols expressing quantities between zero and one are compared, no researcher has assessed the effect of scale on relative frequencies. It may appear that relative frequencies and fractions are identical, and therefore should produce identical results. These two formats, however, differ in important ways. First, the word “in”, as in X “in” Y, may create a different understanding of a quantity than X/Y (X over Y). This subtle change in format could alter the way individuals process or understand the quantity. Furthermore, there exists convincing evidence that relative frequencies are processed differently than other numerical formats expressing values between zero and 1 (e.g., Gigerenzer & Hoffrage, 1995). Thus, it is not at all clear that observers will process fractions and relative frequencies identically. Below we briefly review the data regarding participants’ performance in a relative quantity task using real numbers denoting quantities between zero and 1.

¹For independent random variables, the variance of their sum or difference is the sum of their variances.

Perhaps the most well studied numerical format for symbolizing quantities between zero and one is the fraction. There is data that suggests that the numerator and denominator of fractions are processed separately. For example, Bonato, Fabbri, Umiltà, & Zorzi (2007) assessed fractions and found a “reversed” distance effect. Bonato et al.’s experiment involved comparing fractions in which the numerator remained constant (e.g. 1) and the denominator varied between 1 and 9, excluding 5. The fractions were all compared to a standard of 1/5. The researchers found that the distance effect related to the numerical distance between the denominators, rather than the value of the fraction. The authors concluded that the understanding of fractions is based on the integer components, rather than a true quantity representation of the fraction (see also Kallai & Tzelgov, 2009).

Critical of Bonato et al.’s (2007) conclusions, based on the task presented to the participants, Schneider and Siegler (2010) completed related experiments, but ones in which participants may be more apt to view the stimuli holistically. In Experiment 1, the participants were asked to compare fractions in which the values of the numerators and denominators varied. The authors found results contrary to that of Bonato et al. (2007). The data suggest that the participants made comparisons using the holistic magnitude of the fraction, rather than relying on the values of the numerator and denominator separately. The authors also found that the data demonstrated a logarithmic distance which provides further evidence that the magnitude representation was activated. The authors expanded these findings in Experiment 2 in which they had participants compare fractions that consisted of two digit numerator and denominator values. The results of Experiment 2 are parallel to that of Experiment 1. The participants compared the magnitudes of the fractions, rather than relying solely on the numerator and denominator values, and a logarithmic distance effect emerged. To examine whether fraction magnitude estimations are based on educational training, the authors completed a third experiment involving students from a local Community College. The procedures for Experiment 3 paralleled that of Experiment 2. Though the authors found a much a higher error rate with the participants’ responses, the same pattern of behavioral responses emerged, as in Experiment 2. The authors’ overall conclusion was that adults can compare and represent the magnitudes of fractions.

Congruent with Schneider and Siegler’s (2010) behavioral findings, fMRI data seem to suggest that fractions are processed holistically (Jacob & Nieder, 2009; Ischebeck, Schocke, & Delazer, 2009). Ischebeck, Schocke, and Delazer (2009) examined whether modulation of brain activation in the intraparietal sulcus correlated with the total or partial distance between two fractions presented in a numerical comparison task. Participants were presented with four conditions: a same denominator (SD) condition, a same numerator condition (SN), a congruent condition (CO), and an incongruent condition (IC). In the CO condition, the larger fraction had both a larger numerator and denominator. In this condition the participants could examine either the numerator or the denominator and respond correctly. In the IC condition, the larger fraction could have either a larger or smaller numerator or denominator. This condition was constructed to discourage segmental processing. The findings of the behavioral data suggested that both the partial distance and total distance between the fractions predicted response times in the SD, SN, and CO conditions. However, only total distance predicted the response times in the IC condition. Interestingly, only the total distance between the fractions modulated activation in the intraparietal sulcus. The authors concluded, as did Jacob and Nieder (2009), that fractions are processed holistically as their numerical value (see also Ganor-Stern, Karasik-Rivkin, & Tzelgov, 2011).

Cohen (2010) assessed decimals in a typical numerical distance task. Cohen showed that decimals rounded to a consistent digit (e.g., the third decimal place), will elicit the numerical distance effect in a typical numerical distance task. Cohen also found, however, that when decimals are rounded to different decimal places, participants will complete the numerical

distance task based solely on the physical similarity of the numbers to the standard. Cohen explained this difference by proposing that when the decimals are rounded to a consistent place, the ordinal relation between successive decimals follow the same rules as the ordinal relation between successive integers. As such, participants' knowledge of the ordinal relation between integers can inform them on the ordinal relation between decimals. When the decimals are rounded to inconsistent places, this is no longer true. Here, a one digit decimal (e.g., .5) may be larger or smaller than a two digit decimal (e.g., .32 vs .82). Thus, participants can no longer rely on their knowledge based on the ordinal relation of integers.

The previous research on fractions and decimals provides a mixed review of whether quantities less than one are processed and represented as their componential parts or as their total quantity. In the present experiment, we extend these findings by exploring the association between relative frequencies and their quantities. To assess whether a single, scale independent abstract representation exists, we asked participants to compare relative frequencies on (1) the same scale and (2) on different scales using a traditional relative quantity task (e.g., Moyer & Landauer, 1967). To specify the predictions of the traditional model of numerical distance based on an Analogue, Abstract representation (described above), we first simulated the model using a random walk model.

Simulation

We have simulated the abstract representation model described above assuming the RT-distance hypothesis, whereby RT is a monotonically decreasing function of the Euclidean distance between the percept and the criterion, described by Signal Detection Theory (e.g., Balakrishnan & Ratcliff, 1996; Ratcliff, 1978; Thomas & Myers, 1972; Zakay & Tuvia, 1998). When comparing two analogue representations, the representations may both be ordinally correct, but vary in the degree of distributed noise around each quantity. To assess the effects of these different underlying representations, we varied the relative variance of the representations. We created two different sets of representations: a "more noisy" representation and a "less noisy" representation. The different patterns of results from the comparison of numbers in the two variance conditions provide a template from which to interpret our results. Here, the relations between the more and less noisy conditions provide information about how the pattern of data will look if extra error arises in the system. Extra noise suggests a non-abstract representation because the abstract representation hypothesis predicts that the noise will remain constant across conditions.

We simulated several forms of the underlying representation consistent with the most prevalent theories in the literature: logarithmic constant variance (LogC; e.g., Dehaene, 1992); the linear scalar-variance theory (LinS; e.g., Gallistel & Gelman, 1992); and the linear constant variance theory (LinC; e.g., Verguts et. al, 2005). All three models assume that quantities have a place-value representation. That is, quantity is represented along a continuum, such as a number line (e.g., Dehaene, 1992). Similarly, all models assume perceptual variability, so there exists a psychological distribution of perceptual effects for any single quantity. The three models are specified as follow:

$$\text{LogC: } \psi_i = \log(\Theta_i) + e_i, e_i \sim \mathcal{N}(0, s), \quad (5)$$

$$\text{LinC: } \psi_i = \Theta_i + e_i, e_i \sim \mathcal{N}(0, s), \quad (6)$$

$$\text{LinS: } \psi_i = \Theta_i + e_i, e_i \sim \mathcal{N}(0, \Theta_i^* s) \quad (7)$$

We simulated each of the three theories described above in a standard numerical distance task using a simple random walk model. Perceptual distributions were generated for each proportion between 0.1 and 0.9 in steps of 0.1 consistent with each of three theories (i.e., using formulae 5-7). To assess the influence of noise on the data, for each theory, we generated a set of “less noisy” and “more noisy” perceptual distributions (i.e., larger average SDs). In the present simulations, the standard proportion was 0.5, and the probes were 0.1-0.9 (in steps of 0.1), excluding 0.5. The task was to identify whether the standard was larger than the probe. To do this, a sample value was drawn from the perceptual distribution of the standard and one from the probe. A running total of the difference between the two samples is logged (i.e., $\text{tot}_i = \text{tot}_{i-1} + \text{SQ}_i - \text{PQ}_i$), where tot_i is the running total for sample i , SQ_i is the standard quantity drawn on sample i , and PQ_i is the probe quantity drawn on sample i . A response is when tot_i crosses pre-specified decision threshold (DT). If tot_i crosses positive DT, the standard is identified as larger. If tot_i crosses negative DT, the probe is identified as larger. The model also incorporates a decision bias (*bias*), by adding a constant to tot_i prior to the start of the run ($\text{tot}_i = \text{bias}$, where $i = 0$). For example, if $\text{bias} > 0$, the decision is facilitated toward identifying the probe as smaller. To ensure a realistic threshold, the simulation is optimized to identify a threshold that results in an overall error rate between .01-.15. This simulation was run 2,000 times per probe for each condition. RT is simulated by transforming the number of runs prior to response as follows: $\text{RT} = 150 + 350 * \text{ARN}^{0.2}$, where ARN is the average number of runs across the 10000 simulations per probe². To assess how accurately the models matched human data, we compared the simulated data to the Welford function (Welford, 1960). The Welford function has been shown to approximate human data well in the numerical distance task. Figure 1 shows the simulated RTs and the fit of the Welford function for all three models by condition.

Results

LogC

To assess whether the LogC can simulate data similar to the Welford function, we ran a regression with the mean simulated RTs for each probe as the criterion variable and the Welford transformed inputs (e.g., $\log(\text{Larger}/(\text{Larger}-\text{Smaller}))$) as the predictor variable. Perceptual distributions were simulated consistent with Equation 5. We present the “less noisy” and “more noisy” simulations below.

In the “less noisy” simulation, we set $s = \log(2)$ and $\text{bias} = 0$. The Welford function fit the simulated data extremely well, $F(1,6)=551$, $p<.001$, $r^2 = 0.99$ (slope = 140, intercept = 600). The overall error rate was 14%. Similar to the RT data, the errors were asymmetrically distributed around the standard, with more errors above the standard than below the standard (see Table 1).

In the “more noisy” simulation, we set $s = 2 * \log(2)$ and $\text{bias} = 0$. The Welford function fit the simulated data extremely well, $F(1,6)=253$, $p<.001$, $r^2 = 0.97$ (slope = 177, intercept = 739). The overall error rate was 13%. Similar to the RT data, the errors were asymmetrically distributed around the standard, with more errors above the standard than below the standard (see Table 1).

LinC

To assess whether the LinC can simulate data similar to the Welford function, we ran a regression with the mean simulated RTs for each probe as the criterion variable and the Welford function as the predictor variable. Perceptual distributions were simulated

²This formula was taken from Ashby and Maddox (1994). Nevertheless, the pattern of data remains the same with different values.

consistent with Equation 6. We present the “less noisy” and “more noisy” simulations below.

In the “less noisy” simulation, we set $s = 0.1$ and bias = 0. The Welford function fit the simulated data well but did not mimic the asymmetric feature of the Welford function, $F(1,6)=27$, $p=.001$, $r^2 = 0.79$ (slope = 51, intercept = 482). The overall error rate was 14%. Similar to the RT data, the errors were symmetrically distributed around the standard, (see Table 1). To improve the fit, we added a positive bias that equaled 20% of the decision threshold. The addition of the bias greatly improved the fit, $F(1,6)=59$, $p<.001$, $r^2 = 0.9$ (slope = 55, intercept = 478). The overall error rate was 14%. Similar to the RT data, the errors were asymmetrically distributed around the standard, with more errors above the standard than below the standard (see Table 1).

In the “more noisy” simulation, we set $s = 2*0.1$ and bias = 0.2 (similar to the “less noisy” condition). The Welford function fit the simulated data extremely well, $F(1,6)=142$, $p<.001$, $r^2 = 0.95$ (slope = 88, intercept = 546). The overall error rate was 13%. Similar to the RT data, the errors were asymmetrically distributed around the standard, with more errors above the standard than below the standard (see Table 1).

LinS

To assess whether the LinS can simulate data similar to the Welford function, we ran a regression with the mean simulated RTs for each probe as the criterion variable and the Welford function as the predictor variable. Perceptual distributions were simulated consistent with Equation 7. We present the “less noisy” and “more noisy” simulations below.

In the “less noisy” simulation, we set $s = 0.2$ and bias = 0. The Welford function fit the simulated data well but did not mimic the asymmetric feature of the Welford function, $F(1,6)=53$, $p=.001$, $r^2 = 0.88$ (slope = 54, intercept = 481). The overall error rate was 15%. Unlike the RT data, the errors were asymmetrically distributed around the standard, with more errors above the standard than below the standard (see Table 1). To improve the fit, we added a positive bias that equaled 20% of the decision threshold. The addition of the bias greatly improved the fit, $F(1,6)=190$, $p<.001$, $r^2 = 0.96$ (slope = 67, intercept = 477). The overall error rate was 12%. Similar to the RT data, the errors were asymmetrically distributed around the standard, with more errors above the standard than below the standard (see Table 1).

In the “more noisy” simulation, we set $s = 2*.5$ and bias = 0.2 (similar to the “less noisy” condition). The Welford function fit the simulated data extremely well, $F(1,6)=205$, $p<.001$, $r^2 = 0.97$ (slope = 102, intercept = 554). The overall error rate was 13%. Similar to the RT data, the errors were asymmetrically distributed around the standard, with more errors above the standard than below the standard (see Table 1).

Discussion

The purpose of the present simulation is to determine the performance characteristics of the traditional abstract representation model under the different variance conditions. It is not an attempt to fully explore the validity of the three different quantity representations. Indeed, all three representations performed well in as much as all three manifested the Welford function³.

³The pattern of results remains the same when the models are simulated with integers rather than proportions.

The simulation shows that increased variance of the underlying representations will increase both the slope and the intercept associated with the comparison process. When the variability between the two quantities is small, the results are both an efficient comparison (i.e. small slope values) and a short pre-comparison process (i.e. small intercept). In contrast, as the representations become increasingly “noisy,” the RT values increase, demonstrated by the larger slopes and intercepts. The increase in the intercept results from the general slowing of the comparison process for all stimuli.

Experiment 1

In Experiment 1 we ran a numerical distance task with relative frequencies of different scales. If a single, scale independent abstract representation exists, the slopes of conditions in which different scales are compared will be predictable (i.e., no extra noise will arise) from the same scale conditions. This would occur because the two presented relative frequencies would be converted into the same abstract representation prior to comparison. If representations are scale dependent, there is no expectation that the slopes of the same scale or different scale conditions would be equivalent. Indeed, one signature of different representations is that extra noise will be introduced in the multiple scale comparison conditions which will reduce the efficiency of the comparison, therefore the slopes and intercepts of the different scale conditions should be greater than the slopes of the same scale conditions.

Method

Participants

Fifty one students volunteered to participate. The participants received course credit for their participation. The participants were unaware of the purpose of the experiment.

Apparatus and Stimuli

All stimuli were presented on a 15 inch color monitor with a 60 Hz refresh rate controlled by a Pentium microcomputer running a Windows based operating system. The resolution of the monitor was 1024×768 .

A trial consisted of two relative frequencies, each in the format “X in Y,” presented one above the other⁴, centered on the computer monitor. Each integer subtended 1.33 degrees vertical visual angle. In every condition and for each participant, the quantity represented by each relative frequency was randomly selected from a uniform distribution ranging from .25-.75.

There were five different conditions in which the scale, Y , of the relative frequency was manipulated: two single scale conditions, two dual scale conditions, and one multiple scale condition. In the single scale conditions, participants were asked to compare pairs of relative frequencies that were on the same scale. The two conditions differed in the scale that the relative frequency was presented in. In the 1000-1000 condition, both relative frequencies were presented in the format “X in 1000.” In the 736-736 condition, both relative frequencies were presented in the format “X in 736.” In the dual scale condition, participants were asked to compare pairs of relative frequencies that were on two different scales. The two scales we presented were always 1000 and 736⁵. The conditions differed on the physical position of the scale on the computer screen. In the 1000-736 condition, the relative

⁴We have found that presenting our relatively long stimuli (e.g., “323 in 1000”) one-above-the-other increases readability.

⁵We chose 1000 as a scaling factor because it is relatively commonly used. We chose 736 as a scaling factor because it was not a simple factor of 1000, thus participants could not easily calculate a conversion between 1000 and 736.

frequency on the top was always presented in the format “X in 1000” and the relative frequency on the bottom was always presented in the format “X in 736.” In the 736-1000 condition, the presentation order was reversed. In the multiple scale condition (termed multi-scale condition), participants were asked to compare pairs of relative frequencies that were on unpredictable scales. The values for each relative frequency in the presented pair of this condition were created by generating a relative frequency as “X in 1000” and then reducing the position and scale to the lowest common denominator. For example, the initial generation might create the quantity 540 in 1000. This quantity would then be reduced by the greatest divisor to produce the lowest common denominator, which, in this example, would be 27 in 50. The multi-scale condition was included to assess whether the presence of unpredictable scales influences participants representations (or strategies).

Procedure

Participants were tested in a small dark room that contained a single computer. The participants were instructed that they would be presented two relative frequencies representing different quantities. These two relative frequencies were presented one above the other. Participants were asked to indicate whether the bottom relative frequency represented a larger or smaller quantity than the top relative frequency by pressing one of the two keys on the keyboard. Participants pressed the “L” button, for a response of larger or “S” button, for a response of smaller. Participants were instructed that speed was important but accuracy was essential.

Each trial consisted of the two relative frequencies being presented, one above the other, on a single screen. The stimulus remained visible until a response was made. There was a blank screen between each trial that lasted 2000ms.

Each condition consisted of 10 practice trials and 100 experimental trials. The order of the five conditions was randomized between participants. The participants’ RTs were recorded in milliseconds. Each participant completed all five conditions during a 90 minute session.

Results

Of the fifty-one participants that completed the experiment, the data of forty-five participants were examined. Six participants were removed because their error rate in the two single scale conditions, 1000-1000 and 736-736, was higher than 10% (Mean Error Rate = 74.0%). The high error rates demonstrated that the participants either did not understand the task or simply did not take the task seriously and thus were not included in the statistical analysis.

Prior to analyzing the data, we trimmed the high and low values from the RT data. The top 2% of the response time data, per condition, was excluded from our analysis (Ratcliff, 1993) (See Table 2). The low RT cut-off threshold was determined, by condition, to be the RT below which participants responded with an accuracy of 51% correct or less (Ratcliff, 1993). For this analysis, we examined RTs in bins of 50ms increments⁶ (see Table 2). The final RT data was normalized by taking the $\log(RT)$.

To assess the numerical distance effect, we calculated both the Welford and the symmetrical Log D function ($\log |\text{stimulus-probe}|$) on the RT data by participant and condition. Although both functions fit the data about the same, the symmetrical Log D function fit slightly better as measured by r^2 (see Table 3)⁷. To examine the effects of our manipulations, the slopes

⁶We statistically explored why the multi-scale condition had a lower cut-off than the other conditions and could not find a consistent pattern.

and intercepts for each participant $\times$ condition of the symmetrical Log D regressions were analyzed through a series of mixed model ANOVAs. Figure 2 plots the average best fit functions for each condition.

A mixed model ANOVA with subject as the random effect and condition as the within subject variable revealed a significant main effect of condition on slope, $F(4,220) = 9.13$, $p < .001$, $MSE = .008$. Tukey's HSD post-hoc test ($p < .05$) revealed that the 736-1000 condition and the 1000-736 condition produced significantly steeper slopes than the 1000-1000 condition, 736-736 condition, and the multi-scale condition. The analysis also revealed that the 736-1000 and 1000-736 conditions were not significantly different from each other and the 1000-1000 condition, 736-736 condition, and multi-scale condition were not significantly different from each other. In addition, a one-sample t-test ($t(134) = -16.45$, $p < .01$) demonstrated that the three shallowest slope conditions (1000-1000, 736-736, and multi-scale condition) were significantly different from zero (see Table 2).

In the dual scale and multi-scale condition, neither the position nor the scale elements alone would provide the information necessary to produce the monotone decreasing function our data displayed. It is only through integrating the position and the scale elements into a single quantity that the numerical distance effect in our data could be present in the dual scale and multi-scale condition. In the single scale conditions, participants could have produced the monotone decreasing function by attending only to the position element. If participants attended only to the position element of the relative frequencies, then both the 1000-1000 and 736-736 conditions would be identical only when analyzed as a function of the integer in the position element. When analyzed as a function of the quantity represented by the relative frequencies (e.g., $500/1000 = 0.5$ whereas $500/736 = .679$), the slopes of the 736-736 condition should be about $\frac{3}{4}$ as steep as the 1000-1000 condition (i.e., $736/1000 = 0.736$). If, however, participants attended to the relative frequencies holistically, then both the 1000-1000 and 736-736 conditions would be identical when analyzed as a function of the quantity represented by the relative frequencies (e.g., $500/1000 = 0.5$ and $368/736 = 0.5$). Our data show that, when analyzed as a function of the quantity represented by the relative frequencies, the 736-736 is virtually identical to the 1000-1000 condition (only 3% less steep). In addition, a t-test revealed that the slope of the 736-736 condition was significantly steeper than the slope predicted if participants attended only to the position element of the relative frequencies $t(44) = 23.12$, $p < .01$. Thus, the data support the hypothesis that participants were attending to the information in the combined position and scale elements of the relative frequencies in the single scale conditions.

A mixed model ANOVA with subject as the random effect and condition as the within subject variable revealed a significant main effect of condition on intercept, $F(4,220) = 67.31$, $p < .001$, $MSE = .034$. Tukey's HSD post-hoc test ($p < .05$) revealed that the 1000-1000 condition and the 736-736 condition were not significantly different from each other, however both conditions had a significantly smaller intercept than the 736-1000 condition and the 1000-736 condition. The 736-1000 condition and 1000-736 condition were found not to be significantly different from each other. In addition, the multi-scale condition had a significantly larger intercept than the other four conditions (see Table 2).

A one-way ANOVA revealed a significant main effect of condition on performance errors, $F(4,220) = 21.18$, $p < .0001$. Tukey's HSD post-hoc test ($p < .05$) revealed that the 1000-1000 condition was not significantly different from the 736-736 condition. The 736-1000 condition was not significantly different from the 1000-736 condition and neither of these

⁷The r^2 appear small because the regressions were run on each participant's raw data and the average r^2 is presented. When the x-values are binned and the data is averaged over participants, the r^2 values fall between .8 and .95.

conditions were significantly different from the multi-scale condition. However, both the 1000-1000 and the 736-736 conditions were significantly lower than the 736-1000 condition, the 1000-736 conditions, and the multi-scale condition (see Table 2).

Discussion

The data from Experiment 1 suggests that relative frequencies' numerical representations (1) are analogue, (2) describe the integrated position and scale information, and (3) are non-abstract. We briefly discuss these findings below and expand on them in the General Discussion.

The presented data is the first to demonstrate a numerical distance effect with quantities represented as relative frequencies. Participants' RTs in all conditions were a monotone decreasing function of numerical distance, revealing the traditional *numerical distance effect*. It is only through integrating the position and the scale elements of the relative frequency into a single quantity that the numerical distance effect in our data could be present in the dual scale and multi-scale condition. The relatively long RTs in the dual and multi scale condition are consistent with the times of other researchers who have identified the numerical distance effect with fractions (e.g., Ischebeck et al., 2009). In addition, our data support the hypothesis that participants were integrating the position and the scale elements into a single quantity in the single scale conditions. The presence of the numerical distance effect has traditionally been interpreted as supporting the hypothesis that the underlying numerical representation is analogue (Dehaene, 1990; Dehaene, 1995; Hinrichs et al., 1981; Zhang & Wang, 2005).

Because the abstract representation model hypothesizes that all comparisons are completed in the same, abstract representation, a critical feature of this model is that the speed with which two quantities are compared would be equivalent regardless of numerical format or scale. Our finding that the slope of the dual scale conditions slowed significantly from that of the single scale conditions along with the higher intercepts, support that there was added noise in the system when comparing two scales. This finding calls the abstract representation hypothesis into question.

The slopes for the two single scale conditions (i.e., 1000-1000 and 736-736 conditions) were statistically equivalent and shallow. In addition, the intercept was relatively small. The shallow slopes of the 1000-1000 and 736-736 conditions indicate an efficient comparison process. Theoretically, the most efficient comparisons should be made within a shared numerical representation. Given that one would logically suspect that relative frequencies with the same scale and numerical format would share a quantity representation and the data demonstrate a very efficient comparison process, we conclude that relative frequencies with a common scale share a representation.

The slopes for the dual scale conditions (i.e., 736-1000 and 1000-736 conditions) were equivalent (i.e., there was no effect of presentation order). The data revealed that the quantity comparisons in the dual scale conditions were slow, highly affected by the numerical distance between the relative frequencies, and were statistically steeper than the other conditions. There are at least two possible conclusions one can infer from the dual scale data: (1) each scale of the dual scale comparison has its own scale specific representation and the inefficient comparison is a feature of between representation comparisons, or (2) when dual scale comparisons are required, a unique representation is generated that accommodates both scales, but this representation is vague and indistinct relative to a single scale representation. Both of these conclusions reveal that scale is a critical factor in numerical representation. Although the current data cannot adjudicate

between these two conclusions, the difference between the slopes relating RT to numerical distance in the single and dual scale conditions refutes the prediction of those models espousing a single abstract representation in which comparisons between quantities occurs. We discuss this further in the General Discussion.

The multi-scale condition, in which the scale consistently varied, produced a slope equivalent to that of the single scale conditions. The equivalent slope suggests that when relative frequencies with varying denominators were compared, the comparisons occurred within a single numerical representation. However, the results of the dual scale conditions suggest that multiple scale comparisons should be inefficient. The very high intercept value in the multi-scale condition may explain this inconsistency. Specifically, the high intercept together with the shallow slope suggests that the participants applied a mathematical strategy to convert the numerical stimuli into a common scale prior to comparison. Once the relative frequencies were converted into a common scale, participants could efficiently compare the two quantities within a single numerical representation (see Figure 2). This finding is consistent with Siegler, Thompson, and Schneider (2011), who also found the application of “numerical transformation strategies” when 6th and 8th graders estimated the magnitude of fractions involving a number line estimation task.

To assess whether the numerical representations are scale specific or format specific, in Experiment 2, we asked participants to compare relative frequencies and decimals in a numerical distance task. We examined decimals because they allowed us to separate format and scale. Specifically, decimals are a different numerical format than relative frequencies, but when decimals are rounded to three decimal places they share the same scale as relative frequencies with a scale element of 1000. Therefore, if numerical representations are format specific, we should see steep slopes in a numerical distance task whereby participants are asked to compare decimals to relative frequencies regardless of scale. However, if numerical representations are scale specific, we should see shallow slopes in a numerical distance task whereby participants are asked to compare decimals to relative frequencies with a scale of 1000 and steep slopes when participants are asked to compare decimals to relative frequencies with a scale other than 1000.

Experiment 2

Experiment 2 was similar to Experiment 1. The participants were asked to compare two quantities in a numerical distance task. One of the quantities was presented as a decimal and the other as a relative frequency. The participants' response times were recorded.

Method

Participants

Participants consisted of 54 volunteer college students. The participants received course credit for their participation. The participants were naive to the task.

Apparatus and Stimuli

The apparatus used in Experiment 2 was the same as Experiment 1.

The stimuli in Experiment 2 were identical to that of Experiment 1 with the exception of the addition of a decimal stimulus and the composition of the conditions. The values of the decimal stimuli were randomly chosen from a uniform distribution ranging from .25-.75 and rounded to the third decimal place. The decimals were consistently rounded because Cohen (2010) demonstrated that participants comparing rounded decimals will manifest the numerical distance effect.

There were six conditions in which participants were asked to compare a relative frequency to a decimal. In the 1000-decimal condition, a relative frequency with a scale of 1000 was presented over a decimal. In the decimal-1000 condition, a decimal was presented over a relative frequency with a scale of 1000. In the 736-decimal condition, a relative frequency with a scale of 736 was presented over a decimal. In the decimal-736 condition, a decimal was presented over a relative frequency with a scale of 736. In the multi-scale-decimal condition, a relative frequency with an unpredictable scale was presented over a decimal. In the decimal-multi-scale condition, a decimal was presented over a relative frequency with an unpredictable scale. Similar to Experiment 1, the values for the multi-scale condition were created by generating a relative frequency as X in 1000 and then the position and scale were reduced to the lowest common denominator.

Procedure

The procedure in Experiment 2 was identical to that of Experiment 1 with the exception that participants completed six sessions rather than five.

Results

Of the fifty-four participants that completed the experiment, the data of forty-one participants were examined. One participant was removed for not completing all six conditions. Twelve participants were removed from the data set because their error rates in the two easiest conditions, 1000-deci and deci-1000, were higher than 10% (Mean Error Rate = 51.9%). The high error rates demonstrated that the participants either did not understand the task or simply did not take the task seriously and thus were not included in the statistical analysis.

Similar to Experiment 1, the first ten trials of each condition were practice trials and were not statistically examined. The slowest and quickest response times were removed from the data set by the same criteria used in Experiment 1. The slowest 2% of the response time data were removed per condition (Ratcliff, 1993). Like Experiment 1, the faster response times were eliminated until the responses consistently demonstrated a 51% or better accuracy threshold, per condition (Ratcliff, 1993). Again, this analysis occurred with 50ms increments. In the 1000-deci and deci-1000 conditions, the participants always responded above 51% correct and thus all of the data were included (see Table 4).

Like Experiment 1, the initial distribution of the data was positively skewed. To normalize the distribution of the data, the response time data was again transformed by using the log (base 10) of the response time. To assess the numerical distance effect, we calculated both the Welford and the symmetrical Log D function on the RT data, by participant and condition. Although both functions fit the data about the same, the symmetrical Log D function fit slightly better as measured by r^2 (see Table 5). Therefore to assess the effects of our manipulation, the slopes and intercepts for each participant by condition of the symmetrical Log D regressions were analyzed through a series of 2x3 mixed model ANOVAs.

A 2 (decimal top/ decimal bottom) X 3 (1000/736/ multi-scale) mixed model ANOVA with subject as the random effect and condition and order as the within subject variables revealed a significant main effect of condition on slope, $F(2,240) = 14.86$, $p < .001$, $MSE = .009$. Tukey's HSD post-hoc test ($p < .05$) revealed that the 736 condition ($M = -.15$, $SD = .13$) had a significantly higher slope than both the 1000 condition ($M = -.09$, $SD = .04$) and the multi-scale condition ($M = -.07$, $SD = .10$). There was no significant difference between the 1000 and multi-scale condition. The analysis also revealed that the order of presentation (decimal on top or bottom) was not significant, $F(1,240) = 2.886$, $p = .091$, and there was not a

significant interaction between condition and order, $F(2,240) = .404$, $p = .668$. In addition, a one-sample t -test ($t(163) = -13.49$, $p < .01$) demonstrated that the two most shallow slope conditions (1000, multi-scale condition) were significantly different from zero (see Table 4). A 2 (decimal top/ decimal bottom) $\times$ 3 (1000/736/ multi-scale) mixed model ANOVA with subject as the random effect and condition and order as the within subject variables revealed that there was a significant main effect of condition on intercept, $F(2,240) = 66.56$, $p < .001$, $MSE = .030$. Tukey's HSD post-hoc test ($p < .05$) demonstrated that all three conditions were significantly different from each other. The 1000 condition ($M = 3.05$, $SD = .13$) had the smallest intercept, the 736 condition ($M = 3.27$, $SD = .19$) had the next smallest intercept, and the multi-scale condition ($M = 3.35$, $SD = .19$) produced the largest intercept. There was no significant effect of order, $F(1,240) = .212$, $p = .646$, and there was no significant interaction between condition and order, $F(2,240) = .144$, $p = .866$ (see Table 4).

A one-way ANOVA revealed a significant main effect of condition on performance errors, $F(5,240) = 19.87$, $p < .0001$. Tukey's HSD post-hoc test ($p < .05$) revealed that the 1000-deci condition was not significantly different from the deci-1000 condition, the 736-deci condition was not significant from the deci-736 condition, and the multi-scale-deci condition was not significantly different from the deci-multi-scale. In addition, the deci-736 condition was not significantly different from the multi-scale-deci condition, nor the deci-multiscale condition. However, the 736-deci condition was found to be significantly different from the deci-multi-scale condition. The primary finding was that both the 1000-deci and deci-1000 conditions were significantly different from the other four conditions (see Table 4).

Discussion

The pattern of data in Experiment 2 paralleled that of Experiment 1. Specifically, numerical representations that shared the same scale (i.e. decimal-1000 and 1000-decimal) allowed for efficient comparisons, despite the fact that the quantities were presented in different numerical formats⁸. In addition, the dual scale conditions led to the most inefficient comparisons. These findings confirm that numerical representations are scale specific.

Similar to Experiment 1, the multi-scale condition led to high intercepts and shallow slopes, which suggests a mathematical conversion strategy. However, the intercept of the multi-scale condition appears to be lower in Experiment 2 relative to Experiment 1. The gain in efficiency in the Experiment 2 multi-scale condition relative to the Experiment 1 multi-scale condition was likely the result of the reduced calculation requirements of the task. Recall that in Experiment 1 the scale of both stimuli changed on every trial, whereas in Experiment 2 one of the presented stimuli was always in the same scale (decimals). As such, in the Experiment 2 multi-scale condition, the participants could create and maintain a representation associated with decimals.

The mathematical conversion strategy used in Experiment 2 multi-scale condition was more efficient for *very* small numerical distances (those less than 0.1) than the direct analogue comparison strategy used in the Experiment 2 dual scale condition (see Figure 3). Given this efficiency gain, we must address the question, "why do participants choose not to adopt the mathematical conversion strategy in the dual scale task?" Although our data do not directly address this question, there is a likely explanation. Optimally, the participant should adopt the most efficient strategy on a trial-by-trial basis (e.g., Faulkenberry & Pierce, 2011).

⁸Although scale is relevant for relative frequencies, theoretically decimals are scale independent. However, one can read decimals as the position element in a relative frequency whereby the scale element is implied. In a two digit decimal, the implied scale element is 100. In a three digit decimal, the implied scale element is 1000. In Experiment 2, whereby participants were asked to compare a decimal to a relative frequency, it is not unlikely that the participants conceptualized the decimal as a relative frequency on a scale of 1000.

However, to select the most efficient strategy on a trial-by-trial basis, the participant must know the numerical distance between the stimuli (i.e., direct analogue comparison on most trials and a mathematical conversion strategy on very small numerical distances). Because the participant requires the strategy to extract the numerical distance, the participant cannot adopt a strategy on a trial-by-trial basis. Given this fact, the strategy that is *on average the most efficient* is the next most optimal strategy to adopt. The data suggest that, in both Experiment 1 and Experiment 2, a direct analogue comparison is the most efficient strategy, on average, to adopt in the dual scale condition (it is the most efficient, most of the time). Thus, the participants are, in fact, choosing their strategy optimally given the task requirements.

General Discussion

The results of Experiments 1 and 2 are consistent with the hypothesis that numerical representations of relative frequencies are (1) analogue and (2) scale specific. First, the numerical distance effect was present in all conditions. Second, the slope describing the relation between RT and numerical distance varied as a function of scale. Proportions presented as relative frequencies and/or decimals were compared equally efficiently, as long as they shared a scale, regardless of the numerical format. Proportions presented as relative frequencies and/or decimals were compared equally inefficiently when their scales differed, regardless of the numerical format. The simulation of the logarithmic model produces similar results. The multi-scale condition was the exception, yet was consistent with the rule. Here, multiple unpredictable scale conditions produced efficient comparisons after a likely mathematical conversion of the presented numbers into a common scale, again regardless of the numerical format.

Numerical Distance Effect in Proportions

Although few numerical judgment tasks have used numerical stimuli between zero and one, the ones that did (e.g., Cohen, 2010; Bonato et al., 2007) have found mixed results concerning the numerical distance effect. Some have found the numerical distance effect linked to the component integers rather than the proportions presented (e.g., Bonato et al., 2007). Others have found the numerical distance effect linked to the proportion represented by the integration of both components of a fraction (e.g., Jacob & Nieder, 2009; Ischebeck, et al., 2009). It appears that the numerical distance effect emerges for holistic perception of fractions when the perception of the fractions components is discouraged experimentally (e.g., Meert, Grégoire, Noel, 2008). The perception of the fraction components is discouraged experimentally by asking participants to compare fractions with varying numerators and denominators (equivalent to the scale position in a relative frequency).

The present experiment was the first to assess the numerical distance effect in relative frequencies. The pattern of data in Experiments 1 and 2 reveal that participants responded to the quantity represented by the integrated position and scale elements of the presented relative frequencies. Our data show the holistic perception of relative frequencies when the scale position is held constant and when the scale position is varied. This suggests that observers understand how to integrate the components of the relative frequencies more easily than fractions.

One reason why observers may be better able to interpret the relative frequencies than fractions is the structure of the two numerical formats. The structure of the fractions (i.e., one integer above a horizontal line above a second integer) provides little direct information to the viewer on how to integrate the two integers that compose the fraction. Indeed, even if the participant read the fraction as a mathematical formula (numerator divided by the denominator), the complexity of the calculation may have dissuaded him or her from

calculating it. The structure of relative frequencies (i.e., “x in y”), on the other hand, provides a cue how to interpret the two integers that compose it in a concrete manner. The word “in” cues the viewer that the position element is part of the scale element. In fact, the relative frequency can be read as a sentence involving abstract elements: five *things* in ten *things*, where *things* is implied. Alternatively, the relative frequency can be read as a sentence involving concrete elements: five *apples* in ten *apples*, where the participant substitutes *apples* for the implied *things*. Thus, the structure of relative frequencies can aid in their interpretation.

The hypothesis that observers understand how to integrate the elements of the relative frequencies better than the elements of fractions is supported by the work of Gigerenzer and his colleagues (Gigerenzer & Hoffrage, 1995). Unlike fractions, Gigerenzer and Hoffrage (1995) claim that humans have a natural understanding for frequencies. Natural frequencies are the likelihood of an event occurring in or during an organism’s life. The researchers hypothesize that through experience, humans become natural statisticians in their ability to determine whether an event is likely or unlikely to occur (Gigerenzer & Hoffrage, 1995).

The presence of the numerical distance effect has traditionally been interpreted as supporting the hypothesis that the underlying numerical representation is analogue (Dehaene, 1990; Dehaene, 1995; Hinrichs et al., 1981; Zhang & Wang, 2005). The results of Experiments 1 and 2 suggest that the numerical representation of relative frequencies contains information about the quantity represented by the integration of the components of the relative frequency, rather than the individual components themselves. The present data reveal the participants’ understanding and use of relative frequencies is different from either fractions or inconsistently rounded decimals.

Non-Abstract Representations

Experiments 1 and 2 reveal a robust relation between the slope of the numerical distance effect and the number of scales to be compared. The integration of the behavioral data and computer simulation lends insight into the structure of the underlying analogue representation(s). The pattern of data from the computer simulation of two quantities with low variance is similar to the single-scale behavioral data. The pattern of data from the computer simulation of two quantities with high variance is similar to the dual-scale behavioral data. One must address from where the extra variance in the system arises. The source of the increase in variance can arise either from (1) the comparison process, or (2) in the representations themselves. According to the abstract representation model, the comparison process acts on abstract representations and is constant across conditions. Therefore, if the extra variance arises in the comparison process, the abstract representation model is called into question. Similarly, the variance of an abstract representation for Scale A should not change as a function of what it is being compared to. That is, the transcoding of Scale A should be constant across conditions. Therefore, if the extra variance arises in the representations themselves, the abstract representation model is again called into question.

We hypothesize that the extra variance suggests that the representations of proportions are tied closely to the scale of the presented proportion. There are at least two possible ways this can occur: (1) each scale of the dual scale comparison has its own scale specific representation and the inefficient comparison is a feature of between representation comparisons, or (2) when dual scale comparisons are required, a unique representation is generated that accommodates both scales, but this representation is vague and indistinct relative to a single scale representation. Although our data cannot distinguish between these two conclusions, both conclusions rest on the foundation of multiple numeric representations. Thus, our findings strongly suggest that the numerical representation of numbers between zero and one are not abstract.

The scale dependent slopes revealed in the present experiments are inconsistent with both McCloskey's single abstract model (McCloskey et al., 1986) and Dehaene's (1992) triple code model. Both of these models hypothesize that numerical comparisons occur in a single abstract representation, regardless of numerical format, scale, etc. These models would therefore predict that the comparison process would be independent of the number of scales to be compared.

Our current data is more consistent with Multiple Representations models such as Gonzalez and Koler's (1982). Gonzalez and Koler hypothesize that representations are linked to numerical format. Our data reveal that representations for numbers between zero and one are linked more closely to scale than format (though we do not dispute that numerical format is relevant as well). Although relative frequencies are a single numerical format, our data show that relative frequencies are associated with multiple representations. Furthermore, when two numerical formats share a scale, such as decimals and relative frequencies with a scale of 1000, it appears that they share a single representation. The present data builds on the conclusions of Cohen et al. (2002), who concluded that the numerical representations of decimals and relative frequencies were separate. Here we show that this separation exists when the relative frequencies are on a different scale than the decimals.

Our data are also consistent with Cohen Kadosh and Walsh's (2009) hypothesized dual process theory of numerical representation. Cohen Kadosh and Walsh hypothesize that dual codes are present during numerical representation. There is an initial, automatic code that is tied to the numeric format. Later, there is a refinement of the automatic code that is influenced by intentionality, resources, task demands, etc. It is this later code that has the potential to take on abstract qualities. Because relative frequencies can take on any scale, it is unparsimonious to assume that the cognitive systems that process relative frequencies have pre-stored representations for every possible scale. It is, therefore, likely that the numerical representations of relative frequencies are generated, by the participant, "on the fly" during the second stage of Cohen Kadosh and Walsh's dual code theory (2009). That is, participants generate appropriate (or inappropriate) representations based on their understanding of the numerical system, the task demands, available resources, etc. If the proposed dual code hypothesis is accurate, then our data suggests that the second stage representations may require substantial cognitive resources to generate and maintain over time. A high cognitive load may explain why our participants were able to perform a direct analogue comparison in the dual scale conditions but they had to perform a mathematical conversion prior to comparison in the multi-scale condition.

Limitations

One of the primary limitations to numerical cognition research is the ability to extract the exact source of conditional RT variance. For example, in the dual-scale conditions, an additional variability arises, compared to the single-scale conditions. The current research is unable to specify whether the variability arises due to the comparison of two distinct and noisy abstract representations, the comparison of the quantities based on a single noisy representation that is task specific, or some third unspecified source. Similarly, in the multi-scale condition, we infer that the high intercept values, compared to the single- and dual-scale conditions, suggest a mathematical strategy is being applied to more efficiently compare the presented stimuli. However, more research needs to be completed in this area to clarify the specific strategies being used. Finally, similar to all research that relates to the numerical distance task, we assume a singular comparison process. It is possible that participants are using unspecified strategies that mimic the numerical distance effect.

It should also be understood that the current model assumes that the numerical distance effect arises from a comparison process, rather than a decision process. This is the

traditional understanding of the numerical distance effect (e.g. Dehaene, 1990; Dehaene, 1995; Hinrichs et al., 1981; Zhang & Wang, 2005). Nevertheless, Verguts et al. (2005) propose that the numerical distance effect arises from the decision process. It is important to note, however, that the Verguts model is more similar to the signal detection models described here than is generally acknowledged. That is, although Verguts and colleagues claim that the response stage is not a quantity representation, the response weights linking the place coding and comparison response codes are magnitudes that fall on a continuum. These magnitudes are highly correlated with the magnitude of the quantity denoted by the integer with which they are associated. Furthermore, more than one response weight is activated on every comparison trial, creating activation variance. As such, the response weights themselves carry the properties of the analogue quantity representation. Therefore, it is difficult to distinguish between the inherent nature of the response weights as implemented in the Verguts model and those of the analogue quantity representation the Verguts model is often contrasted with. Finally, it is unclear how the decision model described by Verguts can accommodate the slope and intercept variability data presented in Experiments 1 and 2 through a decision mechanism.

In sum, we conducted two numerical distance experiments to explore how numbers between zero and one are represented and three simulations to explore the abstract representation model of RT and the Welford function. Our data suggest that the numerical representations of numbers between zero and one represented as relative frequencies are (1) analogue, (2) non-abstract, and (3) scale dependent.

Acknowledgments

This work was supported by NIH grant RO1HD047796.

References

- Ashby F, Maddox W. A response time theory of separability and integrality in speeded classification. *Journal of Mathematical Psychology*. 1994; 38:423–466.
- Balakrishnan JD, Ratcliff R. Testing models of decision making using confidence ratings in classification. *Journal of Experimental Psychology: Human Perception and Performance*. 1996; 22:615–633. [PubMed: 8666956]
- Birnbaum M, Jou J. A theory of comparative response time and “difference” judgments. *Cognitive Psychology*. 1990; 22:184–210.
- Bonato M, Fabbri S, Umiltà C, Zorzi M. The mental representation of numerical fractions: real or integer? *Journal of Experimental Psychology: Human Perception and Performance*. 2007; 33(6): 1410–1419. [PubMed: 18085953]
- Buckley P, Gillman C. Comparisons of digits and dot patterns. *Journal of Experimental Psychology*. 1974; 103(6):1131–1136. [PubMed: 4457588]
- Carey S. Cognitive Foundations of Arithmetic Evolution and Ontogenesis. *Mind and Language*. 2001; 16(1):37–55.
- Cohen DJ. The role of physical similarity in “numerical distance” tasks with decimals and integers. *Journal of Experimental Psychology: Human Perception and Performance*. 2010; 36:1389–1398.
- Cohen Kadosh R, Tzelgov J, Henik A. A synesthetic walk on the mental number line: the size effect. *Cognition*. 2008; 106(1):548–557. [PubMed: 17275802]
- Cohen Kadosh R, Walsh V. Numerical representation in the parietal lobes: Abstract or not abstract? *Behavior and Brain Sciences*. 2009; 32:313–328.
- Cohen D, Ferrell J, Johnson N. What Very Small Numbers Mean. *Journal of Experimental Psychology: General*. 2002; 131(3):424–442. [PubMed: 12214756]

- Dehaene S. Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception and Performance*. 1990; 16(3):626–641. [PubMed: 2144576]
- Dehaene S. Varieties of numerical abilities. *Cognition*. 1992; 44(1-2):1–42. [PubMed: 1511583]
- Dehaene S. Electrophysiological evidence for category-specific word processing in the normal human brain. *NeuroReport*. 1995; 6:2153–2157. [PubMed: 8595192]
- Dehaene S, Akhavein R. Attention, Automaticity, and Levels of Representation in Number Processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1995; 21(2):314–326.
- Faulkenberry TJ, Pierce BH. Mental representations in fraction comparison: Holistic versus component-based strategies. *Experimental Psychology*. 2011; 58:480–489. [PubMed: 21592948]
- Gallistel C, Gelman R. Preverbal and verbal counting and computation. *Cognition*. 1992; 44:43–74. [PubMed: 1511586]
- Ganor-Stern D, Karasik-Rivkin I, Tzelgov J. Holistic representation of unit fractions. *Experimental Psychology*. 2011; 58:201–206. [PubMed: 21106472]
- Gigerenzer G, Hoffrage U. How to improve Bayesian Reasoning without instruction: frequency formats. *Psychological Review*. 1995; 102(4):684–704.
- Gonzalez E, Kolers P. Mental Manipulation of Arithmetic Symbols. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1982; 8(4):308–319.
- Hinrichs J, Yurko D, Hu J. Two Digit Number Comparison: Use of Place Information. *Journal of Experimental Psychology: Human Perception and Performance*. 1981; 7(4):890–901.
- Ischebeck A, Schocke M, Delazer M. The processing and representation of fractions within the brain: An fMRI investigation. *Neuroimage*. 2009; 47:403–413. [PubMed: 19328235]
- Jacob S, Nieder A. Notation-Independent Representation of Fractions in the Human Parietal Cortex. *The Journal of Neuroscience*. 2009; 29(14):4652–4657. [PubMed: 19357289]
- Kallai A, Tzelgov J. A Generalized Fraction: An Entity Smaller than One on the Mental Number line. *Journal of Experimental Psychology: Human Perception and Performance*. 2009; 35(6):1845–1864. [PubMed: 19968440]
- Link S. Modeling imageless thought: the relative judgment theory of numerical comparisons. *Journal of Mathematical Psychology*. 1990; 34:2–41.
- McCloskey M, Sokol S, Goodman R. Cognitive processes in verbal-number production: Inferences from the performance of brain damaged subjects. *Journal of Experimental Psychology: General*. 1986; 115(4):307–330. [PubMed: 2949043]
- Meert G, Gregoire J, Noel M. Rational Numbers: Componential versus holistic representation of fractions in a magnitude comparison task. *The Quarterly Journal of Experimental Psychology*. 2008; 62(8):1598–1616. [PubMed: 19123119]
- Moyer R, Landauer T. Time required for judgments of numerical inequality. *Nature*. 1967; 215(5109):1519–1520. [PubMed: 6052760]
- Ratcliff R. A theory of memory retrieval. *Psychological Review*. 1978; 85:59–108.
- Ratcliff R. Models for dealing with reaction time outliers. *Psychological Bulletin*. 1993; 114(3):510–532. [PubMed: 8272468]
- Schneider M, Siegler R. Representations of the Magnitudes of Fractions. *Journal of Experimental Psychology: Human Perception and Performance*. 2010; 36(5):1227–1238. [PubMed: 20873937]
- Siegler R, Thompson C, Schneider M. An integrated theory of whole number and fractions development. *Cognitive Psychology*. 2011; 62:273–296. [PubMed: 21569877]
- Takahashi A, Green D. Numerical judgments with Kanji and Kana. *Neuropsychologia*. 1983; 21(3):259–263. [PubMed: 6877579]
- Thomas E, Myers J. Implications of latency data for threshold and nonthreshold models of signal detection. *Journal of Mathematical Psychology*. 1972; 9:253–285.
- Verguts T, Fias W, Stevens M. A model of exact small-number representation. *Psychonomic Bulletin & Review*. 2005; 12(1):66–80. [PubMed: 15945201]
- Welford AT. The measurement of sensory-motor performance: Survey and reappraisals of twelve years' progress. *Ergonomics*. 1960; 3:189–230.

- Zakay D, Tuvia R. Choice latency times as determinants of post-decisional confidence. *Acta Psychologica*. 1998; 98:103–115.
- Zhang J, Wang H. The effect of external representations on numeric tasks. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*. 2005; 58A(5):817–838.

Figure 1. Simulated data for the Logarithmic constant variance (LogC), Linear constant variance (LinC), and Linear scalar variance (LinS) models as fit to the Welford function.

Figure 2.

A graphical representation of the average slope and intercept of the single, dual, and multiple scale conditions of Experiment 1.

Figure 3.

A graphical representation of the average slope and intercept of the single, dual, and multiple scale conditions of Experiment 2.

Table 1

Error rate by probe for each model by condition.

Model	Condition	Probe							
		0.1	0.2	0.3	0.4	0.6	0.7	0.8	0.9
LogC	Less Noisy Unbiased	0	0	<.01	0.06	0.07	0.01	0	<.01
	More Noisy Unbiased	0	0	0	0.06	0.05	0.01	<.01	0
	Less Noisy Unbiased	0	0	<.01	0.07	0.06	<.01	0	0
	Less Noisy Biased	0	<.01	<.01	0.04	0.09	0.01	0	0
LinC	More Noisy Biased	0	0	<.01	0.03	0.08	0.01	<.01	<.01
	Less Noisy Unbiased	0	0	<.01	0.03	0.1	0.02	<.01	<.01
	Less Noisy Biased	0	0	0	0.01	0.09	0.01	<.01	<.01
	More Noisy Biased	0	0	<.01	0.01	0.09	0.02	<.01	<.01

Table 2

The slope (SD), intercept (SD), percent overall error (SD), upper trim threshold in ms, lower trim threshold in ms, percent below the lower trim threshold, and overall mean in ms (SD), for each condition in Experiment 1.

	1000-1000	736-736	736-1000	1000-736	multi-scale
Slope (SD)	-0.09 (0.04)	-0.09 (0.04)	-0.16 (0.12)	-0.15 (0.12)	-0.08 (0.09)
Intercept (SD)	2.95 (0.12)	2.97 (0.13)	3.25 (0.22)	3.27 (0.24)	3.50 (0.18)
Overall Error (SD)	4.69 (3.81)	4.73 (3.85)	23.11 (16.58)	21.04 (18.58)	19.96 (15.77)
Upper Trim Threshold	3872	5084	17303	18439	21008
Lower Trim Threshold	450	500	550	550	300
% below lower trim threshold	0.87	1.38	12.24	13.04	0.76
Mean (SD)	1190 (564)	1339 (743)	3305 (2763)	3290 (2814)	4598 (3366)

Table 3

The average of participants' r^2 values from regressions predicting their raw $\log(RT)$ by the Log D and by the Welford function for each condition for Experiment 1.

	1000-1000	736-736	736-1000	1000-736	multi-scale
Log D	.051	.042	.054	.038	.019
Welford	.050	.039	.048	.035	.019

Table 4

The slope (SD), intercept (SD), percent overall error (SD), upper trim threshold in ms, lower trim threshold in ms, percent below the lower trim threshold, and overall mean in ms (SD), for each condition in Experiment 2.

	1000-deci	deci-1000	736-deci	deci-736	multi-scale-deci	deci-multi-scale
Slope (SD)	-0.09 (0.04)	-0.08 (0.04)	-0.16 (0.14)	-0.14 (0.12)	-0.09 (0.12)	-0.06 (0.07)
Intercept (SD)	3.04 (0.14)	3.06 (0.13)	3.26 (0.16)	3.27 (0.21)	3.36 (0.16)	3.35 (0.22)
Overall Error (SD)	2.46 (2.3)	2.37 (19)	20.63 (15.8)	19.61 (17.2)	13.63 (111)	13.46 (11.3)
Upper Trim Threshold	6378	6467	16168	15254	16978	15868
Lower Trim Threshold	N/A	N/A	600	650	750	850
% below lower trim threshold	N/A	N/A	4.17	6.46	3.44	3.68
Mean (SD)	1487 (890)	1530 (832)	3405 (2601)	3296 (2452)	3425 (2530)	3481 (2588)

Table 5

The average of participants' r^2 values from regressions predicting their raw $\log(\text{RT})$ by the Log D and by the Welford function for each condition for Experiment 2.

	1000-deci	deci-1000	736-deci	deci-736	multi-scale-deci	deci-multi-scale
Log D	.041	.039	.032	.026	.015	.005
Welford	.040	.037	.028	.024	.014	.004