



ELSEVIER

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp



CrossMark

Individual differences in the components of children's and adults' information processing for simple symbolic and non-symbolic numeric decisions

Clarissa A. Thompson^{a,*}, Roger Ratcliff^b, Gail McKoon^b

^aDepartment of Psychological Sciences, Kent State University, Kent, OH 44242, USA

^bDepartment of Psychology, The Ohio State University, Columbus, OH 43210, USA

ARTICLE INFO

Article history:

Received 17 November 2015

Revised 18 April 2016

Available online 28 May 2016

Keywords:

Diffusion model

Response time

Accuracy

Cognitive development

Numerical ability

Individual differences

ABSTRACT

How do speed and accuracy trade off, and what components of information processing develop as children and adults make simple numeric comparisons? Data from symbolic and non-symbolic number tasks were collected from 19 first graders ($M_{\text{age}} = 7.12$ years), 26 second/third graders ($M_{\text{age}} = 8.20$ years), 27 fourth/fifth graders ($M_{\text{age}} = 10.46$ years), and 19 seventh/eighth graders ($M_{\text{age}} = 13.22$ years). The non-symbolic task asked children to decide whether an array of asterisks had a larger or smaller number than 50, and the symbolic task asked whether a two-digit number was greater than or less than 50. We used a diffusion model analysis to estimate components of processing in tasks from accuracy, correct and error response times, and response time (RT) distributions. Participants who were accurate on one task were accurate on the other task, and participants who made fast decisions on one task made fast decisions on the other task. Older participants extracted a higher quality of information from the stimulus arrays, were more willing to make a decision, and were faster at encoding, transforming the stimulus representation, and executing their responses. Individual participants' accuracy and RTs were uncorrelated. Drift rate and boundary settings were significantly related across tasks, but they were unrelated to each other. Accuracy was mainly determined by drift rate, and RT was mainly determined by boundary separation. We concluded that RT and accuracy operate largely independently.

© 2016 Elsevier Inc. All rights reserved.

* Corresponding author.

E-mail address: cthomp77@kent.edu (C.A. Thompson).

Introduction

One debate in the numerical cognition literature focuses on whether performance across individuals on precise symbolic numerical tasks, in which children and adults compare Arabic numerals such as 45 versus 50, is correlated with performance on approximate non-symbolic numerosity tasks, in which participants compare 45 versus 50 dots. The nature of the relation between these tasks and among these tasks and overall mathematics achievement tests is unclear because there is a discrepancy in the findings reported in the numerical cognition literature. Sometimes performance on symbolic and non-symbolic tasks is correlated, but sometimes it is not (De Smedt, Verschaffel, & Ghesquiere, 2009; Fazio, Bailey, Thompson, & Siegler, 2014; Gilmore, Attridge, & Inglis, 2011; Halberda, Ly, Wilmer, Naiman, & Germine, 2012; Holloway & Ansari, 2009; Maloney, Risko, Preston, Ansari, & Fugelsang, 2010; Price, Palmer, Battista, & Ansari, 2012; Sasanguie, Defever, Van den Bussche, & Reynvoet, 2011). Sometimes performance on non-symbolic comparison tasks is correlated with mathematics achievement scores, and sometimes it is not (De Smedt et al., 2009; Durand, Hulme, Larkin, & Snowling, 2005; Fazio et al., 2014; Gilmore, McCarthy, & Spelke, 2010; Halberda, Mazzocco, & Feigenson, 2008; Holloway & Ansari, 2009; Inglis, Attridge, Batchelor, & Gilmore, 2011; Libertus, Feigenson, & Halberda, 2011; Lyons & Beilock, 2011; Mazzocco, Feigenson, & Halberda, 2011a, 2011b; Mundy & Gilmore, 2009; Price et al., 2012).

Recent research by Inglis and colleagues (2011), Halberda and colleagues (2012), and Fazio and colleagues (2014) illustrates variability in the reported strength of the relationship between non-symbolic tasks and mathematics achievement. Inglis and colleagues (2011) reported a strong correlation for 7- to 9-year-old children between accuracy on the calculation subtests of the Woodcock–Johnson mathematics achievement test and accuracy on a non-symbolic dot comparison task. Accuracy was operationalized as the Weber fraction, w , which is a measure of precision/acuity of an individual's approximate number system (ANS). It is hypothesized that the ANS allows humans of all ages, as well as animals, to approximate numerical magnitudes or values (Dehaene, Dehaene-Lambertz, & Cohen, 1998). Larger values of w indicate lower levels of numerical precision. Inglis and colleagues found a negative correlation between w on the non-symbolic comparison task and standardized mathematics achievement scores for children but not for adults. That is, those children who had larger Weber fractions had lower standardized mathematics achievement scores. Inglis and colleagues' (2011) findings contrast with those of Halberda and colleagues (2012), who used a large internet-based sample of children and adults between 11 and 85 years of age to complete a non-symbolic number comparison task. Halberda and colleagues found a small but significant correlation between ANS precision on the non-symbolic comparison task and school mathematics ability across all ages tested. Fazio and colleagues (2014) found that after controlling for standardized reading performance, a non-mathematics test of basic cognitive ability, both symbolic and non-symbolic performance uniquely predicted mathematics achievement scores. That is, both symbolic knowledge and non-symbolic knowledge contribute to overall mathematics achievement. However, performance on symbolic tasks explained a greater proportion of the variance in mathematics achievement than did performance on non-symbolic tasks. Taken together, these three examples illustrate the variability in the strength of the correlations that exist between non-symbolic task performance and mathematics achievement across the lifespan. Two recent meta-analyses (Chen & Li, 2014; Fazio et al., 2014) provided possible explanations for these variable findings.

Chen and Li's (2014) meta-analysis claims that one source of the discrepant results is that many of the previously conducted studies may have been underpowered, and the small sample sizes led to spurious conclusions. Fazio and colleagues' (2014) meta-analysis found a small reliable correlation between non-symbolic numerical performance and overall mathematics achievement. The size of the correlations reported in the individual studies that comprised the meta-analysis differed with regard to the dependent variables that the studies' authors chose. For example, when accuracy (percentage correct) or accuracy-based measures (w) in addition to response time (RT) was the dependent variable of choice, the correlation between performance on the non-symbolic numerical task and standardized mathematics achievement was stronger than when a measure of RT was used alone.

The relationship between performance on non-symbolic tasks and overall mathematics achievement was also stronger for children under 6 years of age as compared with children between 6 and 18 years of age and adult participants. This finding is consistent with age differences found by [Inglis and colleagues \(2011\)](#). Fazio and colleagues interpreted the age differences to mean that formal schooling might play a role in the developmental differences and offered possible interpretations. First, children under 6 years of age might be more reliant on non-symbolic numerical representations than children in the 6- to 18-year-old age range and adults, who presumably have extensive experience in working with symbolic numbers in formal schooling environments. Second, mathematics achievement tests for children under 6 years of age might be more sensitive to non-symbolic numerical knowledge. In summary, [Fazio and colleagues' \(2014\)](#) meta-analysis indicated that the age of the participants completing the non-symbolic and mathematics achievement tests and the dependent variables chosen to measure the numerical knowledge might contribute to the discrepancies found in the numerical cognition literature.

Recent work by [Ratcliff, Thompson, and McKoon \(2015\)](#), in which the diffusion model ([Ratcliff, 1978](#)) was applied to adults' performance on symbolic and non-symbolic discrimination tasks, agreed with conclusions drawn by [Fazio and colleagues \(2014\)](#). They argued that it is misleading for researchers to report only RT or accuracy from symbolic and non-symbolic discrimination tasks because each dependent variable could present a slightly different view of participants' numerical information processing abilities. Intuition might suggest that RT and accuracy are negatively correlated with one another such that better performance means more accurate responses and faster responses relative to the average values for the group (e.g., see [Siegler's \(1996\)](#) overlapping waves theory that describes how children know and use various strategies that differ in accuracy and efficiency). However, [Ratcliff and colleagues \(2015\)](#) provided evidence that an individual's accuracy and RT were not correlated within symbolic and non-symbolic numerical discrimination tasks (this is also true in other tasks; [Ratcliff, Thapar, & McKoon, 2010, 2011](#)). Knowing how accurately a participant decided whether 45 was greater than or less than 50 did not predict how quickly the participant made the decision. The diffusion model, described in detail below, provides an explanation of this result. The model was fit to correct and error RT distributions and accuracy, and it produced model parameters representing the quality of evidence (drift rate), the amount of evidence needed for a decision (boundary separation), and the duration of non-decision processes (non-decision time). Application of the model showed that boundary settings and drift rates were largely uncorrelated with each other and that accuracy was mainly determined by drift rate and RT was largely determined by boundary separation (and non-decision time). This provided a way of understanding the lack of correlation between RT and accuracy (but did not predict why there was no correlation). People use global boundary settings for decisions, and they do not modulate their settings as a function of whether they are good or bad at a particular task ([Ratcliff et al., 2015](#)).

In this article, we show how, according to the diffusion model, accuracy and RT and dependent measures based on them do not assess independent abilities but instead are manifestations of the same underlying cognitive processes. In the studies cited above ([Fazio et al., 2014](#); [Halberda et al., 2012](#); [Inglis et al., 2011](#); [Ratcliff et al., 2015](#)), the response required of a participant is a decision between two (or more) alternatives (e.g., "is this number larger or smaller than 50?"). Whatever the quality of a participant's numerosity information, a response must be chosen, and the choice will take some amount of time. Accuracy and speed can trade off, and the trade-off is under a participant's control. A participant might decide to respond as quickly as possible, sacrificing accuracy, or as accurately as possible, sacrificing speed. If a participant adopts a speed emphasis, the slope of the RT–difficulty function will be lower than if the participant adopts an accuracy emphasis ([Ratcliff et al., 2015](#), Experiment 2). In consequence, differences among participants in the quality of the numeracy information on which they base their decisions can be obscured by differences in their speed/accuracy settings. The only way to separate information quality from speed/accuracy settings is to understand how they interact. We do that with a sequential sampling decision model, [Ratcliff's \(1978\)](#) diffusion model (see also [Ratcliff & McKoon, 2008](#)), which is described below.

The diffusion model

In Ratcliff's (1978) diffusion model (see also Ratcliff & McKoon, 2008), there is noisy accumulation of stimulus information over time toward one of two boundaries. A response is made when the amount of accumulated information reaches one of two boundaries, or criteria, one for each of the two possible choices (e.g., "is 60 greater than or less than 50?"). The rate of accumulation, called "drift rate," is determined by the quality of the information used in a decision. For example, information about "is 9 greater than 1?" would be stronger than information about "is 2 greater than 1?" and the information available to a high school student would likely be stronger than the information available to a second grader (e.g., Ratcliff, Love, Thompson, & Opfer, 2012). Drift rate corresponds to numerical acuity in the numerosity and number discrimination tasks.

Fig. 1 shows the operation of the model. Total RT is the time it takes to encode a stimulus, transform the stimulus representation to a decision-relevant representation, compare the representation with memory, decide on a response, execute a response, and so forth. The transformation from the stimulus to a decision-relevant representation maps the many dimensions of a stimulus (e.g., size, color, shape, number) onto the task-relevant dimension—the drift rate that drives the decision process. The accumulation of information begins at a starting point (z in Fig. 1) and proceeds until one of the two boundaries is reached (a or 0 in the figure). Because the accumulation process is noisy, for a given value of drift rate, at each instant of time there is some probability of moving toward the correct boundary and some smaller probability of moving toward the incorrect boundary. This variability means that accumulated information can hit the wrong boundary, producing errors, and that stimuli with the same values of drift rate can hit a boundary at different times. For application of the model, non-decision processes (e.g., stimulus encoding, transformation to task-relevant information, response execution) are combined into one parameter, T_{er} in Fig. 1. As illustrated in the figure, the model predicts the right-skewed shapes of RT distributions that are observed empirically in two-choice tasks.

The model decomposes accuracy and RTs into the three main components just described: drift rates, boundary settings, and non-decision processes. The values of these components vary from trial to trial because, it is assumed, participants cannot accurately set identical values from trial to trial (e.g., Laming, 1968; Ratcliff, 1978). Across-trial variability in drift rate is assumed to be normally distributed with SD η , across-trial variability in the starting point (equivalent to across-trial variability in

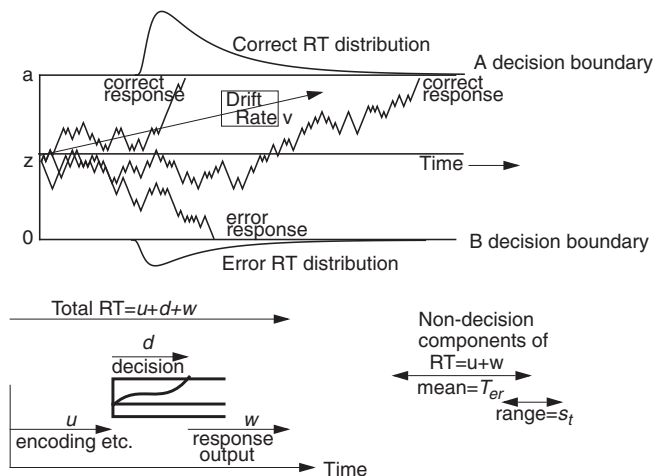


Fig. 1. Illustration of the diffusion model. The top panel shows three simulated paths with drift rate v , starting point z , and boundary separation a . Drift rate is normally distributed with SD η , and starting point is uniformly distributed with range s_z . Non-decision time is composed of encoding processes, processes that turn the stimulus representation into a decision-related representation, and response output processes. Non-decision time has mean T_{er} and a uniform distribution with range s_t .

the boundary positions) is assumed to be uniformly distributed with range s_z , and across-trial variability in the non-decision component is assumed to be uniformly distributed with range s_c . These distributional assumptions are the ones usually made, but they are not critical as long as they are within their usual ranges (Ratcliff, 2013).

Because the model decomposes accuracy and RTs into components and separates out variability in them, the power to observe effects of independent variables on performance can be substantially increased. For example, lexical decision experiments have been used to attempt to identify participants with high anxiety by looking at their performance on “threat” words (e.g., anger, hostility, attack). Significant differences between high-anxiety and low-anxiety participants did not appear with RTs or accuracy but did appear with drift rates. The model analyses increased power by a factor of approximately 2 (White, Ratcliff, Vasey, & McKoon, 2010).

Current theories about numeracy are constrained only by mean RTs for correct responses or only by accuracy. The diffusion model is more tightly constrained and can be falsified. The most powerful constraint comes from the requirement that the model fit the right-skewed shape of RT distributions, as shown in Fig. 1 (Ratcliff, 1978, 2002; Ratcliff & McKoon, 2008; Ratcliff, Van Zandt, & McKoon, 1999). Ratcliff (2002) generated simulated data for which RT distributions behaved across conditions in ways that are plausible but never obtained empirically. In all cases, the model failed to fit the data. In addition, across experimental conditions that vary in difficulty, such as when target numbers differ in absolute distance from a comparison number (e.g., 50) in a symbolic number discrimination task, changes in accuracy, RT distributions, and the relative speeds of correct and error responses must all be captured by changes in only one parameter of the model, namely drift rate. Across experimental conditions that vary in speed/accuracy criteria (e.g., speed vs. accuracy instructions to participants), all changes in accuracy, RT distributions, and the relative speeds of correct and error responses are usually captured by changes only in the settings of the boundaries. The boundaries cannot be adjusted as a function of difficulty because it would be necessary for the system to know which level of difficulty was being tested before boundary settings could be determined. The model explains the ways in which drift rates and boundary settings can interact to determine participants' RTs and accuracy. For a given value of drift rate, a participant can adopt wider boundaries and so be more accurate but slower, whereas the participant can adopt narrower boundaries and so be faster but less accurate. Across participants, drift rates and boundary settings can differ independently. Participants who have high drift rates will have good accuracy and fast responses when their boundaries are close together, and they will have good (perhaps slightly better) accuracy and slow responses when their boundaries are farther apart. Participants who have low drift rates will have poor accuracy and fast responses when their boundaries are close together, and they will have (perhaps) somewhat better accuracy and slow responses when their boundaries are far apart. To put this another way, participants with fast responses can be accurate or inaccurate, and participants with slow responses can be accurate or inaccurate (cf. Ratcliff, Thapar, & McKoon, 2006; Ratcliff et al., 2010, 2011). That boundary settings are under a participant's control has been demonstrated in past studies where participants responded to instructions to maximize speed by decreasing the settings (Ratcliff, Thapar, & McKoon, 2001, 2003, 2004; Thapar, Ratcliff, & McKoon, 2003). In the model, accuracy is related to drift rate and RT is related to speed–accuracy criteria, but drift rate and criteria are not related to each other across participants. This provides a theoretical basis for understanding why the common belief that accurate participants do not always make the quickest decisions does not always hold true. Ratcliff and colleagues (2015) provided evidence that accuracy, RT, and dependent measures based on them do not assess independent abilities but instead are manifestations of the same underlying cognitive processes. The finding that accuracy–RT correlations were not significant calls into question interpretations of results from the many previous studies in the numerical cognition literature for which only RTs or only accuracy values were reported. The finding also puts a stringent constraint on theories about number processing; whatever it is that determines a participant's overall accuracy is not directly related to whatever it is that determines the participant's overall speed. Any theory about performance in numeracy tasks must explain why this is so. In summary, the diffusion model breaks down accuracy and RT into components of information processing, and these resulting model parameters are not significantly correlated with one another. Thus, the model allows researchers to assess speed–accuracy relationships.

The model also solves a scaling problem for RT and accuracy. In the numeracy literature, when a task has high accuracy the performance measure is typically RT (or a measure based on RT), and when a task has low accuracy the measure is typically accuracy (or a measure based on accuracy). The two measures have different scales; accuracy varies from chance to ceiling and asymptotes at 1, and RTs vary from a short minimum to potentially very long duration. The diffusion model resolves this issue because the two measures come from the same underlying processes.

The model also helps to address problems with ceiling and floor effects (Ratcliff, 2014). For some experiments, accuracy might be at chance for several of the most difficult conditions (e.g., “is 48 more or less than 50?”), and it might be at ceiling for several of the easiest conditions (e.g., “is 13 more or less than 50?”). Model parameters (especially parameters that represent across-trial variability in model components) can be estimated when accuracy is not at ceiling or floor (0 or 1). When accuracy is at ceiling or floor, the model can estimate differences in drift rates because the other conditions allow model parameters to be estimated. Because RTs differ across these high-accuracy conditions, RTs are sufficient to determine drift rates. Ceiling effects are likely present in symbolic number discrimination tasks, especially for older children and adults, because participants have extensive experience in thinking about and processing whole numbers.

The current experiment

Ratcliff and colleagues (2015) tested college-aged adults on four numerical tasks: (a) numerosity discrimination (a non-symbolic numerical decision task in which participants decided whether a stimulus array contained more or less than 50 asterisks), (b) number discrimination (a symbolic numerical decision task in which participants decided whether an Arabic numeral was greater than or less than the number 50), (c) memory for two-digit numbers (participants indicated whether two-digit numbers on a test list appeared on an immediately preceding study list), and (d) memory for three-digit numbers (participants indicated whether three-digit numbers on a test list appeared on an immediately preceding study list). Results provided evidence that “on-line” decision tasks, such as numerosity and number discrimination that require perceptual discrimination yet do not require memory access, were correlated with performance on “off-line” tasks that do require memory access, such as memory for two- and three-digit numbers. Accuracy positively correlated across tasks, as did RTs. The more accurate an individual was in deciding whether the number of asterisks on the computer screen was greater than or less than 50 during the numerosity discrimination task, the more accurate the individual was when deciding whether a numeral was greater than or less than 50 in the number discrimination task. The faster a participant responded on one task, the faster the participant responded on the other task. However, if a participant was accurate, it did not mean that the participant was fast (and vice versa). There were significant positive correlations across the tasks between the quality of the numeracy information (drift rate) driving the decision process and between the speed/accuracy criterion settings, suggesting that similar numeracy skills and similar speed–accuracy settings are involved in numerosity and number discrimination and memory for two- and three-digit numbers. The novelty of our current experiment is that the diffusion model can account for correct and incorrect RTs, RT distributions, and accuracy. The model decomposes RTs into three main information processing components, and these parameters can indicate what develops over time as children become more accurate and faster at making numerical decisions. The advantage of applying the diffusion model to cross-sectional data is that the mechanism of change in simple numerical decisions across the lifespan can be highlighted. For instance, if drift rates increase with age or grade level, it could indicate that improvement in children’s magnitude understanding, or numerical acuity/discrimination ability, is driving the decision process. If boundary separation becomes narrower with age, it could indicate that children become more willing to execute a response. Finally, if the non-decision component becomes faster with age, a model-based interpretation is that the children are becoming more skilled at transforming the stimulus array into a form that will allow them to execute a discrimination decision.

In our current experiment, our *first aim* was to replicate the findings of Ratcliff and colleagues (2015) with children by investigating whether elementary and middle school students’ performance on the non-symbolic numerosity discrimination task (“is the number of asterisks in a 10×10 array greater than or less than 50 asterisks?”) was correlated with their performance on the symbolic number

discrimination task (“is an Arabic numeral greater than or less than the Arabic numeral 50?”). We tested four groups of children—first graders, second/third graders, fourth/fifth graders, and seventh/eighth graders—and then compared the children’s performance with performance of the adults reported in Ratcliff and colleagues (2015). A *second aim* was to investigate whether there were developmental progressions in diffusion model parameters across the lifespan on a symbolic numerical discrimination task. Ratcliff and colleagues (2012) reported developmental differences on a non-symbolic numerical discrimination task in drift rate, boundary separation, and non-decision time across the lifespan. Children, much like older adults, made slower decisions than did college-aged adults. However, developmental differences in the diffusion model parameters indicated that children and older adults made slow decisions for different reasons. Younger children required more evidence prior to making a decision, set more conservative decision criteria, and took longer to encode stimulus information, transform the representation to produce a decision related to numerical quantity, and execute a decision than did older children and college-aged adults. It is an open question as to whether drift rate, boundary separation, and non-decision times change across the lifespan for other numeracy tasks such as the number discrimination task, a task on which participants presumably improve because they are learning more about the relations between whole numbers during formal schooling. A *third aim* was to examine individual differences in accuracy and RT between the two tasks, to examine individual differences in model parameters between tasks, and to examine correlations between accuracy and RT within each task. This allowed us to investigate whether RT and accuracy on symbolic and non-symbolic discrimination tasks were relatively independent measures of performance within individual children, as was found previously with adults. If so, we could then argue against the practice of choosing one or the other dependent variable alone to assess children’s numerical discrimination abilities.

Method

Participants

Ratcliff and colleagues (2015) tested 32 college students at the University of Oklahoma ($M_{\text{age}} = 19.4$ years) who participated in one 60-min session in partial fulfillment of class requirements for an introductory psychology course. Results from these students served as a benchmark against which data from 91 newly-tested children in the current experiments were compared. There were four groups of children who each participated in two 30-min sessions. See Table 1 for participant demographics. In total, 19 first graders, 5 second/third graders, and 4 fourth/fifth graders were excluded from analyses because (a) they made a substantial proportion of fast responses (≤ 400 ms) showing that they were just hitting keys as fast as possible and had given up on the task or (b) they did not complete enough trials for modeling. Not completing enough trials for modeling was particularly a problem for first graders. These children either lacked an ability to complete the task or found the tasks to be very difficult/monotonous and, thus, decided to give up.

Procedure and stimuli

Participants completed the symbolic number discrimination task during Session 1 and completed the non-symbolic numerosity discrimination task during Session 2. Stimuli were displayed on the

Table 1
Participant demographics.

| Grade | <i>n</i> | Mean age (<i>SD</i>) | Sex | Ethnicity breakdown | Mean days between testing sessions |
|----------------|----------|------------------------|-----------|--|------------------------------------|
| First | 19 | 7.12 (0.21) | 53% girls | 79% Caucasian, 16% biracial, 5% Asian | 10.01 |
| Second/Third | 26 | 8.20 (1.83) | 62% girls | 85% Caucasian, 8% Asian, 4% biracial, 4% Hispanic | 22.62 |
| Fourth/Fifth | 27 | 10.46 (0.63) | 59% girls | 89% Caucasian, 4% African American, 4% Native American, 4% biracial | 28.59 |
| Seventh/Eighth | 19 | 13.22 (0.58) | 37% girls | 63% Caucasian, 11% Native American, 11% biracial, 5% African American, 5% Asian, 5% Hispanic | 2.47 |

screen of a laptop computer, and responses were collected from the laptop's keyboard using the “?” key for “large” responses and the “Z” key for “small” responses. Prior to beginning each experiment, participants were told that sometimes decisions would be difficult to make, but they were to respond as quickly and accurately as possible. Participants were encouraged to take a brief rest break between each block of trials. During the rest break, participants saw two progress bars that displayed their proportion of correct and error responses.

During Session 3, second/third and fourth/fifth graders completed the matrix reasoning and vocabulary subtests of the Wechsler Intelligence Scale for Children–Fourth Edition (WISC-IV). The second/third graders from our sample had a mean IQ score of 115 on the matrix reasoning and vocabulary subtests; the fourth/fifth graders had a mean IQ score of 117. We do not have IQ scores from the first graders, the seventh/eighth graders, or the college-aged adults. However, the IQ scores of the second/third and fourth/fifth graders are almost the same as those of a comparable college-aged sample ($M = 119$) who completed the Wechsler Adult Intelligence Scale–Revised (WAIS-R) matrix reasoning and vocabulary subtests (Ratcliff et al., 2001). Average achievement test scores (Basic Early Assessment of Reading [BEAR] and Oklahoma Core Curriculum Tests [OCCT]), obtained from the public school district administrators, and average IQ scores are listed in the [Appendix](#).

Symbolic number discrimination

White Arabic numerals between 10 and 90 were displayed on a black background in the middle of a laptop screen. Participants pushed the “Z” key if a small number between 10 and 49 was displayed and pushed the “?” key if a large number between 51 and 90 was displayed. Each numeral remained visible until a response key was pressed. After a key was pressed, the screen was cleared, a smiling (correct response) or frowning (incorrect response) face was displayed for 500 ms, a blank screen appeared for 100 ms, and then the next trial began. There were eight blocks of trials, 80 trials per block, with each of the possible numbers tested once in each block in random order.

Non-symbolic numerosity discrimination

On each trial, some number of white asterisks, between 11 and 90, was displayed against a black background. The asterisks occupied randomly selected positions in a 10×10 grid in the center of the laptop screen that subtended a visual angle of 7.5 degrees horizontally and 7.0 degrees vertically. Arrays containing fewer than 11 asterisks were not included to deter participants from counting.

Participants pushed the “Z” key if a small number of asterisks between 11 and 49 was displayed and pushed the “?” key if a large number of asterisks between 51 and 90 was displayed. The asterisks remained on the screen until a response key was pressed. Then the screen was cleared, a smiling (correct response) or frowning (incorrect response) face was displayed for 500 ms, the screen was cleared, a blank screen appeared for 100 ms, and then the next trial began. There were eight blocks of trials, 80 trials per block, with each of the possible numbers of asterisks tested once in each block in random order.

Results

Responses shorter than 300 ms for children and 250 ms for adults were eliminated from analyses, as were responses longer than 4000 ms for children and 2000 ms for adults. The percentages eliminated were 19.1%, 6.1%, 5.1%, 4.4%, and 3.0% for first graders, second/third graders, fourth/fifth graders, seventh/eighth graders, and college-aged adults, respectively, for the numerosity discrimination task, and 11.5%, 3.4%, 1.6%, 1.1%, and 3.0% for these corresponding groups in the number discrimination task.

[Table 2](#) shows the proportion correct and mean RT for children in first, second/third, fourth/fifth, and seventh/eighth grades and college-aged adults in the numerosity and number discrimination tasks. The left-most column of [Table 2](#) shows that experimental stimuli (e.g., number of asterisks or numerals presented on the computer screen) were grouped into conditions (six for the numerosity discrimination task and eight for the number discrimination task) to provide more observations per condition for application of the diffusion model. The conditions were created by grouping “small”

responses to small stimuli with “large” responses to large stimuli (e.g., 17–22 asterisks grouped with 71–76 asterisks). In this way, our conditions collapsed over absolute distance from the standard (e.g., 50 asterisks). In addition, we made sure that “small” responses to small stimuli were similar to “large” responses to large stimuli. Recent computational modeling results suggest that an interaction of stimulus ratio (e.g., Weber fraction) and absolute distance best explains adults’ performance on non-symbolic numerical discrimination tasks (Prather, 2014). But this appears to be task specific. In our numerosity and number discrimination tasks, “large” responses to large stimuli are symmetric to “small” responses to small stimuli (Ratcliff, 2014). Thus, using stimulus ratios would combine conditions with data that are not equivalent.

The participants were slightly biased in the numerosity discrimination task such that their chance level of accuracy for responding “small” and “large” was approximately 47 and not 50 (as in Ratcliff, 2014, Experiments 1 and 2; Ratcliff et al., 2015). It is for this reason that there are slightly more numbers in the first bin as compared with subsequent bins shown in Table 2.

Note that we refer interchangeably to age and grade because older children were in higher grades than younger children. However, we did not use age as a continuous variable in any of the analyses. Across all grades on both the numerosity and number discrimination tasks, accuracy increased and RT decreased as difficulty level decreased. That is, the closer the stimuli were to 50, the lower the accuracy and the longer the RT. Participant accuracy was at or near ceiling levels for the three easiest conditions in the numerosity discrimination task that contained the most/least asterisks (.799–.894 for first graders, .876–.936 for second/third graders, .891–.938 for fourth/fifth graders, .904–.947 for seventh/eighth graders, and .908–.943 for adults). RTs continued to decrease in the three easiest conditions as the stimuli became more discriminable, that is, as their distance from 50 increased (1076–994 ms for first graders, 911–845 ms for second/third graders, 795–705 ms for fourth/fifth graders, 660–593 ms for seventh/eighth graders, and 491–455 ms for adults). In the number discrimination task, accuracy was near ceiling, 80% to 97% range, for each of the conditions (.802–.906 for first graders, .854–.955 for second/third graders, .858–.957 for fourth/fifth graders, .870–.970 for seventh/eighth graders, and .879–.972 for adults). As in the numerosity discrimination task, RTs from the number discrimination task decreased substantially as the conditions went from hardest to easiest (1302–1135 ms for first graders, 1239–1015 ms for second/third graders, 1034–811 ms for fourth/fifth graders, 758–629 ms for seventh/eighth graders, and 632–501 ms for adults).

Fig. 2 plots accuracy and median RT against level of difficulty. Note that the most difficult conditions were the ones with the smallest absolute distance from 50. For both tasks, accuracy decreased as a function of difficulty, although the decrease in accuracy was more apparent for the

Table 2
Response proportions and mean RTs for the numerosity and number discrimination tasks.

| Task | Condition | 1 grade | | 2/3 grade | | 4/5 grade | | 7/8 grade | | Adult | |
|------------|-------------|----------|---------|-----------|---------|-----------|---------|-----------|---------|----------|---------|
| | | Pr. Cor. | Mean RT | Pr. Cor. | Mean RT | Pr. Cor. | Mean RT | Pr. Cor. | Mean RT | Pr. Cor. | Mean RT |
| Numerosity | 11–16/77–88 | .894 | 994 | .936 | 845 | .938 | 705 | .947 | 593 | .943 | 455 |
| | 17–22/71–76 | .856 | 1049 | .916 | 895 | .909 | 737 | .940 | 623 | .934 | 479 |
| | 23–28/65–70 | .799 | 1076 | .876 | 911 | .891 | 795 | .904 | 660 | .908 | 491 |
| | 29–34/59–64 | .757 | 1110 | .813 | 959 | .825 | 838 | .846 | 701 | .874 | 527 |
| | 35–40/53–58 | .671 | 1109 | .710 | 1009 | .735 | 864 | .753 | 747 | .783 | 558 |
| | 41–46/47–52 | .570 | 1139 | .590 | 1034 | .594 | 901 | .600 | 799 | .617 | 591 |
| Number | 10–14/86–90 | .906 | 1135 | .955 | 1015 | .957 | 811 | .970 | 629 | .972 | 501 |
| | 15–19/81–85 | .887 | 1179 | .930 | 1016 | .950 | 861 | .948 | 616 | .963 | 512 |
| | 20–24/76–80 | .867 | 1155 | .934 | 1029 | .945 | 868 | .956 | 616 | .968 | 510 |
| | 25–29/71–75 | .852 | 1203 | .938 | 1056 | .943 | 867 | .948 | 637 | .952 | 529 |
| | 30–34/66–70 | .860 | 1209 | .916 | 1099 | .929 | 904 | .937 | 664 | .956 | 534 |
| | 35–39/61–65 | .854 | 1246 | .908 | 1149 | .919 | 946 | .924 | 676 | .943 | 563 |
| | 40–44/56–60 | .821 | 1253 | .859 | 1176 | .880 | 1000 | .896 | 723 | .904 | 590 |
| | 45–49/51–55 | .802 | 1302 | .854 | 1239 | .858 | 1034 | .870 | 758 | .879 | 632 |

Note. Pr. Cor., proportion correct; Mean Cor. RT, mean correct response time.

non-symbolic task than for the symbolic task in which performance was at ceiling. RT also increased as difficulty increased across both tasks. There was no evidence of a size effect, with smaller numbers discriminated from the 50 cutoff better than larger numbers (also see Figs. 2 and 4 in Ratcliff, 2014). Dehaene, Dupoux, and Mehler (1990) provided evidence that 1 and 2 are easier to discriminate from one another than are larger numbers equidistant apart, such as 8 and 9, but we did not see this result with the larger numbers used in our experiments. This may also reflect a task difference; judging differences between two numbers or numerosities is different from judging whether one number or numerosity is different from a criterion number or numerosity. Importantly, our numerosity task is tapping a basic numeracy skill because it has been shown to correlate with number discrimination as well as number memory tasks (Ratcliff et al., 2015).

When accuracy was averaged across all of the easy and difficult conditions, first graders were less accurate than all other groups on the numerosity and number discrimination tasks (see Table 3). There were no other significant differences in averaged accuracy across the groups of participants. These data averaged across conditions indicated that children in second/third grade were responding as accurately as older children and adults on the two tasks; however, there were significant grade differences in median RT averaged across conditions (see Table 3) for both tasks. First through third graders

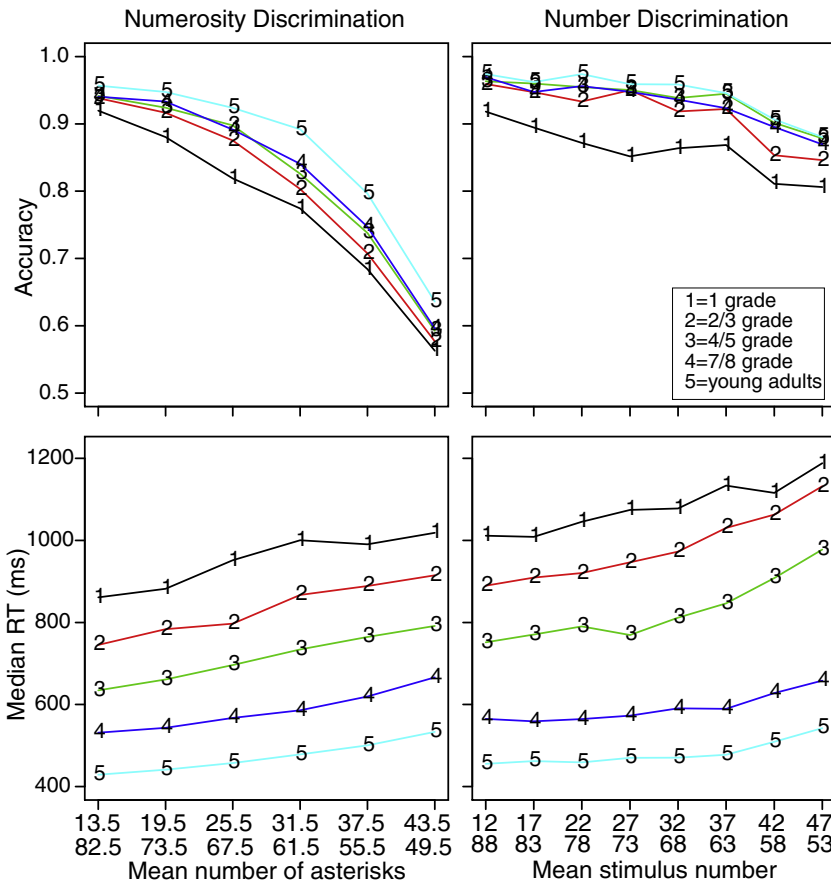


Fig. 2. The top panels are plots of average accuracy across conditions for the numerosity (left) and number (right) discrimination tasks for first, second/third, fourth/fifth, and seventh/eighth graders and adults. The bottom panels are plots of median RT across conditions for the numerosity (left) and number (right) discrimination tasks. As task difficulty increased (e.g., arrays of dots/numbers near 50), participants' accuracy decreased and their RTs increased.

did not differ from one another in median RTs for the number discrimination task, nor did seventh/eighth graders and adults. Fourth/fifth graders, seventh/eighth graders, and adults made faster responses than did first graders and second/third graders. Seventh/eighth graders and adults made faster responses than did fourth/fifth graders. For the numerosity discrimination task, participants in all grades made faster responses than did first graders. Fourth/fifth graders, seventh/eighth graders, and adults made faster responses than did second/third graders. Seventh/eighth graders and adults made faster responses than did fourth/fifth graders. Adults made faster responses than did seventh/eighth graders.

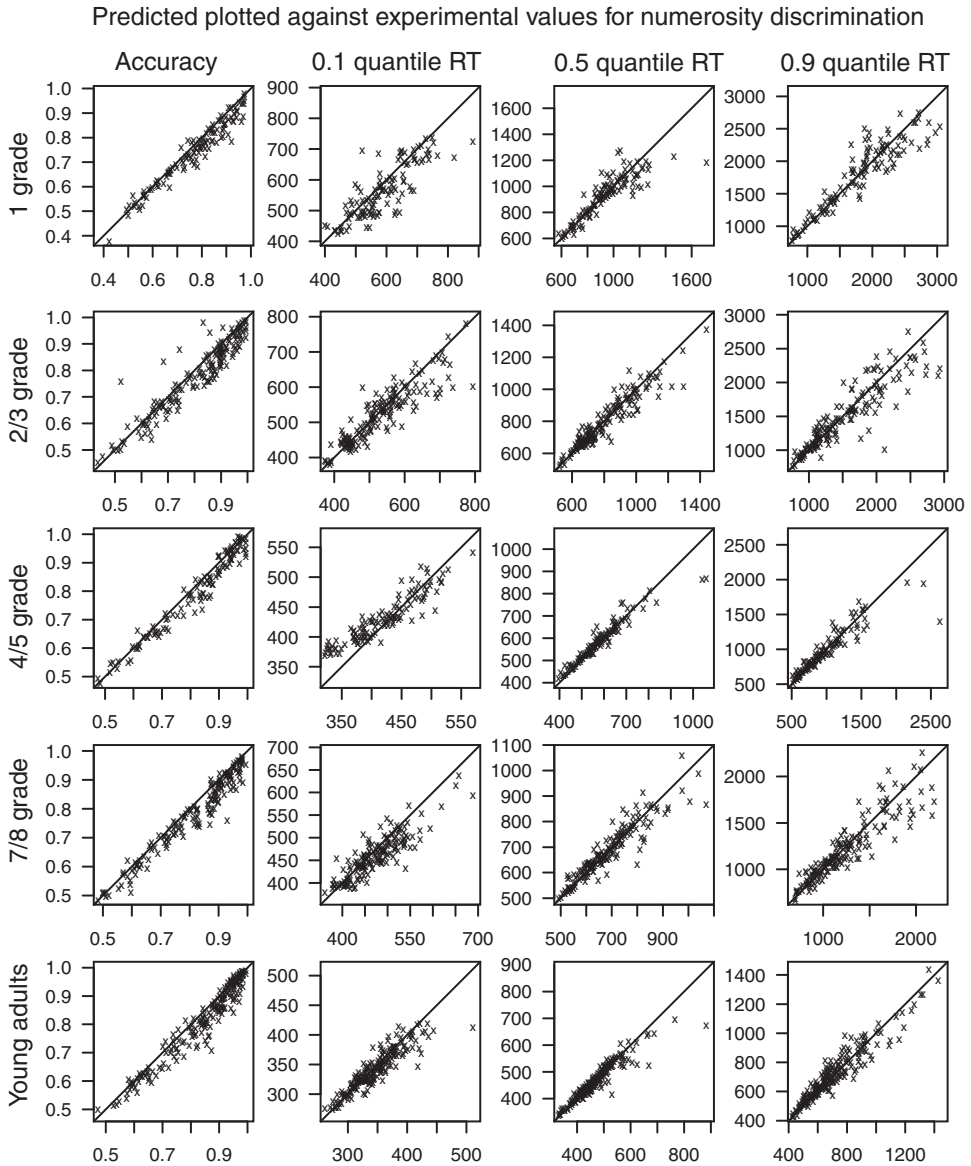


Fig. 3. Model predictions plotted against experimental data for the numerosity discrimination task. Accuracy and RT data from .1, .5, and .9 quantiles are plotted.

The model was fit to the data for each participant individually using a standard chi-square method. The values of all the components of processing identified by the model are estimated simultaneously from the data for all the conditions in an experiment. The fitting method uses quantiles of the RT distributions for correct and error responses for each condition (the .1, .3, .5, .7, and .9 quantile RTs). The model predicts the cumulative probability of a response at each RT quantile. Subtracting the cumulative probabilities for each successive quantile from the next higher quantile gives the proportion of responses between adjacent quantiles. For a chi-square computation, these are the expected values, to be compared with the observed proportions of responses between the quantiles (i.e., the

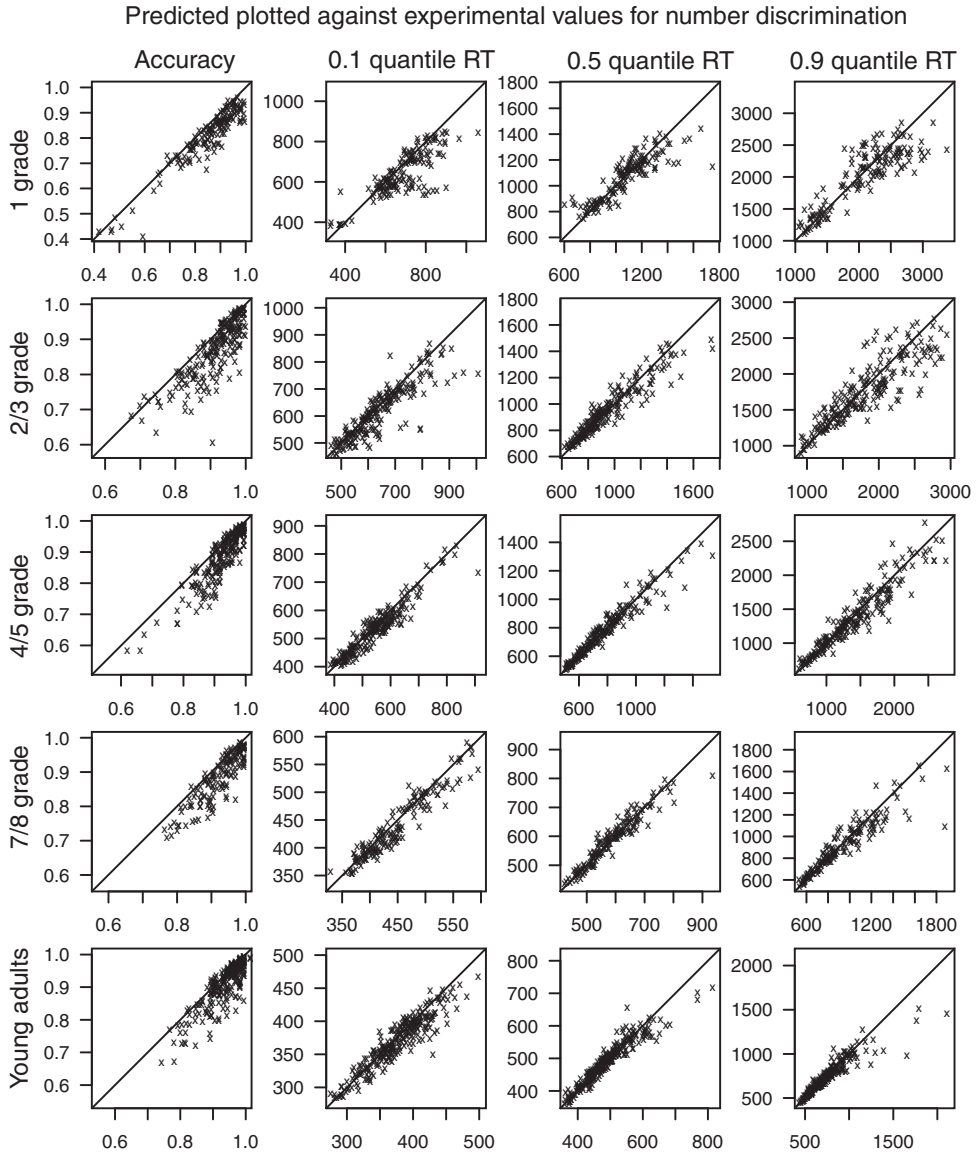


Fig. 4. Model predictions plotted against experimental data for the number discrimination task. Accuracy and RT data from .1, .5, and .9 quantiles are plotted.

Figs. 3 and 4 demonstrate that the model accounted for the data in both the numerosity and number discrimination tasks well. These figures plot the predicted values from the diffusion model against the experimental values for accuracy at the .1, .5, and .9 quantile RTs. A developmental progression is clear for both tasks. The predicted values from the model are a closer fit for older children and adults than for the very youngest participants.

Model-based interpretations of the data

The values of the parameters that generated the best fit of the model to the data averaged over participants are shown in Tables 4 and 5. It should be noted that the mean parameter values across the two tasks were fairly similar. For example, the boundaries that first graders set on the numerosity discrimination task were similar to the boundary separations that they set on the number discrimination task. Likewise, non-decision times were similar across tasks, as were drift rates for equivalent conditions (e.g., the easiest conditions in which numbers or arrays of asterisks were furthest from 50). Similar parameter settings across the two tasks for individuals at each grade level indicate that the tasks are likely tapping a similar underlying numerical ability (see Ratcliff et al., 2015). Statistical analyses comparing the parameter values across grade levels in the numerosity and number discrimination tasks are included in Tables 6 and 7, respectively. First we discuss the parameter values averaged over participants, and then we discuss individual differences in the parameter values.

Values of model parameters averaged across participants

Fig. 5 shows the psychometric functions that relate drift rate (v) to difficulty for the numerosity (top) and number (bottom) discrimination tasks. Average drift rate (y -axis) is plotted against the independent variable of number of asterisks or number shown on the computer screen (x -axis). Linear functions characterize each grade level for both experiments. Performance is near chance for numbers of asterisks around 50, so the intercept of the drift rate function is near zero for the numerosity discrimination task. The intercept is well above zero for the number discrimination task. Children in lower grades had significantly lower drift rates than did children in higher grades and adults on both the numerosity and number discrimination tasks. This indicates that numerical acuity/discrimination abilities increased with increasing grade level. Children in lower grades set wider boundaries (a) than did children in higher grades and adults for the numerosity and number discrimination tasks (see Fig. 6, top left panel). Thus, the decision criteria narrowed with increasing grade level. Children in lower grades had larger non-decision times (T_{er}) than did children in higher grades and adults for both

Table 4
Diffusion model parameters.

| Task | Group | a | T_{er} | η | s_z | s_t | χ^2 |
|------------|-----------|------|----------|--------|-------|-------|----------|
| Numerosity | 1 grade | .195 | .450 | .074 | .102 | .293 | 69.6 |
| | 2/3 grade | .193 | .429 | .125 | .092 | .249 | 64.6 |
| | 4/5 grade | .170 | .396 | .131 | .085 | .248 | 67.1 |
| | 7/8 grade | .154 | .380 | .180 | .085 | .134 | 56.2 |
| | Adult | .117 | .322 | .134 | .067 | .156 | 64.3 |
| Number | 1 grade | .210 | .497 | .034 | .109 | .290 | 99.9 |
| | 2/3 grade | .227 | .462 | .085 | .102 | .209 | 84.8 |
| | 4/5 grade | .194 | .420 | .106 | .085 | .180 | 75.7 |
| | 7/8 grade | .152 | .392 | .124 | .077 | .150 | 73.1 |
| | Adult | .129 | .336 | .099 | .072 | .140 | 64.2 |

Note. The parameters were as follows: boundary separation a (starting point $z = a/2$), mean non-decision component of response time T_{er} , SD in drift across trials η , range of the distribution of starting point s_z , and range of the distribution of non-decision times s_t . Critical values of chi-squares are 67.5 for 50 degrees of freedom for the numerosity discrimination task and 72.2 for 54 degrees of freedom for the number discrimination task.

Table 5Drift rates, v_1 through v_8 correspond to the various conditions that differ in difficulty.

| Task | Group | v_1 | v_2 | v_3 | v_4 | v_5 | v_6 | v_7 | v_8 |
|------------|-----------|-------|-------|-------|-------|-------|-------|-------|-------|
| Numerosity | 1 grade | .193 | .163 | .123 | .095 | .062 | .022 | | |
| | 2/3 grade | .268 | .228 | .194 | .145 | .092 | .042 | | |
| | 4/5 grade | .314 | .265 | .218 | .161 | .110 | .035 | | |
| | 7/8 grade | .414 | .382 | .298 | .222 | .144 | .059 | | |
| | Adult | .482 | .392 | .351 | .249 | .167 | .060 | | |
| Number | 1 grade | .142 | .127 | .116 | .100 | .100 | .102 | .079 | .072 |
| | 2/3 grade | .218 | .210 | .203 | .196 | .166 | .151 | .118 | .108 |
| | 4/5 grade | .304 | .272 | .248 | .255 | .227 | .194 | .147 | .124 |
| | 7/8 grade | .391 | .361 | .370 | .341 | .285 | .290 | .230 | .189 |
| | Adult | .434 | .409 | .414 | .365 | .346 | .305 | .253 | .189 |

the numerosity and number discrimination tasks (see Fig. 6, bottom left panel). That is, with increasing grade, participants were quicker at encoding, transforming the stimulus array, and executing their decisions. There was a significant grade difference in the range in non-decision times (s_e) across trials for both the numerosity and number discrimination tasks (see Fig. 6, bottom left panel). Variability in non-decision time got smaller as grade level increased. There was a significant grade difference in the standard deviation in drift across trials (η) for the numerosity and number discrimination tasks (see Fig. 6, top right panel). This means that the numerical acuity/discrimination became less variable with increasing grade level. There was no statistically significant difference in the range in across-trial variability in starting point (s_z) across grade levels for the numerosity task; however, there was a significant difference in s_z across grade levels for the number discrimination task, with children in lower grades showing greater variability in starting points across trials than adults (see Fig. 6, bottom right panel).

Differences among individuals in data and model parameters

The more accurate the participant was on the numerosity discrimination task, the more accurate the participant was on the number discrimination task (see Table 8), and this was true across grade levels. All of these correlations were significant at the $p < .05$ level except for the performance of first graders. Similarly, the faster a participant responded (mean RT) on the numerosity discrimination task, the faster the participant responded (mean RT) on the number discrimination task (see Table 8). Again, this was true across grades. As reported previously by Ratcliff and colleagues (2015), there were extremely weak correlations between mean RT and accuracy for the numerosity discrimination task for all grades except seventh/eighth in which a moderate, but not statistically significant, correlation was found. Similarly, there were weak correlations between mean RT and accuracy on the number discrimination task for all grades except fourth/fifth in which a moderate statistically significant correlation was found. The most accurate participants were not always the participants who responded the quickest. Importantly, the common belief would be that accuracy and RT are negatively correlated with one another (e.g., accurate participants produce shorter response times). The correlations that were found between accuracy and RT, however, were positive instead of negative. In fact, only 2 of the 10 correlations comparing accuracy and RT on the numerosity and number discriminations tasks were negative (see Table 8).

Fig. 7 shows scatter plots and the correlations between model parameters (e.g., drift rate on the numerosity discrimination task vs. drift rate on the number discrimination task) across the two tasks for each grade. Drift rate was highly correlated across the two tasks for all grade levels. Non-decision time was highly correlated across tasks for first graders and seventh/eighth graders and was moderately so for second/third graders and fourth/fifth graders. Boundary separation was strongly correlated across tasks for first graders and fourth/fifth graders and was moderately correlated for second/third graders and seventh/eighth graders. These results showed that the main components of processing

Table 6

Grade differences in parameter values for numerosity task.

| Parameter | <i>F</i> statistic | Post-hoc <i>t</i> -tests (Bonferroni) | Post-hoc <i>p</i> -values |
|--|---|--|---|
| Drift rate (<i>v</i>) | $F(4,118) = 19.27, p < .0001,$ $\eta^2 = .395$ | First (.093) < Fourth/Fifth (.158) First (.093) < Seventh/Eighth (.221) First (.093) < Adults (.247) Second/Third (.14) < Seventh/Eighth (.221) Second/Third (.14) < Adults (.247) Fourth/Fifth (.158) < Seventh/Eighth (.221) Fourth/Fifth (.158) < Adults (.247) | $p < .05$ $p < .0001$ $p < .0001$ $p < .01$ $p < .0001$ $p < .05$ $p < .0001$ |
| Boundaries (<i>a</i>) | $F(4,118) = 14.83, p < .0001,$ $\eta^2 = .334$ | First (.20) > Seventh/Eighth (.154) First (.20) > Adults (.118) Second/Third (.193) > Seventh/Eighth (.154) Second/Third (.193) > Adults (.118) Fourth/Fifth (.17) > Adults (.118) Seventh/Eighth (.154) > Adults (.118) | $p < .05$ $p < .0001$ $p < .05$ $p < .0001$ $p < .0001$ $p = .051$ |
| Non-decision times (<i>T_{er}</i>) | $F(4,118) = 29.96, p < .0001,$ $\eta^2 = .504$ | First (.45) > Fourth/Fifth (.396) First (.45) > Seventh/Eighth (.38) First (.45) > Adults (.323) Second/Third (.429) > Seventh/Eighth (.38) Second/Third (.429) > Adults (.323) Fourth/Fifth (.396) > Adults (.323) Seventh/Eighth (.38) > Adults (.323) | $p < .01$ $p < .0001$ $p < .0001$ $p < .01$ $p < .0001$ $p < .0001$ $p < .0001$ |
| Range in non-decision times across trials (<i>s_r</i>) | $F(4,118) = 19.63, p < .0001,$ $\eta^2 = .40$ | First (.293) > Seventh/Eighth (.134) First (.293) > Adults (.154) Second/Third (.249) > Seventh/Eighth (.134) Second/Third (.249) > Adults (.154) Fourth/Fifth (.248) > Seventh/Eighth (.134) Fourth/Fifth (.248) > Adults (.154) | $p < .0001$ $p < .0001$ $p < .0001$ $p < .0001$ $p < .0001$ $p < .0001$ |
| Standard deviation in drift rates across trials (<i>η</i>) | $F(4,118) = 3.16, p < .05,$ $\eta^2 = .096$ | First (.074) < Seventh/Eighth (.18) | $p < .01$ |
| Range of starting points across trials (<i>s_z</i>) | $F(4,118) = 1.86, p > .05$ | | |

were related across tasks. If an individual had a high drift rate, large boundary separation, or long non-decision time on one task, the individual tended to have a high drift rate, large boundary separation, or long non-decision time on the other task.

The diffusion model parameters were not highly correlated with one another, and this was true across grade level (see Table 9). That is, boundary settings were not strongly correlated with non-decision times or drift rates, and non-decision times were not strongly correlated with drift rates. The only exception was that for first graders non-decision time and drift rate were significantly correlated. These results are consistent with those reported by Ratcliff and colleagues (2015), in which it was suggested that the diffusion model decomposes the dependent variables—accuracy and correct and error RT distributions—into separate and relatively independent components of processing.

Table 7
Grade differences in parameter values for number task.

| Parameter | F statistic | Post-hoc <i>t</i> -tests (Bonferroni) | Post-hoc <i>p</i> -values |
|--|---|---|---------------------------|
| Drift rate (v) | $F(4,118) = 37.77$, $p < .0001$, $\eta^2 = .561$ | First (.10) < Second/Third (.165) | $p < .05$ |
| | | First (.10) < Fourth/Fifth (.21) | $p < .0001$ |
| | | First (.10) < Seventh/Eighth (.295) | $p < .0001$ |
| | | First (.10) < Adults (.326) | $p < .0001$ |
| | | Second/Third (.165) < Seventh/Eighth (.295) | $p < .0001$ |
| | | Second/Third (.165) < Adults (.326) | $p < .0001$ |
| | | Fourth/Fifth (.21) < Seventh/Eighth (.295) | $p < .01$ |
| | | Fourth/Fifth (.21) < Adults (.326) | $p < .0001$ |
| Boundaries (a) | $F(4,118) = 19.8$, $p < .0001$, $\eta^2 = .402$ | First (.211) > Seventh/Eighth (.152) | $p < .01$ |
| | | First (.211) > Adults (.129) | $p < .0001$ |
| | | Second/Third (.227) > Seventh/Eighth (.152) | $p < .0001$ |
| | | Second/Third (.227) > Adults (.129) | $p < .0001$ |
| | | Fourth/Fifth (.194) > Seventh/Eighth (.152) | $p < .05$ |
| Fourth/Fifth (.194) > Adults (.129) | $p < .0001$ | | |
| Non-decision times (T_{er}) | $F(4,118) = 33.6$, $p < .0001$, $\eta^2 = .532$ | First (.497) > Fourth/Fifth (.42) | $p < .0001$ |
| | | First (.497) > Seventh/Eighth (.392) | $p < .0001$ |
| | | First (.497) > Adults (.336) | $p < .0001$ |
| | | Second/Third (.462) > Seventh/Eighth (.392) | $p < .0001$ |
| | | Second/Third (.462) > Adults (.336) | $p < .0001$ |
| | | Fourth/Fifth (.42) > Adults (.336) | $p < .0001$ |
| | | Seventh/Eighth (.392) > Adults (.336) | $p < .01$ |
| Range in non-decision times across trials (s_t) | $F(4,118) = 9.73$, $p < .0001$, $\eta^2 = .248$ | First (.29) > Second/Third (.209) | $p < .05$ |
| | | First (.29) > Fourth/Fifth (.18) | $p < .001$ |
| | | First (.29) > Seventh/Eighth (.15) | $p < .0001$ |
| | | First (.29) > Adults (.14) | $p < .0001$ |
| | | Second/Third (.209) > Adults (.14) | $p < .05$ |
| Standard deviation in drift rates across trials (η) | $F(4,118) = 4.05$, $p < .01$, $\eta^2 = .121$ | First (.034) < Fourth/Fifth (.106) | $p < .05$ |
| | | First (.034) < Seventh/Eighth (.124) | $p < .01$ |
| | | First (.034) < Adults (.099) | $p < .05$ |
| Range of starting points across trials (s_z) | $F(4,118) = 4.02$, $p < .01$, $\eta^2 = .119$ | First (.109) > Adults (.072) | $p < .05$ |
| | | Second/Third (.102) > Adults (.072) | $p < .05$ |

Table 10 shows correlations among the behavioral measures, RT and accuracy, and model parameters. Our previous work with adults (Ratcliff et al., 2015) showed that accuracy is mainly determined by drift rate and RT is mainly determined by boundary settings, and that was replicated across grade levels in our current experiments.

Discussion

Previous research (Ratcliff et al., 2012) showed a developmental progression in children's simple two-choice non-symbolic numerical discrimination decisions. For example, children accumulated a lower quality of evidence than did adults when attempting to make a non-symbolic numerical decision, set wider decision criteria than did adults, and were slower at encoding the stimulus array, transforming the array into task-relevant information, and executing a motor response than were adults. One goal of our current experiments was to determine whether this pattern of results could be replicated with a new group of children and extended to a symbolic number discrimination task. We also investigated whether RT and accuracy on the symbolic and non-symbolic discrimination tasks was related across individuals. Indeed, our current results provide evidence for a developmental progression in diffusion model parameters toward higher drift rates, narrower boundary settings, and faster non-decision components across the lifespan in both non-symbolic and symbolic tasks. In contrast, previous research found that only boundary separation and non-decision criteria differed between

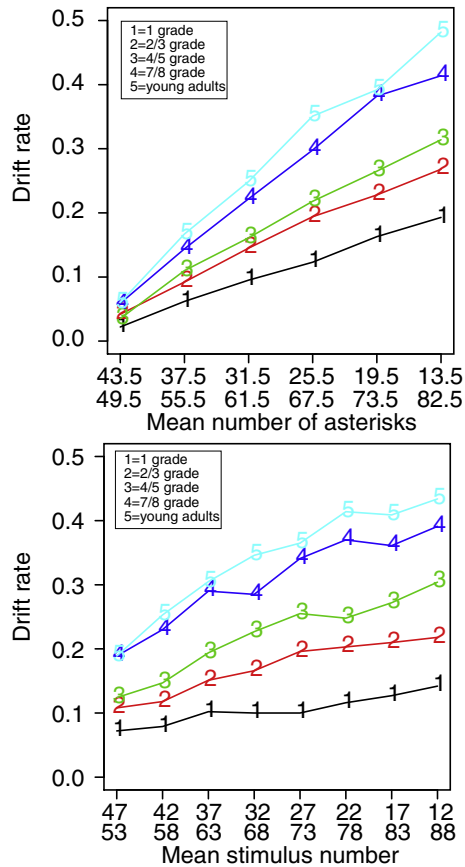


Fig. 5. Drift rate plotted against number of asterisks (numerosity discrimination task) and number (number discrimination task). Drift rate increased as condition difficulty became easier (e.g., numbers further from 50).

aging adults and college-aged students who completed a non-symbolic discrimination task (Ratcliff et al., 2001). Our current findings indicate that with increasing age, children's symbolic and non-symbolic magnitude knowledge improves and results in higher drift rates.

RTs decreased and accuracy increased from the hardest conditions (comparisons closest to 50) to the simplest conditions (comparisons furthest from 50), and this provided evidence for the numerical distance effect (Dehaene et al., 1998) in both the symbolic and non-symbolic discrimination tasks. Across age ranges, participants were at ceiling levels of accuracy on the number discrimination task and also on the simplest conditions in the numerosity discrimination task. By the time children are in first grade, they have already acquired some adult-like abilities in comparing Arabic numerals and asterisk arrays.

We found (in line with Ratcliff et al., 2015) that RT was strongly correlated across the symbolic number discrimination task and the non-symbolic numerosity discrimination task, as was accuracy. This means that participants who made fast responses on one task tended to make fast responses on the other task, and participants who made accurate decisions on one task made accurate decisions on the other task. Diffusion model drift rate, boundary separation, and non-decision time parameters were also strongly correlated with the same model parameter on the two tasks. For example, a participant who had a high drift rate (accuracy) on one task was able to abstract high-quality information from

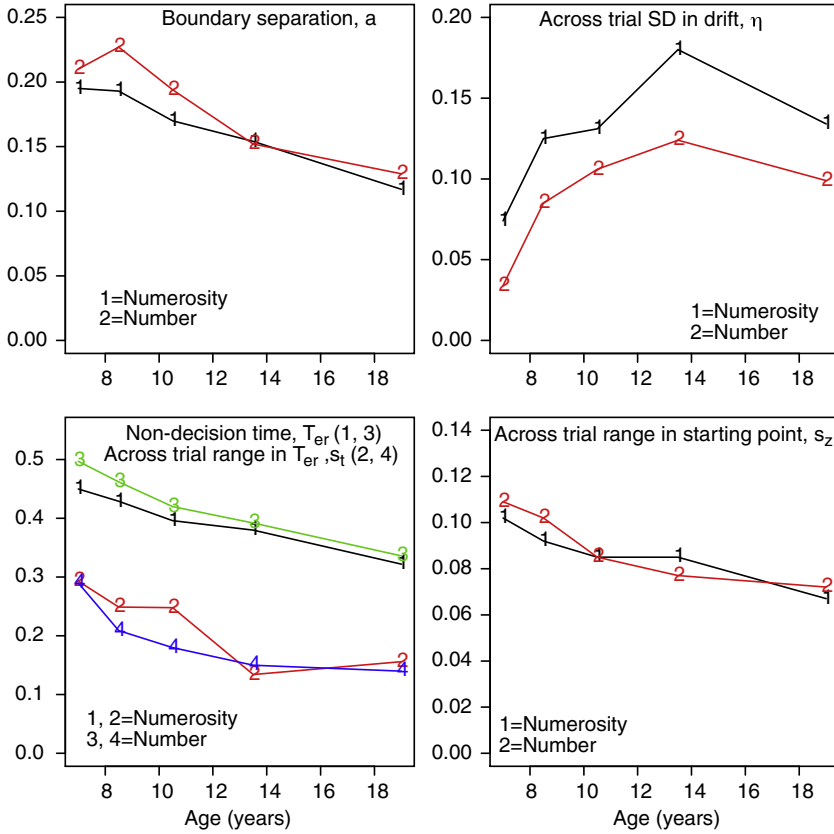


Fig. 6. Developmental differences in boundary separation (upper left panel), SD in drift rate (upper right panel), non-decision time and across-trial range in non-decision time (bottom left panel), and across-trial range in starting point (bottom right panel).

Table 8
Correlations for accuracy and mean RT averaged across conditions.

| Group | Accuracy–accuracy | Mean RT–mean RT | Numerosity accuracy–mean RT | Number accuracy–mean RT |
|-----------|-------------------|-----------------|-----------------------------|-------------------------|
| 1 grade | .372 | .667 | –.049 | .335 |
| 2/3 grade | .609 | .822 | .045 | –.236 |
| 4/5 grade | .341 | .838 | .187 | .480 |
| 7/8 grade | .762 | .550 | .404 | .299 |
| Adult | .634 | .592 | .078 | .205 |

Note. Critical values for the correlations are as follows: first grade, .46; second/third grade, .39; fourth/fifth grade, .38; seventh/eighth grade, .46; and adult, .35.

the stimulus array on the other task, and a participant with conservative boundary settings on one task tended to set conservative decision criteria on the other task.

In contrast, RT and accuracy were not correlated with each other within a task so that whether a participant was fast did not predict whether the participant was accurate. In addition, drift rate, boundary setting, and non-decision time parameters were not correlated with each other within a

Correlations between the numerosity and number discrimination tasks

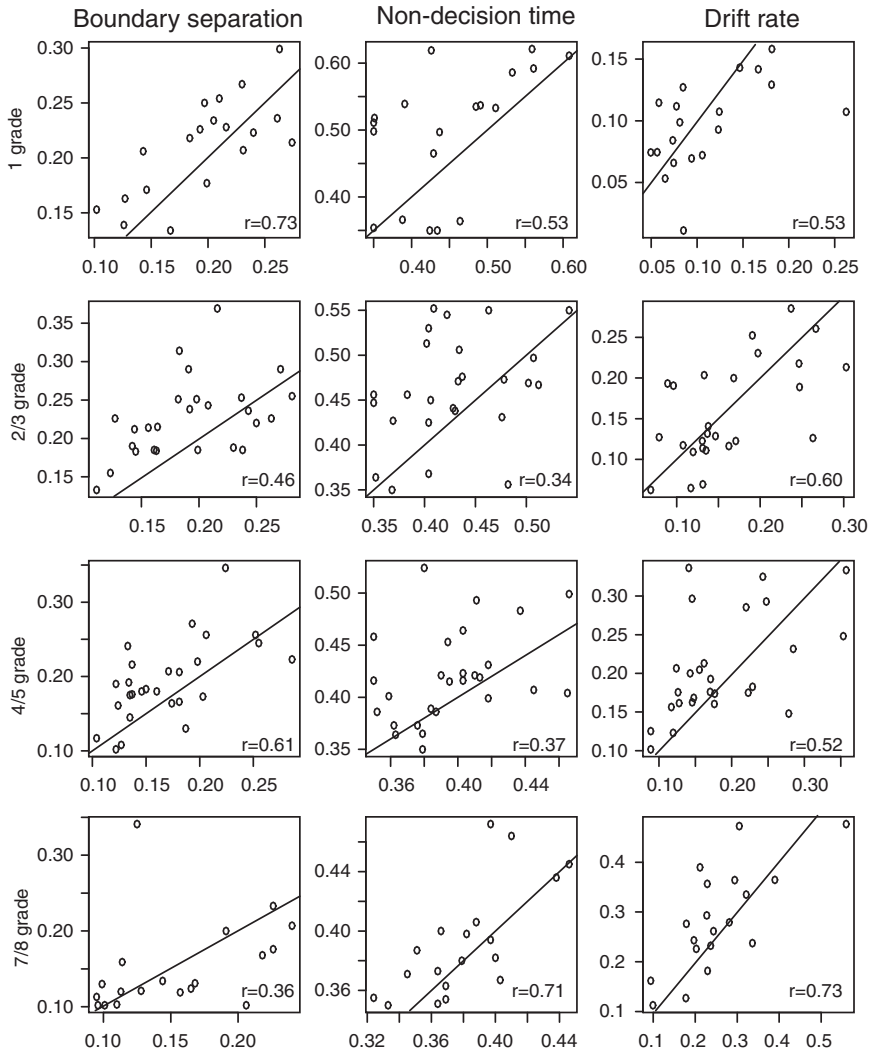


Fig. 7. Plots of model parameters between the numerosity and number discrimination tasks for boundary separation (left column), non-decision time (middle column), and drift rate (right column). The diagonal lines have slope 1 and intercept 0. The critical values for the correlations are as follows: first graders = .46, second/third graders = .39, fourth/fifth graders = .38, seventh/eighth graders = .46.

task. This provided an explanation within the diffusion model framework of the lack of correlation between RT and accuracy (see Ratcliff et al., 2015). That is, the model successfully differentiates the quality of information on which participants base their decisions from their willingness to make decisions and other components of information processing such as their ability to encode and transform a stimulus and execute a decision.

Our results may help to reconcile discrepant findings in the numerical cognition literature. Sometimes there are significant correlations across symbolic and non-symbolic tasks in the numerical

Table 9
Correlations between model parameters averaged across the two tasks.

| Group | a/T_{er} | a/v | T_{er}/v |
|-----------|------------|-------|------------|
| 1 grade | -.40 | -.24 | .51 |
| 2/3 grade | -.23 | .21 | -.08 |
| 4/5 grade | -.27 | .09 | .09 |
| 7/8 grade | -.17 | .27 | .19 |
| Adult | .16 | -.13 | .18 |

Note. Critical values for the correlations are as follows: first grade, .46; second/third grade, .39; fourth/fifth grade, .38; seventh/eighth grade, .46; and adult, .35.

cognition literature, and sometimes there are not. Fazio and colleagues' (2014) meta-analysis indicated the circumstances under which correlations between non-symbolic numerical discrimination tasks and mathematics achievement are likely to be found. There were stronger correlations between non-symbolic task performance and mathematics achievement when accuracy-based measures or accuracy- and RT-based measures of non-symbolic task performance were combined rather than when RT-based measures of non-symbolic task performance were considered alone.

Therefore, the discrepancy in the numerical cognition literature may at least partly stem from the fact that researchers primarily use either accuracy- or RT-based measures, but rarely both, to describe performance on numerical decision tasks. The classic study of symbolic numerical discrimination conducted by Sekuler and Mierkiewicz (1977) is a prime example in which changes in RT only were tracked across age cohorts. Our results point to the fact that accuracy and RT vary independently and may stem from different underlying processes; we found no evidence in support of the common belief that accurate participants should make faster responses relative to the average values for the group. Therefore, basing claims about participants' numerical decision making on either of these dependent variables alone may lead to a skewed understanding of children's and adults' numerical information processing abilities. The explanation for the lack of negative correlations between RT and accuracy within the framework of the diffusion model was that drift rates, boundary settings, and non-decision times are independent of one another, accuracy is mainly determined by drift rate, and RT is mainly determined by boundary settings. One advantage of modeling the data with the diffusion model is that it accounts for the quality of the accumulated information (drift rate) separately from speed/accuracy settings (boundary separation).

Ceiling- and floor-level performance may also contribute to the inconsistent findings in the numerical cognition literature. That is, if all participants make highly accurate decisions but there is some variability in RTs, researchers may then choose to report RTs instead of accuracy because of the increased variability in the RT data as compared with the accuracy data. Drift rates can be obtained for participants when their RTs vary even though their accuracy may be at ceiling or floor levels.

Our current results indicate that an information-processing model, such as the diffusion model, that decomposes simple two-choice decisions into an encoding/response execution component, an evidence accumulation component, and speed/accuracy settings can inform theories about numerical decision making across the lifespan. Our model-based analysis of children's performance on symbolic and non-symbolic numerical decisions suggests evidence that the number and numerosity discrimination tasks are tapping participants' underlying numerical ability. Future research should investigate whether diffusion model parameters are correlated with mathematics achievement test scores and IQ scores. These analyses could point to the mechanisms by which basic numeracy skills, such as deciding whether an Arabic numeral is greater than or less than 50, are related to more complex mathematical abilities indexed on achievement tests. We were unable to obtain achievement test scores for all of

Table 10
Correlations of model parameters and accuracy and RT across grades.

| | Accuracy for numerosity discrimination | | | Accuracy for number discrimination | | | RT for numerosity discrimination | | | RT for number discrimination | | |
|-----------|--|-----------------------|----------|------------------------------------|-----------------------|----------|----------------------------------|-----------------------|----------|------------------------------|-----------------------|----------|
| | <i>a</i> | <i>T_{er}</i> | <i>v</i> | <i>a</i> | <i>T_{er}</i> | <i>v</i> | <i>a</i> | <i>T_{er}</i> | <i>v</i> | <i>a</i> | <i>T_{er}</i> | <i>V</i> |
| 1 grade | -0.013 | 0.464 | 0.613 | 0.165 | 0.414 | 0.814 | 0.843 | -0.316 | -0.581 | 0.632 | 0.095 | -0.205 |
| 2/3 grade | 0.055 | -0.067 | 0.419 | 0.116 | 0.299 | 0.542 | 0.749 | 0.172 | -0.421 | 0.227 | 0.331 | -0.652 |
| 4/5 grade | 0.51 | -0.237 | 0.474 | 0.554 | -0.25 | -0.068 | 0.536 | -0.027 | -0.367 | 0.882 | -0.048 | -0.637 |
| 7/8 grade | 0.48 | -0.154 | 0.56 | 0.46 | 0.017 | 0.456 | 0.79 | -0.058 | -0.148 | 0.705 | 0.225 | -0.297 |
| Adult | 0.431 | -0.117 | 0.535 | 0.372 | 0.248 | 0.309 | 0.84 | 0.427 | -0.335 | 0.835 | 0.294 | -0.585 |

our participants because the very youngest children were not state-mandated to take achievement tests, and not all parents gave permission for their children's scores to be released for research purposes. Given findings from [Chen and Li's \(2014\)](#) meta-analysis, we were concerned that any significant or non-significant correlations that we found between non-symbolic task performance and mathematics achievement would be underpowered in our sample. One "proof of existence" example from the domain of reading illustrates that in a large individual differences study with struggling adult readers, drift rates from a lexical decision task were correlated with standardized reading scores and IQ scores ([McKoon & Ratcliff, 2016](#)). Future research should involve the testing of many more children from each grade level to establish whether there are also significant relations among model parameters for symbolic and non-symbolic numerical decision tasks and standardized achievement/IQ scores in the domain of mathematics. This could be especially important to tease apart the reasons why so many first graders were unable to complete the tasks (e.g., general lack of ability vs. task difficulty vs. boredom).

It is an open question as to whether diffusion model parameters that model non-symbolic discrimination data are related to overall mathematics achievement. Like [Fazio and colleagues \(2014\)](#), we hypothesize that simple two-choice symbolic and non-symbolic numerical decision tasks may tap more rudimentary information processing skills than the more complex variety of skills necessary to answer questions on standardized achievement tests and IQ assessments. Standardized tests of mathematics achievement may test a range of skills, some of which are related to basic numerical abilities and some of which are unrelated to numerical abilities. Future research could focus on cataloging the constructs being tested on standardized achievement tests such that the same basic underlying processes at play in the achievement tests can be more directly compared with similar cognitive tasks.

Another step for future research will be to use stimuli that control for continuous extent (e.g., yellow and blue dot arrays available from <http://panamath.org>) to determine whether diffusion model parameters differ when these aspects of the stimuli are controlled. Pilot data that we have collected in our labs with college-aged adults have shown that drift rates for non-symbolic tasks that do and do not control for continuous extent are correlated. In addition, drift rates on the symbolic and non-symbolic tasks used in the current experiments are highly correlated for participants across all of the grades we tested; therefore, these tasks are likely tapping a common underlying numerical ability.

Acknowledgments

The authors thank Brian Pickens, Mary Rischar, Sarah Wiatrek, and Heather Smith for their help with data collection and thank Ashley Stafford and Craig Cruzan for their help in administering the WISC-IV subtests.

Appendix A

Average achievement test and IQ scores.

| Grade | Achievement test and maximum possible score | Average score and <i>SD</i> | Percentage at or above grade level proficiency | Average matrix reasoning and vocabulary IQ and <i>SD</i> |
|----------------|---|--------------------------------|--|--|
| First | BEAR, 41 | 30.11, 5.54 | 89 | |
| Second/Third | BEAR, 35 | 28.77, 7.22 | 73 | 115.24, 12.28 |
| Fourth/Fifth | OCCT Math, 990 OCCT Reading, 990 | 854.74, 76.18 822.35, 61.81 | 100 96 | 117, 12.81 |
| Seventh/Eighth | OCCT Math, 990 OCCT Reading, 990 | 804.47, 55.15 817.94, 62.59 | 100 94 | |

Note. BEAR, Basic Early Assessment of Reading; OCCT, Oklahoma Core Curriculum Tests.

References

- 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.
- De Smedt, B., Verschaffel, L., & Ghesquiere, P. (2009). The predictive value of numerical magnitude comparison for individual differences in mathematics achievement. *Journal of Experimental Child Psychology, 103*, 469–479.
- Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neuroscience, 21*, 355–361.
- Dehaene, S., Dupoux, E., & Mehler, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 626–641.
- Durand, M., Hulme, C., Larkin, R., & Snowling, M. (2005). The cognitive foundations of reading and arithmetic skills in 7- to 10-year-old children. *Journal of Experimental Child Psychology, 91*, 113–136.
- Fazio, L. K., Bailey, D. H., Thompson, C. A., & Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *Journal of Experimental Child Psychology, 123*, 53–72.
- Gilmore, C., Attridge, N., & Inglis, M. (2011). Measuring the approximate number system. *Quarterly Journal of Experimental Psychology, 64*, 2099–2109.
- Gilmore, C. K., McCarthy, S. W., & Spelke, E. S. (2010). Non-symbolic arithmetic abilities and mathematics achievement in the first year of formal schooling. *Cognition, 115*, 394–406.
- Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., & Germine, L. (2012). Number sense across the lifespan as revealed by a massive internet-based sample. *Proceedings of the National Academy of Sciences of the United States of America, 109*, 1116–1120.
- Halberda, J., Mazocco, M. M. M., & Feigenson, L. (2008). Individual differences in nonverbal number acuity predict maths achievement. *Nature, 455*, 665–668.
- Holloway, I. D., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children's math achievement. *Journal of Experimental Child Psychology, 103*, 17–29.
- Inglis, M., Attridge, N., Batchelor, S., & Gilmore, C. K. (2011). Non-verbal number acuity correlates with symbolic mathematics achievement: But only in children. *Psychonomic Bulletin & Review, 18*, 1222–1229.
- Laming, D. R. J. (1968). *Information theory of choice reaction time*. New York: John Wiley.
- Libertus, M. E., Feigenson, L., & Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Developmental Science, 14*, 1292–1300.
- Lyons, I. M., & Beilock, S. L. (2011). Mathematics anxiety: Separating the math from the anxiety. *Cerebral Cortex, 22*, 2102–2110.
- Maloney, E., Risko, E., Preston, F., Ansari, D., & Fugelsang, J. (2010). Challenging the reliability and validity of cognitive measures: The case of the numerical distance effect. *Acta Psychologica, 134*, 154–161.
- Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011a). Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child Development, 82*, 1224–1237.
- Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011b). Preschoolers' precision of the approximate number system predicts later school mathematics performance. *PLoS ONE, 6*(9), e23749.
- McKoon, G., & Ratcliff, R. (2016). Adults with poor reading skills: How lexical knowledge interacts with scores on standardized reading comprehension tests. *Cognition, 146*, 453–469.
- Mundy, E., & Gilmore, C. (2009). Children's mapping between symbolic and nonsymbolic representation of number. *Journal of Experimental Child Psychology, 103*, 490–502.
- Prather, R. W. (2014). Numerical discrimination is mediated by neural coding variation. *Cognition, 133*, 601–610.
- Price, G. R., Palmer, D., Battista, C., & Ansari, D. (2012). Nonsymbolic numerical magnitude comparison: Reliability and validity of different task variants and outcome measures, and their relationship to arithmetic achievement in adults. *Acta Psychologica, 140*, 50–57.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*, 59–108.

- Ratcliff, R. (2002). A diffusion model account of reaction time and accuracy in a two choice brightness discrimination task: Fitting real data and failing to fit fake but plausible data. *Psychonomic Bulletin & Review*, 9, 278–291.
- Ratcliff, R. (2013). Parameter variability and distributional assumptions in the diffusion model. *Psychological Review*, 120, 281–292.
- Ratcliff, R. (2014). Measuring psychometric functions with the diffusion model. *Journal of Experimental Psychology: Human Perception and Performance*, 40, 870–888.
- Ratcliff, R., & Childers, R. (2015). Individual differences and fitting methods for the two-choice diffusion model. *Decision*, 2, 237–279.
- Ratcliff, R., Love, J., Thompson, C. A., & Opfer, J. (2012). Children are not like older adults: A diffusion model analysis of developmental changes in speeded responses. *Child Development*, 83, 367–381.
- Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, 20, 873–922.
- Ratcliff, R., Thapar, A., & McKoon, G. (2001). The effects of aging on reaction time in a signal detection task. *Psychology and Aging*, 16, 323–341.
- Ratcliff, R., Thapar, A., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on brightness discrimination. *Perception & Psychophysics*, 65, 523–535.
- Ratcliff, R., Thapar, A., & McKoon, G. (2004). A diffusion model analysis of the effects of aging on recognition memory. *Journal of Memory and Language*, 50, 408–424.
- Ratcliff, R., Thapar, A., & McKoon, G. (2006). Aging and individual differences in rapid two-choice decisions. *Psychonomic Bulletin & Review*, 13, 626–635.
- Ratcliff, R., Thapar, A., & McKoon, G. (2010). Individual differences, aging, and IQ in two-choice tasks. *Cognitive Psychology*, 60, 127–157.
- Ratcliff, R., Thapar, A., & McKoon, G. (2011). Effects of aging and IQ on item and associative memory. *Journal of Experimental Psychology: General*, 140, 464–487.
- Ratcliff, R., Thompson, C. A., & McKoon, G. (2015). Modeling individual differences in response time and accuracy in numeracy. *Cognition*, 137, 115–136.
- Ratcliff, R., & Tuerlinckx, F. (2002). Estimating the parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review*, 9, 438–481.
- Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review*, 106, 261–300.
- Sasanguie, D., Defever, E., Van den Bussche, E., & Reynvoet, B. (2011). The reliability of and the relation between non-symbolic numerical distance effects in comparison, same–different judgments, and priming. *Acta Psychologica*, 136, 73–80.
- Sekuler, R., & Mierkiewicz, D. (1977). Children's judgments of numerical inequality. *Child Development*, 48, 630–633.
- Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. New York: Oxford University Press.
- Thapar, A., Ratcliff, R., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on letter discrimination. *Psychology and Aging*, 18, 415–429.
- White, C. N., Ratcliff, R., Vasey, M. W., & McKoon, G. (2010). Using diffusion models to understand clinical disorders. *Journal of Mathematical Psychology*, 54, 39–52.