Empirical Research

Individual Differences in Numerical Comparison Is Independent of Numerical Precision

Richard Prather*[a]

[a] University of Maryland, College Park, MD, USA.

Abstract

Numeracy, as measured by performance on the non-symbolic numerical comparison task, is a key construct in numerical and mathematical cognition. The current study examines individual variation in performance on the numerical comparison task. We contrast the hypothesis that performance on the numerical comparison task is primarily due to more accurate representations of numbers with the hypothesis that performance dependent on decision-making factors. We present data from two behavioral experiments and a mathematical model. In both behavioral experiments we measure the precision of participant's numerical value representation using a free response estimation task. Taken together, results suggest that individual variation in numerical comparison performance is not predicted by variation in the precision of participants' numerical value representation.

Keywords: estimation, non-symbolic, precision, modeling

Learners' performance on non-symbolic numerical comparison tasks is used to define the learner's *numeracy*, a key construct in research on numerical cognition and early math learning (e.g., Feigenson, Dehaene, Spelke, Feigenson, & Spelke, 2004; Libertus, Feigenson, & Halberda, 2013; Lukowski et al., 2017). Though the non-symbolic numerical comparison task plays a central role in the study of numerical cognition, there is not a comprehensive characterization of the processes involved in completing the task. Many variations of the numerical comparison task exist. The task typically involves a forced choice comparison between two numerical values that are displayed as a set of shapes on a screen. The participant then indicates which set is larger without counting. The non-symbolic numerical comparison task has been adapted for a wide range of participants including pre-literate children and infants (e.g., Halberda & Feigenson, 2008; Xu & Spelke, 2000), human populations without formal number systems (Pica, Lemer, Izard, & Dehaene, 2004), and non-human primates (Andreas Nieder & Merten, 2007; Brannon & Terrace, 2000).

Characterization of the cognitive processes involved in completing the numerical comparison task is essential for theory and application and will contribute to an explanation of variation in learner's numerical comparison ability. Additionally, it will help researchers to determine *why* performance on numerical comparison correlates with performance on other mathematical tasks and make principled predictions about this relation comes about.

The non-symbolic comparison task has been used as a predictor of mathematical outcomes. Children's performance on the numerical comparison task predicts performance on symbolic arithmetic tasks and standardized math assessments (Chen & Li, 2014; Gilmore, McCarthy, & Spelke, 2010) and training with non-symbolic comparison leads to improvement on symbolic tasks (Park & Brannon, 2013, 2014; Ramani, Jaeggi, Daubert, & Buschkuehl, 2017). Learners' performance on a non-symbolic comparison task predicts later performance on symbolic arithmetic tasks (Libertus, Feigenson, & Halberda, 2011) and later educational outcomes (Mazzocco, Feigenson, & Halberda, 2011). Since non-symbolic numerical comparison performance is associated with children's performance on other mathematical tasks a characterization of the cognitive processes may also be leveraged to develop improved interventions for poor mathematical performance.

Cognitive processes and representations that may contribute to performance include executive functioning and inhibition (Cragg & Gilmore, 2014; Gilmore et al., 2013; Gilmore, Keeble, Richardson, & Cragg, 2017), visuo-spatial processes (Crollen, Collignon, & Noël, 2017),the mental number line (Moeller, Neuburger, Kaufmann, Landerl, & Nuerk, 2009), neural tuning curves (Prather, 2014), decision making evidence accumulation (e.g., Purcell et al., 2010), amongst others. In the current study, we attempt to move towards a more comprehensive characterization of the processes involved while acknowledging that any progress made will undoubtedly still be incomplete and require further work.

The Current Study Approach

The current study includes two behavioral experiments and a mathematical model of the non-symbolic numerical comparison and non-symbolic numerical estimation tasks. Our approach assumes that participants have a mechanism for representing relative values that can be used in completing these numerical tasks. We focus on value representation that may be constructed from a combination of numerical and non-numerical information in order to make accurate and ecologically valid conclusions. Participants' use of non-numerical information on numerical tasks is supported by prior work (e.g., Cohen Kadosh, Cohen Kadosh, & Henik, 2008; Van Opstal & Verguts, 2013; Walsh, 2003), and participants' representation of value is not limited to numerical tasks. Any task in which assessing relative values is useful may involve value representation, such as decision-making tasks (e.g., Behrens, Woolrich, Walton, & Rushworth, 2007; Rangel, Camerer, & Montague, 2008; Sugrue, Corrado, & Newsome, 2005), and reward processing (e.g., Gottfried, O'Doherty, & Dolan, 2003; Silvetti, Seurinck, & Verguts, 2011). While there has been some interest in defining a "pure" number sense the primary concern in the current study is to evaluate how participants construct the relative values in completing number tasks regardless of the perceptual information used.

We evaluate evidence for two cognitive processes, precision of number representation and decision-making threshold, which may contribute to completing the numerical comparison task. Both are characterized at an algorithmic level of analysis (Marr, 1982) using a combination of behavioral experimentation and mathematical modeling. The goal is to create a reproducible formal model of variation in the cognitive processes and behavior relevant to numerical comparison. We consider the hypothesis that the *precision* of number value representation is the primary driver of individual variation in numerical comparison performance (Prather, 2014), where more precise representations are associated with better performance. Learners have some internal representation of number values, be it via neural tuning curves (e.g., Nieder, Freedman, & Miller, 2002; Prather, 2012) or an internal space-to-number mapping such as the mental number line (e.g., Siegler & Opfer, 2003), in which numbers are represented as relative spatial positions, similar to a physical number line.

Learners with a more precise representation of number values are better able to make distinctions between numerical stimuli and answer correctly on a number comparison task (e.g., Prather, 2014). We model number representation as neural tuning curves associated with numbers as reported in both non-human primates and humans (e.g., Moskaleva & Nieder, 2014; Nieder & Dehaene, 2009). The precision hypothesis is consistent with prior mathematical modeling work that demonstrates how increases in the precision of neural coding are associated with improved performance on numerical tasks (DeWind & Brannon, 2012; Prather, 2012).

We also consider the hypothesis that individual differences in numerical comparison task performance are primarily due to variation in *decision*-making threshold, independent of numerical representation. Variations in thresholds for evidence accumulation contribute to performance in perceptual decision-making tasks (Busemeyer & Townsend, 1993; Pleskac & Busemeyer, 2010; Purcell et al., 2010). The numerical comparison task can be framed simply as a version of perceptual decision-making task in which numerical information is relevant.

We do not assume the Precision and Decision processes to be mutually exclusive. The current study evaluates the degree to which these two processes account for behavioral data across two numerical tasks. The behavioral experiments examine the relation between participants' accuracy and precision of number estimation as it relates to numerical comparison (e.g., Libertus et al., 2016). We draw on recent work that focuses on comparisons between performance on non-symbolic numerical comparison and free response non-symbolic estimations (Castronovo & Göbel, 2012; Chesney, Bjalkebring, & Peters, 2015; Guillaume, Gevers, & Content, 2016; Libertus, Odic, Feigenson, & Halberda, 2016). For the free response estimation task, participants are shown a set of objects and asked to estimate how many there are.

Across two experiments we combine behavioral and modeling data to examine the possibility that variation in numerical comparison performance is driven primarily by individual differences in the precision of numerical representations. We also consider the possibility that variation in numerical comparison is primarily driven by variation in the decision-making processes and not representations of number value. There is mixed evidence in prior work regarding the relationship between numerical comparison and estimation task performance. In some cases, no relationship between numerical comparison accuracy and estimation accuracy is found (Guillaume et al., 2016; Pinheiro-Chagas et al., 2014), in others a small positive correlation was reported (Chesney et al., 2015). In other studies a significant relationship between estimation variability and number comparison was found, but not between estimation accuracy and number comparison (Libertus et al., 2016).

In addition to estimation accuracy, we calculated the variation of participants' estimates, e.g., the precision (Izard & Dehaene, 2008). A participant could have very precise estimates while being overall inaccurate. That participant would perform poorly on free response estimation accuracy but perform well on numerical comparison. Such as a participant who tends to estimate 50 dots as about 150 dots has high precision but low accuracy on the estimation task. Precise, but not necessarily accurate performance, should primarily rely on the precision of number value representation. If a participant overestimated values on the estimation task, that participant could still perform well in numerical comparison. For any two values, 20 and 22, increasing both values to 30 and 35, does not necessarily change their relative values. Participants' estimation error need not be a constant proportion for estimation precision and comparison performance to be unrelated.

If the variation in participants' performance is primarily due to variation in the precision of their numerical representations we expect a strong correlation between accuracy on the comparison task and estimation precision in

the estimation task. If individual differences are due to variation in decision-making thresholds, we would not expect a significant correlation between accuracy on the comparison task and estimation precision. Numerical representation precision may not be all that there is to the estimation task or comparison task. The participant must map their internal representation of the stimuli to an output. In the case of the estimation task, the output is a specific cardinal value. For each participant, the model will fit their performance on the comparison and estimation tasks simultaneously. The question to be examined is if how well adjustments to the neural representation precisions fit participants' data relative to the fit when decision-making parameters are also adjusted. Does including a decision-making evidence parameter significantly improve model fit to participant's data one or both tasks?

Behavioral Experiments

Experiment 1

Method

Participants — Participants (N = 71) were adults (age range from 19 to 70, median 32) recruited online through Amazon Mechanical Turk. Protocols were approved by a university Internal Review Board.

Numerical comparison task — Stimuli were 96 visually presented pairs of square arrays with a midline separator. Shape arrays ranged in number from 23 to 111 (see the Appendix). The difference between the two values being compared ranged from a ratio of 1.05 to 1.36. Stimuli were balanced for total area, and size of the largest square. The location of the squares was randomly selected before the experiment. Participants were instructed to indicate which side contained more shapes via button press. Stimuli were displayed for 2 seconds after which the screen was blank. There was no response time limit; participants were instructed to respond as quickly as possible.

Free response estimation — Stimuli were 64 visually presented shape arrays. The number of objects ranged from 23 to 111 (see Appendix). Participants were instructed to respond with an estimate of how many shapes were in the display. The stimuli were displayed for 2 seconds after which the participants were presented with a prompt to type their response. There was no response time limit.

Results

Performance on the numerical comparison task — Participants' performance was calculated as the number of correct responses on the task. Performance ranged from 32% to 84% correct. For the remaining analysis, we only consider participants with performance statistically above chance (58%) on the numerical comparison task (n = 53). For this subset of participants' median performance on the task is 69% correct. Age ranged from 19 to 70 with a median of 31. Given the regression analysis to be performed 53 participants is sufficient for the expected medium effect size (d = 0.38, power = 0.81).

Free response estimation — Participants' performance was calculated using the deviations between the participants' response and the actual number of shapes displayed. Participants' mean deviation ranged from 7.78 to 44.51 with a median of 17.33. Deviations can also be calculated in terms of proportions (e.g., a response of

50 when 40 items were displayed would be a deviation of 0.25). Participants' mean deviation in terms of proportion difference ranged from 0.125 to 0.811 with a median of 0.255.

We also calculated the variation in participants' responses, separately from accuracy (Figure 1). The stimuli in this task included multiple stimuli with the same number of objects in different configurations. This allows us to evaluate how consistent participants' estimations for the eight target values. Participant's precision score was the average coefficient of variation across the eight target values. Participants' precision scores ranged from 0.09 to 0.42 with a median of 0.17. Precision and accuracy scores were not significantly correlated where increased precision was associated with higher accuracy, $r(51) = .21$, $p = .13$.

The relationship between numerical comparison and estimation tasks — We evaluated the relationship between participants' behavior on the two tasks using a linear regression with estimation deviation (e.g., accuracy), estimation precision, and participant age in predicting numerical comparison (Figure 2). Neither estimation deviation nor estimation precision significantly predicted numerical comparison score (Table 1).

Table 1

Results for a Linear Regression Using Participant Scores for Estimation Accuracy and Estimation Precision to Predict Numerical Comparison Task Score

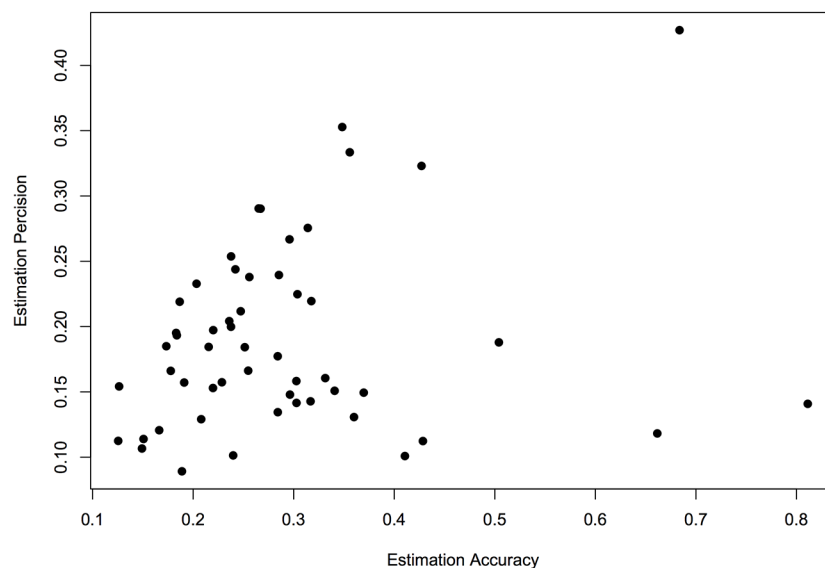| Independent Variable	B	95% CI	t(45)	p	d
Estimation Accuracy	-0.04	[-0.03, 0.04]	0.32	.753	0.15
Estimation Precision	0.14	[-0.85, 0.53]	0.44	.647	0.06
Age	0.58	[-0.21, 1.38]	1.47	.147	0.11
Accuracy*Precision	-0.01	[-0.03, 0.01]	0.85	.397	0.28
Precision*Age	-0.82	[-2.09, 0.45]	1.29	.201	0.48
Accuracy*Age	-0.03	[-0.08, 0.01]	1.67	.101	0.80
Accuracy*Precision*Age	0.02	[-0.04, 0.09]	0.77	.440	0.23

Figure 1. Graph of participants' performance on the estimation task by the Precision and Accuracy measures.

Note. Each dot represents one participant.

The data suggest that there is not a strong relationship between non-symbolic numerical comparison performance and the precision of participants' numerical representations. A Bayes factor analysis suggests evidence for the null hypothesis, BF = 0.126 for the regression model with Estimation Accuracy and Precision as predictors.
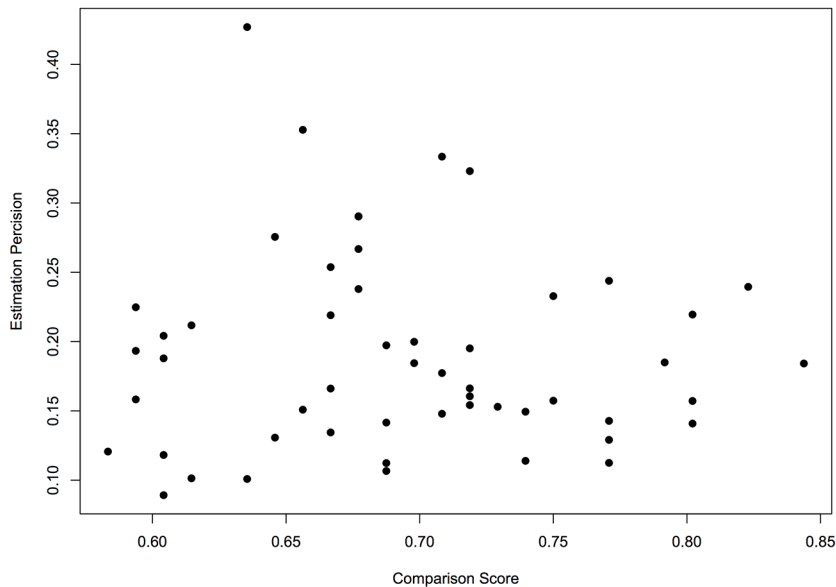

Figure 2. Scatter plot of participants' scores on the numerical comparison task (ANS Score) and their free estimation task precision score.

Note. Estimation precision was calculated as the mean variation in estimation for the target value expressed as a ratio of that value. Larger values represent less consistent estimation responses.

Experiment 2

Method

Participants — Participants (N = 30, 17 male) were children aged 7 years to 8 years, 9 months. Parents of the children were recruited through a university participant pool. Protocols were approved by the University of Internal Review Board. Participants' caregivers were informed of any risks in the study and an age appropriate assent protocol was used for the children.

Participants completed three tasks during the experimental session; the Numerical Comparison, Free Response Estimation and the *Test of Early Mathematics Ability, 3rd Edition* (Ginsburg & Baroody, 2003).

Numerical comparison task — Stimuli were 90 visually presented pairs of shape arrays with a midline separator. Shape arrays ranged in number from 23 to 111 (see Appendix). The difference between the two values being compared ranged from a ratio of 1.05 to 1.85. Participants were instructed to indicate which side contained more shapes via button press. Stimuli were displayed for 2 seconds after which the screen was blank. There was no response time limit; participants were instructed to respond as quickly as possible while being as accurate as possible. No feedback was given, and there were no practice trials. Participants completed all 90 comparisons. Stimuli were constructed to control for the overall area of presented shapes.

Free response estimation — Stimuli were 40 visually presented shape arrays. Arrays included randomly places black squares of varying sizes. The number of objects ranged from 23 to 111 (see Appendix). Participants were instructed to respond with an estimate of how many shapes were in the display. The stimuli were displayed for 2 seconds after which the participants were presented with a prompt to type their response. There was no response time limit.

Mathematical ability — Participants completed the *Test of Early Mathematics Ability* 3rd Edition (TEMA), a standardized early mathematics assessment (Ginsburg & Baroody, 2003). The TEMA is designed to assess children's overall mathematical knowledge including formal and informal mathematics.

Results

Performance on the numerical comparison task — Participant's performance on the numerical comparison task ranged from 60% to 94% correct with a median of 80%.

Performance on the estimation task — We eliminated any trial for which participants did not make a response, representing 4% of trials. We also eliminated responses that represented the top 5% of estimates as many of these responses appeared to by types, e.g., '500000'. Participants' performance on the estimation task was calculated in the same manner as in Experiment 1. Accuracy was calculated by taking the absolute value of the difference between the target value and participants' given estimate and dividing by the target value. This gives us a ratio-difference score, e.g., an estimation of 13 for the target value 10 would produce a score of 0.3. Participant accuracy ranged from 0.30 to 1.06 with a median of 0.60.

We also calculated the variation in participants' responses, separately from accuracy. Participants' precision score calculation was the same as described in Experiment 1. Participants' precision score ranged from 0.29 to 0.84 with a median of 0.56. Precision and accuracy scores significantly correlated where increased precision was associated with higher accuracy, $r(28) = .57$, $p < .001$ (Figure 3).
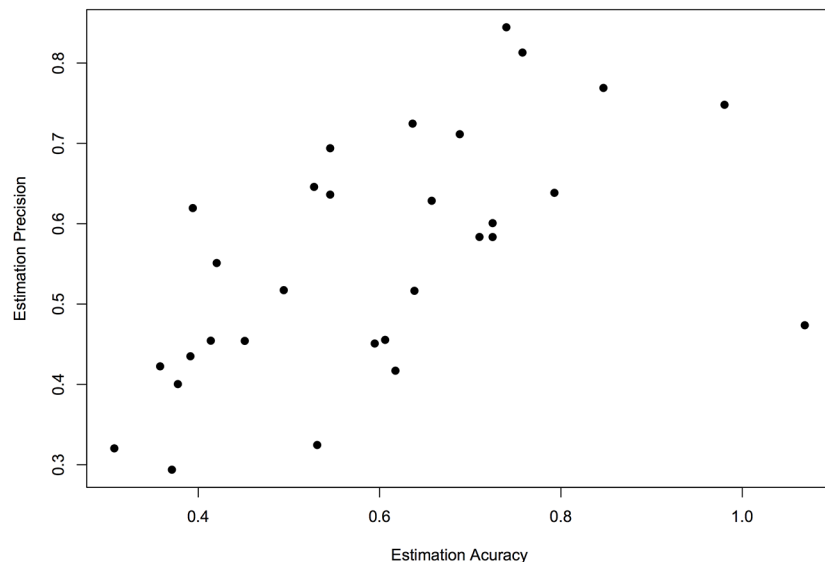

Figure 3. Graph of participants performance on the Estimation task by the Precision and Accuracy measures.

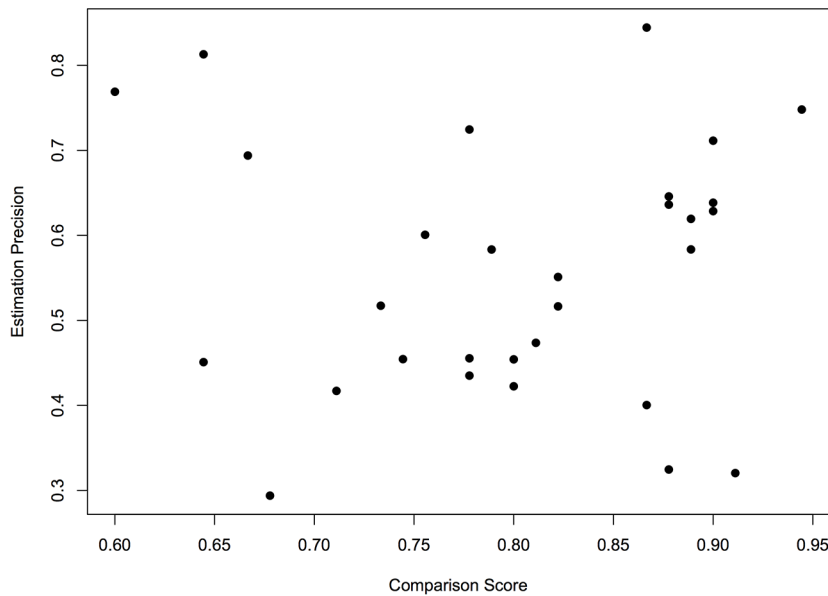*Note.* Each dot represents one participant.

Figure 4. Scatter plot of participants' scores on the numerical comparison task (ANS Score) and their free estimation task precision score.

Note. Estimation precision was calculated as the mean variation in estimation for the target value expressed as a ratio of that value. Larger values represent less consistent estimation responses.

Performance on TEMA — Participant's performance on the TEMA was calculated using the scoring instructions. Participant's scores ranged from 85 to 132 with a median of 114.

Relationship between tasks — We conducted a linear regression to predict participants' numerical comparison score (arcsine transformation of the proportion of correct responses) using estimation accuracy, estimation precision and age and TEMA score as predictors. Given the regression analysis to be performed 30 participants is sufficient for a large effect size (d = .70, power = 0.74). We found no significant predictors of numerical comparison score (see Table 2). A bivariate correlation between numerical comparison task and estimation accuracy score was non-significant, $r(28)$ = .12, p = .52. A correlation between numerical comparison task and estimation precision was also non-significant $r(28)$ = .02, p = .91 (see Figure 4). A Bayes factor analysis suggests evidence for the null hypothesis, BF = 0.072 for the regression model with Estimation Accuracy, Precision, TEMA score and age as predictors.

Table 2

Results Based on a Linear Regression Using Participant Scores for Estimation Accuracy, Estimation Precision, TEMA Score and Participant Age to Predict Numerical Comparison Task Score

Independent Variable	B	95% CI	$t(25)$	p	d
Estimation Accuracy	-0.02	[-0.35, 0.29]	0.17	.86	0.06
Estimation Precision	0.06	[-0.32, 0.46]	0.35	.73	0.04
TEMA score	0.0002	[-0.004, 0.005]	0.10	.91	0.04
Age	-0.04	[-0.13, 0.05]	0.86	.39	0.34

Comparisons between adult and child participants — We found that adults' performance on the estimation task was significantly different in terms of accuracy (M = 0.289) compared to child (M = 0.60), $t(81)$ = 8.67, d =

1.45, *p* < .001. Adults estimation task precision score (*M* = 0.19) was also significantly different than children's (*M* = 0.56), *t*(81) = 15.21, *d* = 3.21 *p* < .001. Adults did not show a significant correlation between estimation accuracy and precision, while child participants did. There are two possible explanations; the child participant data is a Type 1 error due in part to the lower sample size, there is a developmental change in the relationship between estimation accuracy and precision.

Behavioral Experiments Discussion

For both Experiments 1 and 2 we find no statistically significant relationship between participants' performance on the non-symbolic numerical comparison task and estimation task scores. The behavioral data reported in data sets for Experiments 1 and 2 rely on the interpretation of a null effect, thus we used a Bayes Factor approach. While the two experiments had a range of participant ages, we found the same pattern across both analyses for adults and children. In both cases, Bayes factor values suggest evidence for the null effect when compared to the tested regression models. We also do not find that Experiment 2 participants' TEMA scores significantly predicted numerical comparison scores, despite prior evidence of a connection (Schneider et al., 2017). The relatively small age range used in this experiment (7.0 to 8.75 yrs) may affect the measured relationship.

Current study results differ from prior work which reported a significant relationship between estimation variability and number comparison but not estimation accuracy and number comparison (Libertus et al., 2016). Several differences between the studies may contribute to the difference in results. The range of numbers used here for estimation is larger than prior work. Estimation stimuli range was up to 111 whereas prior work was limited to no more than 20 (Libertus et al., 2016). The participant age range was both older and broader than the 5 to 8 years of prior work. The current study combines data from participants 7 to 9 years old for Experiment 2 and 25 to 70 years old for Experiment 1.

We interpret the results of this experiment as inconsistent with the Precision hypothesis.

The lack of significant correlation between estimation precision and numerical comparison suggests that numerical representation precision is not the *primary* driver of behavior. We consider the aforementioned alternative hypothesis, that general decision-making processes not specifically tied to number primarily drive numerical comparison performance. We elaborate on this potential process using a mathematical model in Experiment 2.

Mathematical Modeling

The purpose of the modeling experiment is to demonstrate how well the processes proposed by the Precision and Decision hypotheses fit the behavioral data from both Experiments 1 and 2. We evaluated the Precision hypothesis and the alternative Decision hypotheses using a dynamic neural field model (e.g., McClelland et al., 2010). We evaluate the two hypotheses using two versions of the same model. One model condition is designed to implement the Precision hypothesis; the other model implements the Decision hypothesis. Both model conditions were fit to each participant's behavioral data independently. Model optimization was implemented via an evolutionary algorithm that minimized the deviation between behavioral data and model output. The optimization procedure included adjustments to a subset of model specifications while others remained fixed. For

the Precision hypothesis model, the specification for the width of the tuning curves was variable, while specifications for the decision layer did not vary. For the Decision hypothesis model, the tuning curve widths are fixed while the intra-layer timing of evidence accumulation within the decision layer varies. This also changes the accuracy in detecting differences from the neural turning curves that connect to the decision layer.

An important point here is that the current model is much more strict than prior models of numerical comparison (Prather, 2014). As opposed to modeling participants' performance on only the numerical comparison task, the current model must simultaneously predict behavior on numerical comparison and estimation tasks for each participant.

Method

Model Specifications and Procedure

The model was implemented using MATLAB (MathWorks). The architecture was a multilayered dynamic systems model (e.g., Simmering & Perone, 2013; Spencer, Smith, & Thelen, 2001). Layers included two perceptual neural tuning curves and a decision layer. Perceptual layers modeled neural tuning curves associated with neural coding of stimuli (e.g., Prather, 2012, 2014; Tudusciuc & Nieder, 2007). For the numerical comparison task, the external inputs for the model were the two numerical values to be compared, taken from the stimuli in Experiment 1. The two values were represented by proportionally scaled Gaussian curves that reproduce the ratio dependent distance effect. Perceptual layers of the model reproduced the stimuli while activity was forwarded to the decision layer. The internal decision layer connections were specified to produce competition within the layer through lateral inhibition and self-excitation. Thus the two perceptual layers output created competition within the decision layer. This dynamic corresponded to the "decision" which was the index of the first stable activation peak in the decision layer. For the estimation task, the external input for the model was the target value to be estimated, taken from the stimuli in Experiment 1.

Each trial was comprised of 600 time-steps, which was selected to be large enough for activity from the input layers to create a decision in the output layer. Decisions were defined as when the decision layer produced a steady peak (activity with a peak value at the same layer index for 10 straight time-steps). The time-step of the decision was converted to the predicted reaction time of the decision. Thus on trials in which the model predicted a fast decision the steady peak was reached a relatively low time-step. On trials in which the model predicted a slower decision, the steady peak was reached on a higher time-step.

Model instantiations were fit to behavioral data from Experiment 2 using an evolutionary optimization algorithm. Model instantiations were completed in batches of 10, each corresponding to a generation. For each generation, the model instantiations were ranked based on their deviation from the behavioral data. Model instantiations with smaller deviations, smaller error, were ranked higher. For each generation instantiations ranked, 1–2 were moved forward as is to the next generation. Instantiations ranked 3–5 were 'mutated' by adjusting the specifications by a small random amount. Instantiations ranked 6–10 were discarded. Thus each generation included 5 new instantiations were randomly generated specifications, 3 'mutated' instantiations and 2 instantiations carried over from the previous generation. The specifications of the evolutionary algorithm were selected to maximize the efficiency of the algorithm to keep the number of batches needed relatively low.

The same process was employed for modeling behavioral data from Experiment 1 (adult participants) and Experiment 2 (child participants).

Results

Experiment 1 Model

What are the model conditions performances on the tasks? — The precision condition model instantiations performance on the numerical comparison task ranged from 0.46 to 0.72 with a median of 0.62. Performance on the estimation task, in terms of average proportional deviation from the target, ranged from 0.08 to 0.27 with a median of 0.15. We found that performance on the comparison task correlated with the neural tuning curve width $r(69) = -.37$, $p < .01$, where smaller tuning curve widths were associated with higher scores on the task. We found that performance on the estimation task correlated with the neural tuning curve width, $r(69) = .36$, $p < .01$, where smaller tuning curve widths were associated with better performance on the task.

The decision condition model instantiations performance on the numerical comparison task ranged from 0.50 to 0.86 with a median of 0.65. Performance on the estimation task, in terms of average proportional deviation from the target, ranged from 0.08 to 0.23 with a median of 0.14. We found that performance on the comparison task correlated with the neural tuning curve width $r(69) = -.27$, $p = .02$, where smaller tuning curve widths were associated with higher scores on the task. Performance on the comparison task was significantly correlated with the evidence rate parameter, $r(69) = .67$, $p < .01$. We found that performance on the estimation task was not correlated with the neural tuning curve width, $r(69) = .11$, $p = .36$. Performance on the estimation task was not significantly correlated with the evidence rate parameter, $r(69) = .08$, $p = .50$.

How well does each model fit participants' data? — Model data was evaluated using a similar analysis to the behavioral data. Each model version produced independent simulations of the numerical comparison and estimation task. We compared results for the Precision condition models ($n = 71$) to the Decision condition models ($n = 71$). For the Precision condition models, the median numerical comparison error was 0.02, with a range from 0.18 to 0.0. The median estimation error was 11.5 with a range from 5 to 41. For the Decision condition models, the median numerical comparison error was 0.01, with a range from 0.17 to 0. The median estimation error was 12 with a range from 6 to 39.

How does the model fit compare between precision and decision conditions? — To compare model fit for Precision and Decision conditions we compared the deviation from human data for both tasks. The model error for the numerical comparison task was significantly lower for the Decision condition models (median = 0.01) than for the Precision condition models (median = 0.2), $t(70) = 5.41$, $p < .001$, $d = 0.70$. For the estimation task the model error not significantly different between the Decision condition models (median = 11.5) and the Precision condition models (median = 12), $t(70) = 1.33$, $p = .19$, $d = 0.049$.

We calculated overall model error by combining numerical comparison and estimation error amounts. The overall model error was calculated as $Error_{Comparison} + Error_{Estimation} / 100$. The equation was created to equally weight error on both tasks. The overall model error for Decision condition models (median = 0.14) was significantly lower than overall model error for Precision condition models (median = 0.16), $t(70) = 5.53$, $p < .001$, $d = 0.32$ (Figure 5).
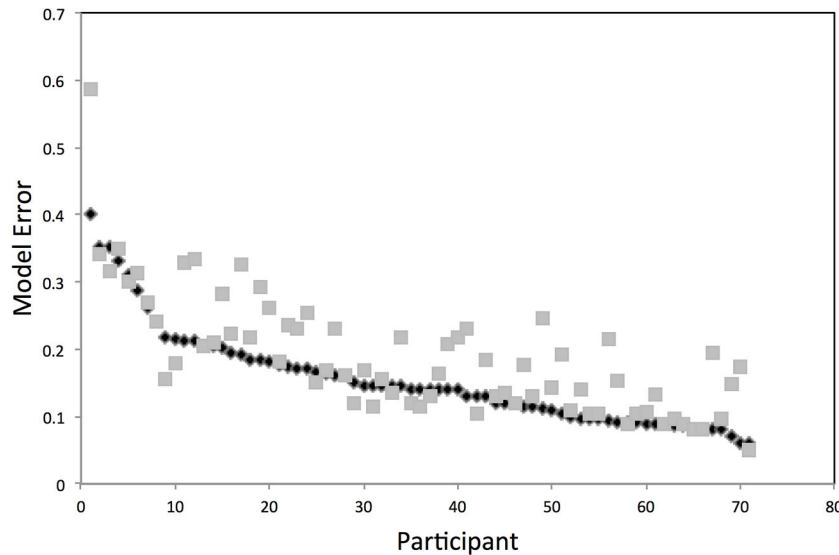
Figure 5. Total Error for Precision (grey-square) and Decision (black-diamond) model instantiations.

Note. The vertical axis represents the calculated error for each model instantiation. The horizontal axis represents the individual human participants (n = 71). The order is sorted by difference between Precision and Decision model performances.

These results show that the Decision condition models are better able to fit adult participants data for the numerical comparison task. Fit to participant data for the estimation task was equivalent. This suggests that the additional decision layer specification was only relevant to model fit to numerical comparison task and that it leads to superior model fit compared to the use of neural tuning curve precision.

Experiment 2 Model

What Are the Model Conditions Performances on the Tasks? — The precision condition model instantiations performance on the numerical comparison task ranged from 0.58 to 0.72 with a median of 0.67. Performance on the estimation task, in terms of average proportional deviation from the target, ranged from 0.25 to 0.08 with a median of 0.13. We found that performance on the comparison task did not significantly correlate with the neural tuning curve width, $r(28)$ = -.23, p = .22, where smaller tuning curve widths were associated with higher scores on the task. We found that performance on the estimation task did significantly correlated with the neural tuning curve width, $r(28)$ = .75, p < .01, where smaller tuning curve widths were associated with better performance on the task.

The decision condition model instantiations performance on the numerical comparison task ranged from 0.57 to 0.93 with a median of 0.77. Performance on the estimation task, in terms of average proportional deviation from the target, ranged from 0.08 to 0.22 with a median of 0.12. We found that performance on the comparison task not significantly correlated with the neural tuning curve width $r(28)$ = -.25, p = .18, where smaller tuning curve widths were associated with higher scores on the task. Performance on the comparison task was significantly correlated with the evidence rate parameter, $r(28)$ = .69, p < .01. We found that performance on the estimation task correlated with the neural tuning curve width, $r(28)$ = .42, p = .02, where smaller tuning curve widths were associated with better performance on the task. Performance on the estimation task was not significantly correlated with the evidence rate parameter, $r(28)$ = .14, p = .46.

How well does each model fit participants' data? — Model data was evaluated using a similar analysis to the behavioral data. Each model version produced independent simulations of the numerical comparison and estimation task. We compared results for the Precision condition models (n = 30) to the Decision condition models (n = 30). For the Precision condition models, the median numerical comparison error was 0.15, with a range from 0.26 to 0.01. The median estimation error was 26 with a range from 14 to 42. For the Decision condition models, the median numerical comparison error was 0.03, with a range from 0.15 to 0. The median estimation error was 25 with a range from 9.5 to 44.5.

How does the model fit compare between precision and decision conditions? — To compare model fit for Precision and Decision conditions we compared the deviation from human data for both tasks. The model error for the numerical comparison task was significantly lower for the Decision condition models (median = 0.03) than for the Precision condition models (median = 0.15), t(29) = 7.78, p < .001, d = 1.76. For the estimation task the model error not significantly different between the Decision condition models (median = 25) and the Precision condition models (median = 26), t(29) = 1.63, p = .11, d = 0.37.

We calculated the overall model error by combining numerical comparison and estimation error amounts. The overall model error was calculated as $Error_{Comparison} + Error_{Estimation} / 100$. The equation was created to equally weight error on both tasks. The overall model error for Decision condition models (median = 0.28) was significantly lower than overall model error for Precision condition models (median = 0.39), t(29) = 4.17, p < .001, d = 1.01 (Figure 6).
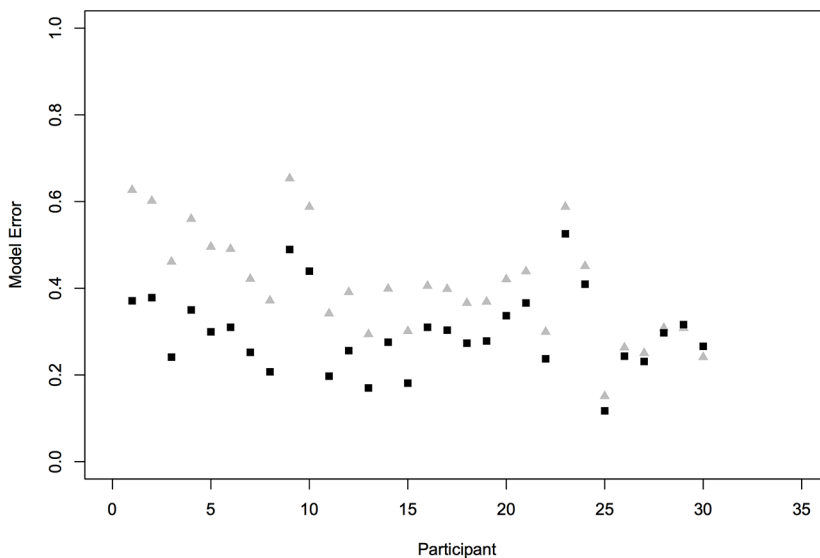

Figure 6. Total Error for Precision (grey-triangle) and Decision (black-squares) model instantiations.

Note. The vertical axis represents the calculated error for each model instantiation. The horizontal axis represents the individual human participants (n = 30). The order is sorted by difference between Precision and Decision model performances.

These results show that the Decision condition models are better able to fit participants data for the numerical comparison task. Fit to participant data for the estimation task was equivalent. This suggests that the additional

decision layer specification was only relevant to model fit to numerical comparison task and that it leads to superior model fit compared to the use of neural tuning curve precision.

Mathematical Modeling Discussion

The mathematical modeling results demonstrate that the Decision model instantiations fit both adult's and children's data significantly closer than the Precision model instantiations. Put more generally, a mathematical model that includes specifications for both numerical representation and decision-making is a better fit to human data than a model that only includes numerical representation. The modeling results suggest that behavioral data reported in behavioral Experiments 1 and 2 cannot be well-characterized using only neural tuning curve precision. This is in contrast with the apparent success of using neural tuning curves to model numerical comparison task (Prather, 2014) or the number-line estimation task (Prather, 2012). The current modeling study is different in two crucial respects; first, we model individual participant data, not group means; second performance on multiple tasks are modeled *simultaneously*. The fit of tuning curve precision models to behavioral data seems to be limited given these two considerations. The experiment demonstrates that the addition of a decision-making parameter allows for a far more accurate fit to participants' data. This suggests that neural tuning curve precision may be a necessary but not sufficient part of modeling the cognitive processes involved in completing numerical comparison and number-line estimation tasks.

General Discussion

The current study evaluated two models of the processes involved in comparing non-symbolic numbers. Results from both empirical and mathematical experiments are inconsistent with the hypothesis that numerical comparison performance is better characterized by variation in neural tuning curve precision. We find that participant's performance on free response estimation, used as an estimate of tuning curve precision, does not correlate with numerical comparison performance. Mathematical modeling results demonstrate that variation in the decision-making process can better account for participants' numerical comparison scores above and beyond variations in neural tuning curve precision. We interpret these results as inconsistent with the Precision hypothesis. Individual variation in performance on the numerical comparison task is not primarily due to variation in tuning curve precision.

The current results provide important evidence regarding the processes involved in non-symbolic numerical comparison. The current and recent results suggest that numerical representation precision does not play the primary role in the numerical comparison task. This contradicts some previous speculation about the role of neural tuning curve precision in numeral tasks (Prather, 2012, 2014). Of course, the precision and decision hypothesis are not mutually exclusive. We expect many factors relating to attention or inhibition may in part account for behavior on the comparison task. It is also possible that the processes involved in numerical comparison can change with experience or development.

If the individual variation in numerical comparison accuracy is due to decision-making more so than number representation what does that tell us? The importance of decision-making in numerical comparison may be informative in the design of interventions to improve learner's performance on numerical tasks. Individual variation in numerical decision-making may contribute to the association between numerical comparison skill and

general mathematical skill. Learners' skill at numerical decision-making may contribute to performance in a wide range of numerical and arithmetic task.

If learners' performance on the numerical comparison task can be characterized without invoking their representations of number values it calls into question the source of the correlation between numerical comparison skill and later arithmetic skills. Recent meta-analysis show mixed evidence that numerical comparison skill, in and of itself, predicts later performance (Chen & Li, 2014; Gilmore et al., 2010). It is possible that these correlations capture variations in domain general skills that happen to be involved in completing the task, such as inhibition (e.g., Gilmore et al., 2013; Purpura & Simms, 2018).

How do the numerical comparison measures used here relate to other work? The non-symbolic numerical comparison task has varying relationships to other measures depending on the details of the task (e.g., Dietrich, Huber, & Nuerk, 2015). The stimuli in the current study were controlled for item size but not item density of the display. This is not the same set up as some other studies (e.g., Panamath; Halberda, Mazzocco, & Feigenson, 2008). Of course, there is evidence that precisely controlling for non-numerical cues may be somewhat beside the point. Participants develop an internal representation of the numerical values of the stimulus that may be informed in part by density, area, perimeter, or convex hull. The point here is that the accuracy of such comparisons does not have a significant relationship with the precision of the representations of the same stimuli. We are concerned with the relationship between individual learner's behavior on these tasks and what that may say about the cognitive processes involved. Other work has even challenged if numerical comparison can be thought of a purely numerical task regardless of the controls employed (Gilmore, Attridge, & Inglis, 2011; Smets, Gebuis, Defever, & Reynvoet, 2014).

Potential Limitations

The child participant data lower sample size may contribute to a possible Type 1 error. It is also possible that adult and child participant results vary because of developmental changes in the relationship between estimation accuracy and precision. Given the scope of the current data we suggest caution in interpreting differences between the adult and child participants.

Reliability of measures calculated using a split-half Spearman correlation. For the data in Experiment 1, we calculated the Spearman coefficient using split half as $r = .70$. This reliability level is similar to what was reported in Chesney et al. (2015), $r = .74$. This suggests an acceptable level of reliability for the current measures. For the estimation task, we can calculate what the confidence interval for the measure of the participants' standard deviation (SD), which is used in calculating their estimation precision score. With 64 trials for the estimation task the 95% confidence interval for the SD [$0.85*SD$, $1.21*SD$].

Other models of decision-making such as drift diffusion models are fairly successful for weighing evidence in two-alternative decision-making (Park & Starns, 2015; Pirrone, Marshall, & Stafford, 2017; Purcell et al., 2010). The current approach does not contradict a drift diffusion model; there are some similarities in implementation. The model implementation of decision-making is comparable to the drift diffusion approach. The current model's implementation of evidence accumulation is the adjustment of thresholds for competition between two potential choices. Though the mathematics of the model implementations differs, we do not see the models as in conflict with each other. However, the current approach allows for a model implementation that can be applied to a two alternative forced choice task and a free response estimation task simultaneously. A considerable mo-

tivation of the use of a dynamic systems model is the potential to be broadly applied to behavioral and neural data for a variety of tasks. It is unclear how to adopt a drift-diffusion model, typically used for two alternative forced choice tasks to a free response task. Only recently has work been done using drift diffusion for multiple alternative choice tasks (Slezak, Sigman, & Cecchi, 2018).

Funding

The author has no funding to report.

Competing Interests

The author has declared that no competing interests exist.

Acknowledgments

The author has no support to report.

Data Availability

For Experiment 1 and Experiment 2 supplementary materials are freely available (see the Supplementary Materials section).

Supplementary Materials

The following Supplementary Materials are available via the PsychArchives repository (for access see Index of Supplementary Materials below):

1. *Stimulus Files:* Picture files for the numerical comparison task as described in the manuscript and appendix.

2. *Estimation Task Stimuli Files:* Picture files of the estimation task as described in the manuscript and appendix.

Index of Supplementary Materials

Prather, R. (2019). *Supplementary materials to "Individual differences in numerical comparison is independent of numerical precision".* PsychOpen. https://doi.org/10.23668/psycharchives.2518

References

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience, 10*(9), 1214-1221. https://doi.org/10.1038/nn1954

Brannon, E. M., & Terrace, H. S. (2000). Representation of the numerosities 1-9 by rhesus macaques (Macaca mulatta). *Journal of Experimental Psychology: Animal Behavior Processes, 26*(1), 31-49. https://doi.org/10.1037/0097-7403.26.1.31

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review, 100*(3), 432-459. https://doi.org/10.1037/0033-295X.100.3.432

Castronovo, J., & Göbel, S. M. (2012). Impact of high mathematics education on the number sense. *PLOS ONE, 7*(4), Article e33832. https://doi.org/10.1371/journal.pone.0033832

Chen, Q., & Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: A meta-analysis. *Acta Psychologica, 148*, 163-172. https://doi.org/10.1016/j.actpsy.2014.01.016

Chesney, D., Bjalkebring, P., & Peters, E. (2015). How to estimate how well people estimate: Evaluating measures of individual differences in the approximate number system. *Attention, Perception & Psychophysics, 77*(8), 2781-2802. https://doi.org/10.3758/s13414-015-0974-6

Cohen Kadosh, R., Cohen Kadosh, K., & Henik, A. (2008). When brightness counts: The neuronal correlate of numerical-luminance interference. *Cerebral Cortex, 18*(2), 337-343. https://doi.org/10.1093/cercor/bhm058

Cragg, L., & Gilmore, C. (2014). Skills underlying mathematics: The role of executive function in the development of mathematics proficiency. *Trends in Neuroscience and Education, 3*, 63-68. https://doi.org/10.1016/j.tine.2013.12.001

Crollen, V., Collignon, O., & Noël, M.-P. (2017). Visuo-spatial processes as a domain-general factor impacting numerical development in atypical populations. *Journal of Numerical Cognition, 3*(2), 344-364. https://doi.org/10.5964/jnc.v3i2.44

DeWind, N. K., & Brannon, E. M. (2012). Malleability of the approximate number system: Effects of feedback and training. *Frontiers in Human Neuroscience, 6*, Article 68. https://doi.org/10.3389/fnhum.2012.00068

Dietrich, J. F., Huber, S., & Nuerk, H. C. (2015). Methodological aspects to be considered when measuring the approximate number system (ANS) – A research review. *Frontiers in Psychology, 6*, Article 295. https://doi.org/10.3389/fpsyg.2015.00295

Feigenson, L., Dehaene, S., Spelke, E. S., Feigenson, L., & Spelke, E. S. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*, 307-314. https://doi.org/10.1016/j.tics.2004.05.002

Gilmore, C. K., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., . . . Inglis, M. (2013). Individual differences in inhibitory control, not non-verbal number acuity, correlate with mathematics achievement. *PLOS ONE, 8*(6), Article e67374. https://doi.org/10.1371/journal.pone.0067374

Gilmore, C. K., Attridge, N., & Inglis, M. (2011). Measuring the approximate number system. *Quarterly Journal of Experimental Psychology, 64*(11), 2099-2109. https://doi.org/10.1080/17470218.2011.574710

Gilmore, C. K., Keeble, S., Richardson, S., & Cragg, L. (2017). The interaction of procedural skill, conceptual understanding and executive functions in early mathematics achievement. *Journal of Numerical Cognition, 3*(2), 400-416. https://doi.org/10.5964/jnc.v3i2.51

Gilmore, C. K., McCarthy, S. E., & Spelke, E. S. (2010). Non-symbolic arithmetic abilities and mathematics achievement in the first year of formal schooling. *Cognition, 115*(3), 394-406. https://doi.org/10.1016/j.cognition.2010.02.002

Ginsburg, H., & Baroody, A. J. (2003). *Test of Early Mathematics Ability* (3rd ed.). Austin, TX, USA: Pro-Ed.

Gottfried, J. A., O'Doherty, J., & Dolan, R. J. (2003). Encoding predictive reward value in human amygdala and orbitofrontal cortex. *Science, 301*(5636), 1104-1107. https://doi.org/10.1126/science.1087919

Guillaume, M., Gevers, W., & Content, A. (2016). Assessing the approximate number system: No relation between numerical comparison and estimation tasks. *Psychological Research, 80*(2), 248-258. https://doi.org/10.1007/s00426-015-0657-x

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the "Number Sense": The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology, 44*(5), 1457-1465. https://doi.org/10.1037/a0012682

Halberda, J., Mazzocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature, 455*(7213), 665-668. https://doi.org/10.1038/nature07246

Izard, V., & Dehaene, S. (2008). Calibrating the mental number line. *Cognition, 106*(3), 1221-1247. https://doi.org/10.1016/j.cognition.2007.06.004

Libertus, M. E., Feigenson, L., & Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Developmental Science, 14*(6), 1292-1300. https://doi.org/10.1111/j.1467-7687.2011.01080.x

Libertus, M. E., Feigenson, L., & Halberda, J. (2013). Is approximate number precision a stable predictor of math ability? *Learning and Individual Differences, 25*, 126-133. https://doi.org/10.1016/j.lindif.2013.02.001

Libertus, M. E., Odic, D., Feigenson, L., & Halberda, J. (2016). The precision of mapping between number words and the approximate number system predicts children's formal math abilities. *Journal of Experimental Child Psychology, 150*, 207-226. https://doi.org/10.1016/j.jecp.2016.06.003

Lukowski, S. L., Rosenberg-Lee, M., Thompson, L. A., Hart, S. A., Willcutt, E. G., Olson, R. K., & Pennington, B. F. (2017). Approximate number sense shares etiological overlap with mathematics and general cognitive ability. *Intelligence, 65*, 67-74. https://doi.org/10.1016/j.intell.2017.08.005

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information.* San Francisco, CA, USA: W. H. Freeman and Company.

Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011). Preschoolers' precision of the approximate number system predicts later school mathematics performance. *PLOS ONE, 6*(9), Article e23749. https://doi.org/10.1371/journal.pone.0023749

McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., & Smith, L. B. (2010). Letting structure emerge: Connectionist and dynamical systems approaches to cognition. *Trends in Cognitive Sciences, 14*(8), 348-356. https://doi.org/10.1016/j.tics.2010.06.002

Moeller, K., Neuburger, S., Kaufmann, L., Landerl, K., & Nuerk, H.-C. (2009). Basic number processing deficits in developmental dyscalculia: Evidence from eye tracking. *Cognitive Development, 24*(4), 371-386. https://doi.org/10.1016/j.cogdev.2009.09.007

Moskaleva, M., & Nieder, A. (2014). Stable numerosity representations irrespective of magnitude context in macaque prefrontal cortex. *The European Journal of Neuroscience, 39*(5), 866-874. https://doi.org/10.1111/ejn.12451

Nieder, A., & Dehaene, S. (2009). Representation of number in the brain. *Annual Review of Neuroscience, 32*, 185-208. https://doi.org/10.1146/annurev.neuro.051508.135550

Nieder, A., Freedman, D., & Miller, E. K. (2002). Representation of quantity of visual items in the primate prefrontal cortex. *Science, 297*, 1708-1711. https://doi.org/10.1126/science.1072493

Nieder, A., & Merten, K. (2007). A labeled-line code for small and large numerosities in the monkey prefrontal cortex. *The Journal of Neuroscience, 27*(22), 5986-5993. https://doi.org/10.1523/JNEUROSCI.1056-07.2007

Park, J., & Brannon, E. M. (2013). Training the approximate number system improves math proficiency. *Psychological Science, 24*(10), 2013-2019. https://doi.org/10.1177/0956797613482944

Park, J., & Brannon, E. M. (2014). Improving arithmetic performance with number sense training: An investigation of underlying mechanism. *Cognition, 133*(1), 188-200. https://doi.org/10.1016/j.cognition.2014.06.011

Park, J., & Starns, J. J. (2015). The approximate number system acuity redefined: A diffusion model approach. *Frontiers in Psychology, 6*, Article 1955. https://doi.org/10.3389/fpsyg.2015.01955

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science, 306*(5695), 499-503. https://doi.org/10.1126/science.1102085

Pinheiro-Chagas, P., Wood, G., Knops, A., Krinzinger, H., Lonnemann, J., Starling-Alves, I., . . . Haase, V. G. (2014). In how many ways is the approximate number system associated with exact calculation? *PLOS ONE, 9*(11), Article e111155. https://doi.org/10.1371/journal.pone.0111155

Pirrone, A., Marshall, J. A. R., & Stafford, T. (2017). A drift diffusion model account of the semantic congruity effect in a classification paradigm. *Journal of Numerical Cognition, 3*(1), 77-96. https://doi.org/10.5964/jnc.v3i1.79

Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review, 117*(3), 864-901. https://doi.org/10.1037/a0019737

Prather, R. W. (2012). Connecting neural coding to number cognition: A computational account. *Developmental Science, 15*(4), 589-600. https://doi.org/10.1111/j.1467-7687.2012.01156.x

Prather, R. W. (2014). Numerical discrimination is mediated by neural coding variation. *Cognition, 133*(3), 601-610. https://doi.org/10.1016/j.cognition.2014.08.003

Purcell, B. A., Heitz, R. P., Cohen, J. Y., Schall, J. D., Gordon, D., Palmeri, T. J., & Palmeri, T. J. (2010). Neurally constrained modeling of perceptual decision making. *Psychological Review, 117*(4), 1113-1143. https://doi.org/10.1037/a0020311

Purpura, D. J., & Simms, V. (2018). Approximate Number System development in preschool: What factors predict change? *Cognitive Development, 45*, 31-39. https://doi.org/10.1016/j.cogdev.2017.11.001

Ramani, G. B., Jaeggi, S. M., Daubert, E. N., & Buschkuehl, M. (2017). Domain-specific and domain-general training to improve kindergarten children's mathematics. *Journal of Numerical Cognition, 3*(2), 468-495. https://doi.org/10.5964/jnc.v3i2.31

Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews: Neuroscience, 9*(7), 545-556. https://doi.org/10.1038/nrn2357

Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, S., Stricker, J., & De Smedt, B. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: A meta-analysis. *Developmental Science, 20*(3), Article e12372. https://doi.org/10.1111/desc.12372

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science, 14*(3), 237-243. https://doi.org/10.1111/1467-9280.02438

Silvetti, M., Seurinck, R., & Verguts, T. (2011). Value and prediction error in medial frontal cortex: Integrating the single-unit and systems levels of analysis. *Frontiers in Human Neuroscience, 5*, Article 75. https://doi.org/10.3389/fnhum.2011.00075

Simmering, V. R., & Perone, S. (2013). Working memory capacity as a dynamic process. *Frontiers in Psychology, 3*, Article 567. https://doi.org/10.3389/fpsyg.2012.00567

Slezak, D. F., Sigman, M., & Cecchi, G. (2018). An entropic barriers diffusion theory of decision-making in multiple alternative tasks. *PLOS Computational Biology, 14*(3), Article e1005961. https://doi.org/10.1101/245308

Smets, K., Gebuis, T., Defever, E., & Reynvoet, B. (2014). Concurrent validity of approximate number sense tasks in adults and children. *Acta Psychologica, 150*, 120-128. https://doi.org/10.1016/j.actpsy.2014.05.001

Spencer, J. P., Smith, L. B., & Thelen, E. (2001). Tests of a dynamic systems account of the A-not-B Error: The influence of prior experience on the spatial memory abilities of two-year-olds. *Child Development, 72*(5), 1327-1346. https://doi.org/10.1111/1467-8624.00351

Sugrue, L. P., Corrado, G. S., & Newsome, W. T. (2005). Choosing the greater of two goods: Neural currencies for valuation and decision making. *Nature Reviews: Neuroscience, 6*(5), 363-375. https://doi.org/10.1038/nrn1666

Tudusciuc, O., & Nieder, A. (2007). Neuronal population coding of continuous and discrete quantity in the primate posterior parietal cortex. *PNAS, 104*(36), 14513-14518. https://doi.org/10.1073/pnas.0705495104

Van Opstal, F., & Verguts, T. (2013). Is there a generalized magnitude system in the brain? Behavioral, neuroimaging, and computational evidence. *Frontiers in Psychology, 4*, Article 435. https://doi.org/10.3389/fpsyg.2013.00435

Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences, 7*(11), 483-488. https://doi.org/10.1016/j.tics.2003.09.002

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition, 74*, B1-B11. https://doi.org/10.1016/S0010-0277(99)00066-9

Appendix

Comparison Task Values

Experiment 1: 60 44, 30 35, 17 15, 30 22, 45 51, 20 22, 52 40, 30 34, 90 105, 78 60, 60 64, 25 28, 15 17, 20 25, 20 26, 40 52, 60 44, 30 35, 44 60, 64 60, 56 50, 26 20, 11 10, 60 78, 105 90, 84 75, 50 56, 10 11, 75 60, 20 22, 90 96, 45 51, 51 45, 25 20, 90 66, 60 64, 30 22, 15 17, 52 40, 60 78, 33 30, 105 90, 30 32, 50 56, 50 40, 20 26, 66 90, 44 60, 70 60, 30 33, 40 50, 50 40, 11 10, 75 84, 84 75, 35 30, 28 25, 40 50, 28 25, 90 96, 17 15, 90 66, 60 75, 20 25, 40 52, 66 90, 22 30, 35 30, 30 33, 90 105, 60 70, 75 84, 22 30, 33 30, 22 20, 34 30, 60 75, 30 32, 78 60, 60 70, 32 30, 96 90, 32 30, 56 50, 22 20, 25 20, 26 20, 64 60, 30 34, 70 60, 96 90, 51 45, 34 30, 25 28, 10 11, 75 60

Experiment 2: 51 30, 18 15, 35 25, 56 35, 13 10, 45 30, 22 20, 26 14, 18 15, 34 20, 30 25, 51 30, 11 10, 15 10, 45 30, 75 50, 18 15, 32 20, 42 40, 30 20, 34 20, 56 35, 39 30, 42 40, 13 10, 13 10, 75 50, 18 15, 33 30, 13 10, 60 40, 54 45, 60 40, 11 10, 56 35, 18 15, 35 25, 54 45, 39 30, 35 25, 56 35, 33 30, 56 35, 51 30, 26 14, 26 14, 60 40, 56 35, 75 50, 54 45, 15 10, 32 20, 35 25, 56 35, 30 25, 54 45, 34 20, 51 30, 18 15, 33 30, 35 25, 60 40, 45 30, 11 10, 22 20, 19 14, 13 10, 35 25, 13 10, 51 30, 33 30, 30 20, 42 40, 39 30, 39 30, 39 30, 75 50, 54 45, 15 10, 18 15, 34 20, 15 10, 18 10, 34 20, 51 30, 30 20, 11 10, 11 10, 45 30, 30 20

Stimuli: see Supplementary Materials

Estimation Task Values

23, 29, 33, 37, 69, 87, 99, 111