Journal of Experimental Psychology: General
2015, Vol. 144, No. 5, 1021–1035

0096-3445/15/$12.00 http://dx.doi.org/10.1037/xge0000094

Rethinking the Implications of Numerical Ratio Effects for Understanding the Development of Representational Precision and Numerical Processing Across Formats

Ian M. Lyons
University of Western Ontario

Hans-Christoph Nuerk
University of Tübingen; Knowledge Media Institute IWM-KMRC, Tübingen, Germany

Daniel Ansari
University of Western Ontario

Numerical ratio effects are a hallmark of numerical comparison tasks. Moreover, ratio effects have been used to draw strong conclusions about the nature of numerical representations, how these representations develop, and the degree to which they generalize across stimulus formats. Here, we compute ratio effects for 1,719 children from Grades K–6 for each individual separately by computing not just the average ratio effect for each person, but also the variability and statistical magnitude (effect-size) of their ratio effect. We find that individuals' ratio effect-sizes in fact increase over development, calling into question the view that decreasing ratio effects over development indicate increasing representational precision. Our data also strongly caution against the use of ratio effects in inferring the nature of symbolic number representation. While 75% of children showed a statistically significant ratio effect for nonsymbolic comparisons, only 30% did so for symbolic comparisons. Furthermore, whether a child's nonsymbolic ratio effect was significant did not predict whether the same was true of their symbolic ratio effect. These results undercut the notions (a) that individuals' ratio effects are indicative of representational precision in symbolic numbers, and (b) that a common process generates ratio effects in symbolic and nonsymbolic formats. Finally, for both formats, it was the variability of an individual child's ratio effect (not its slope or even effect-size) that correlated with arithmetic ability. Taken together, these results call into question many of the long-held tenets regarding the interpretation of ratio effects—especially with respect to symbolic numbers.

Keywords: numerical ratio effect, numerical comparison, numerical development, arithmetic, individual differences

Supplemental materials: http://dx.doi.org/10.1037/xge0000094.supp

Numerical comparison tasks, in which the participant decides which of two numerical stimuli represents the greater quantity, have been a mainstay in the field of numerical cognition for nearly half a century (Moyer & Landauer, 1967), and have significantly informed current thinking about the nature and development of numerical processing. Numerical comparison tasks are presented with either symbolic (e.g., Arabic numerals) or nonsymbolic (e.g., dot arrays) representations of numerical quantity. It has long been recognized that the numerical ratio between the two stimuli is an important predictor of performance on these tasks: when the ratio is closer to 1, performance tends to be worse (typically, longer response-times and higher error-rates; e.g., Moyer & Landauer, 1967; Buckley & Gillman, 1974). This relation between ratio and performance (hereafter, simply "ratio effect") has been a staple assumption for many number researchers, and its ubiquity is often all but taken for granted.[1]

Of particular interest is the fact that ratio effects have been used to draw major theoretical conclusions about the nature of number

This article was published Online First July 13, 2015.

Ian M. Lyons, Department of Psychology, Brain and Mind Institute, University of Western Ontario; Hans-Christoph Nuerk, Department of Psychology and LEAD Graduate School, University of Tübingen and Knowledge Media Research Institute IWM-KMRC, Tübingen, Germany; Daniel Ansari, Department of Psychology, Brain and Mind Institute, University of Western Ontario.

This research was supported by funding from the Canadian Institutes of Health Research (CIHR), The National Sciences and Engineering Research Council of Canada (NSERC), and Canada Research Chairs program (CRC) to D.A., and by funding from the DFG (German Research Foundation) and the Science Campus Tuebingen to Hans-Christoph Nuerk. Data collection costs were paid in part by Boom Test Uitgevers Amsterdam BV.

Correspondence concerning this article should be addressed to Daniel Ansari, Department of Psychology, The University of Western Ontario, Westminster Hall, Room 325, London, ON, N6A 3K7, Canada. E-mail: daniel.ansari@uwo.ca

[1] Here it is worth pointing out that many studies compute instead a close cousin of the ratio effect, referred to as the "distance effect" (where *distance* $= |n_1 - n_2|$). It is important to note that *distance* and *ratio* are closely related. For example, for the range 1 to 10 (exhaustive sampling of all nonequal combinations), where *distance* $= |n_1 - n_2|$, and $ratio = \frac{min(n_1, n_2)}{max(n_1, n_2)}$, *distance* and *ratio* are correlated at $R^2 = .71$. That said, we prefer *ratio* here over *distance* because the former implicitly takes into account the average size of the two numbers in addition to the difference between them.

representation (Dehaene, 1997, 2008; Dehaene & Changeux, 1993; Feigenson et al., 2004; Halberda & Feigenson, 2008; Nieder & Dehaene, 2009; Verguts & Fias, 2004; though see also Verguts et al., 2005; van Opstal et al., 2008, for alternative explanations in certain cases). Nearly all of these conclusions hinge on the notion of *representational precision*. Specifically, individual numbers are thought to be represented in an overlapping or interfering manner whereby the degree of overlap or interference systematically increases as a function of the ratio between the numbers in question (Dehaene & Changeux, 1993; Nieder & Dehaene, 2009; Verguts & Fias, 2004). In this view, the ratio effect is an index of this underlying precision: a smaller ratio effect indicates a more precise underlying representation, and a larger ratio effect indicates a less precise representation.

From this assumption in turn, at least three major claims have been made regarding the theoretical import of numerical ratio effects. First, the magnitude of ratio effects tends to decrease over development, which has been interpreted as evidence for increasing precision over development of the underlying numerical representations (e.g., Halberda & Feigenson, 2008; Holloway & Ansari, 2009; Sekuler & Mierkiewicz, 1977). Second, both symbolic and nonsymbolic comparisons have been shown to yield ratio effects, a fact that has been used to argue that the two formats are derived from a common underlying representation (e.g., Dehaene, 2008). Third, individual differences in distance and ratio effects have recently become popular correlates for other, typically more complex numerical tasks such as mental arithmetic and math achievement (e.g., Bonny & Lourenco, 2013; Bugden et al., 2012; De Smedt et al., 2009; Feigenson et al., 2013; Fuhs & McNeil, 2013; Halberda et al., 2008, 2012; Holloway & Ansari, 2009; Libertus et al., 2011; Lonnemann et al., 2011; Lyons & Beilock, 2011; Mazzocco et al., 2011; Piazza et al., 2010; Sasanguie et al., 2012). These studies have found that smaller distance and ratio effects are associated with relatively higher performance on standardized test of arithmetic. The interpretation of such a relationship is that smaller ratio/distance effects represent more precise representations of numerical magnitude, which in turn are associated with relatively better performance on formal tests of arithmetic/math achievement.

Given the central position ratio effects take in shaping current theoretical accounts of numerical representation, its development, and potential implications for math education (e.g., De Smedt et al., 2013), a deeper characterization of these effects is warranted. In other words, just how valid are these inferences about number representation based on the observation of ratio effects? Current methods for assessing ratio (and related) effects typically rely on an individual's average ratio effect (i.e., the raw slope, or the raw difference between "hard" and "easy" ratios), and then compute population level statistics based on this average. Note that this approach implicitly takes the view that there is some underlying principle or tendency that governs human behavior (ratio drives comparison performance), and individual variation around this central governing tendency is essentially noise. For this to be valid, however, it is crucial to obtain evidence that this principle is indeed influencing a given individual's performance to begin with. More concretely, as noted above, researchers have interpreted small ratio effects as evidence for highly precise numerical representation. Another interpretation, however, is that ratio has no reliable influence over some or the majority of individuals' performance to begin with. Notice that from an individual differences perspective in particular, these two interpretations yield fundamentally different conclusions with respect to underlying cognitive ability: does a small ratio effect indicate high numerical precision, or the absence of the influence of ratio altogether? This is of particular relevance when considering ratio effects as correlates of other measures of numerical ability such as math scores. If the majority of individuals do not show meaningful ratio effects, then, strictly speaking, can we really distinguish these individuals' ratio "effects" from random noise? In that case, one might be better off using a simpler measure of performance such as overall accuracy or response-time (as has been suggested by Inglis & Gilmore, 2014).

To address this issue, we computed not only each individual's ratio effect (slope of the relation between ratio and response times), but also the variability for each individual of that slope. For every child, we can then assess whether they show a statistically meaningful ratio effect (specifically, the corresponding effect-size, *d*). We examined this in a large cross-sectional sample of over 1,700 Dutch children (Grades K–6, $n > 200$ in each grade). This in turn allowed us to estimate just how prevalent the influence of ratio is on numerical comparison performance. In other words, are individuals who fail to show a reliable ratio effect the exception (whose ratio effects can and should be treated as noisy deviations from a central tendency), or is it the other way around? If the majority of individuals show little or no evidence that ratio influences their comparison performance, then perhaps the correct interpretation of a "small" ratio effect is that ratio is of little relevance for understanding or predicting performance in that individual. Moreover, this would give pause when making the assumption that ratio effects index something meaningful about representation. In this way, we essentially use information about the influence of ratio on performance at the individual level to examine the validity of inferences about ratio effects at the population level.

By measuring the variability and statistical effect-size of ratio effects, we can also examine how each of these measures varies over development, which in turn allows us to examine more carefully previous statements about greater representational precision in older versus younger children (e.g., Halberda & Feigenson, 2008; Holloway & Ansari, 2009; Sekuler & Mierkiewicz, 1977). If decreasing ratio effects are accompanied by decreasing variability in performance in general, one might in fact see ratio effect-sizes stay the same or even *increase* over development—indicating numerical ratio actually has *greater* influence over comparison performance in older children. This in turn might lead us to question the logic behind directly linking the magnitude of individuals' ratio effects and developmental change in their underlying representational precision.

Similarly, we can more carefully examine the view that ratio effects index similar representations or processes across symbolic and nonsymbolic formats. Under the assumptions (a) that ratio effects are the central underlying effect (i.e., failure to show a ratio effect is the exception), and (b) that symbolic and nonsymbolic ratio effects arise from the same source, then an individual showing a statistically meaningful effect in one format should be more likely to do so in the other format. Or, if we assume that little can be learned from those whom fail to show a meaningful effect and restrict our analyses to just those individuals who do show a statistically meaningful effect, then the magnitudes of at least these individuals' ratio effects should be correlated across formats. On the other hand, given the large sample size here, failure to show either of the above results would cast strong doubt on the notion that symbolic and nonsymbolic ratio effects should be thought of as indexing related underlying processes. This latter result would be consistent with a growing body of behavioral and neural evidence that has called into question the link between symbolic and nonsymbolic number representations (Bulthé, De Smedt, & Op de Beeck, 2014; Damarla & Just, 2013; Lyons et al., 2012; Lyons, Ansari, & Beilock, 2015).

With respect to the relation between ratio effects and other forms of numerical processing—such as math scores, recent studies cast doubt on whether representational precision is the correct explanation for this relation. Inglis and Gilmore (2014) showed that simple accuracy across all trials is a more robust correlate of math scores than derivations based on ratio (for a review of these issues, see Dietrich et al., 2015). In addition, Park et al. (2014) recently showed that training on nonsymbolic arithmetic but not nonsymbolic comparison led to improvements in symbolic math scores. Both studies indicate that factors beyond basic representational precision drive the link between basic numerical tasks and more complex math ability. The Inglis and Gilmore (2014) and Park et al. (2014) results suggest that the link between math and basic number processing may not be driven by representational precision per se, but perhaps some other more general mechanism. Here we can extend the Inglis and Gilmore (2014) results by examining whether, for example, ratio effect-sizes or simple overall mean performance better predict standardized arithmetic scores.

Finally it is worth noting that the analytical approach adopted here is not without precedent: work on other well-known numerical processing effects has shown the importance of taking into account how much an individual shows variation in the effect in question. For instance, Nuerk et al. (2004) examined developmental trends in the unit-decade compatibility effect (when comparing two two-digit symbolic numbers, individuals tend to be slower and more error-prone when the ones and tens digits give incongruent as opposed to congruent information). After taking into account each individual's variability in as well as the average of their compatibility effect, Nuerk et al. (2004) showed that compatibility effects actually *increased* over development. In another example, space-number-association-of-response-code (SNARC) effects are usually computed using raw slope estimates for each participant (b_1). Tzelgov et al. (2013) instead recommend computing SNARC effects using correlation coefficients, which take into account the variability of and individual's SNARC slope as well as the slope itself (b_1), and are directly related to effect-sizes expressed as Cohen's *d*.

In the current study, we examine the variability of each individual's ratio effects in symbolic and nonsymbolic comparison tasks in a large, cross-sectional dataset spanning Kindergarten to Grade 6. By doing so we provide the first examination of the true *effects* of numerical ratio on symbolic and nonsymbolic number comparison performance and how this changes over developmental time. Specifically, we assess how this variability may alter and contribute to current thinking about the meaning of ratio effects with respect to the development of number processing and number representation, as well as the similarities and differences between symbolic and nonsymbolic number processing. We do this in four ways. First, we assess whether within-subject variability of ratio effects decreases over development and what this means for understanding how the overall influence of numerical ratio on performance changes as a function of chronological age. Second, we assess the statistical reliability of ratio effects on a child-by-child basis. Does a small ratio effect indicate high numerical precision, or the absence of the influence of ratio altogether? Computing the statistical reliability of a given person's ratio effect can help untangle these two possibilities. Third, we examine whether symbolic and nonsymbolic number comparisons are linked in terms of the properties that govern the comparison process (e.g., representational precision): Is the statistical reliability of a given child's nonsymbolic ratio effect predictive of the same for their symbolic ratio effect. Finally, we assess whether a clearer understanding of the within-subject variability of ratio effects can help explain why ratio effects are inconsistently associated with other measures of numerical and mathematical processing.

Method

Participants

The data collection protocol was approved by the ethics review board at Maastricht University. Data were collected cross-sectionally from 1,739 Dutch children in Kindergarten through Grade 6. Data in this study are a subset of the overall dataset; specifically, here we focus on the Numeral comparison and Dot comparison tasks. Chance performance is difficult to interpret, so we removed children who performed at chance on either of these task (>49% error-rate; results were consistent if this criterion was removed or made more stringent, e.g., >40%). The overall final sample size was $N = 1{,}719$ (see Table 1 for a summary of sample statistics).

Arithmetic achievement was only collected for Grades 1–6, so analyses concerning the relation between ratio effects and arithmetic omit Kindergarteners. Of the 1,503 children in Grades 1–6 from above, we were unable to collect data on the arithmetic or one or more of the control tasks from 13 children (4, 2, 1, 1, 4, 1 children in Grades 1–6, respectively), so these analyses proceeded with an *N* of 1,490.

It is important to note that the data reported here are part of a larger dataset, a portion of which has been previously described in Lyons et al. (2014). Crucially, both the theoretical questions and data analyses described here are completely separate from those in

Table 1
Participant Information for Each Grade

Grade	Removed	Final *n*	Female	Mean age
K	10	216	100	6.02
1	4	231	104	7.06
2	1	227	116	8.12
3	2	265	138	9.16
4	1	279	150	10.32
5	2	255	136	11.10
6	0	246	133	12.19
Total	20	1719	877	9.26

Note. Age (years) and gender statistics are computed based on the final sample.

Lyons et al. The reader may also notice that sample *n*s may not match exactly those in Lyons et al. This is because the previous study relied upon a larger set of tasks; requiring above chance performance on the wider set of tasks resulted in omission of a few more participants in each grade in that article. In addition, Lyons et al. did not assess Kindergarten children.

Procedure

Children were from seven different primary schools in the Netherlands. Parents denied consent by returning an enclosed form. Trained project workers administered all measures to children separately in a quiet room at school. Data were collected in a single session for Grades K–2, 5–6, and in two sessions separated by no more than 5 days for Grades 3–4.

The Ravens, and arithmetic achievement tasks were paper-and-pencil tests. All other measures were computerized. In all tasks, children were told to respond as quickly and accurately as possible. No feedback was given during the main experimental trials. Several practice trials were given for each of the numerical comparison tasks. Reading, Ravens, and Arithmetic were each scored as the total number of correctly completed items.

Numerical Comparison Tasks

Numeral comparison. In the Numeral comparison task, children saw two numbers presented horizontally as Indo-Arabic-numerals, and their task was to decide which number represented the larger quantity. Children saw 64 trials. Thirty-two trials were single-digit; 32 trials were double-digit. Ratios were equated across single-and double-digit items. Quantities ranged from 1 to 45. Ratios (R = min/max) ranged from .25 to .80 (the exact quantities for all trials can be found in the Appendix). Note that only single-digit items were shown to Kindergarteners. Stimuli remained on the screen until the child responded. For response-times, only correct trials were considered. To eliminate abnormally short or long response-times biased by processes likely unrelated to numerical processing (e.g., accidental button presses, distraction from the computer screen, etc.), trials with response-times $<$400 ms or $>$10,000 ms were eliminated (these arbitrary cut-off values were yoked to the DotComp task below). This removed a total of 0.36% of all Numeral comparison trials. For response-times, overall reliability was excellent for this task (Cronbach's α = .943, on average; see Table 2); for errors, reliability was substantially lower (Cronbach's α = .616, on average; see Table 2).

Dot comparison. In the Dot comparison task, children saw two arrays of dots—one on either side of the screen—and their task was to decide which array contained more dots. The quantities and ratios used were the same as those in the Numeral comparison task (with the exception that Kindergarteners saw all 64 trials in the Dot comparison task). Stimuli remained on the screen until the child responded. Children were instructed to estimate which array contained more dots without counting. Children likely complied with this instruction, as response-times in fact decreased slightly as set size increased. If children were counting, one would expect an increase in response-times as the number of dots in the arrays increased; but we found the opposite to be the case, with response-times decreasing by an average of 6.5 ms for every dot added to the comparison arrays. In other words, it is very likely that, despite the lack of a time-limit to respond, children were indeed estimating the relative number of dots in each array, instead of relying on a counting strategy. To eliminate abnormally short or long response-times biased by processes likely unrelated to numerical processing (e.g., accidental button presses, distraction from the computer screen, etc.), trials with response-times $<$400 ms or $>$10,000 ms were eliminated. These arbitrary cut-offs were adopted to remove ~0.25% of trials in each direction, and resulted in removal a total of 0.49% of all DotComp trials. For response-times, overall reliability was excellent for this task (Cronbach's α = .950, on average; see Table 2); for errors, reliability was substantially lower (Cronbach's α = .742, on average; see Table 2).

Table 2
Task Reliability Estimates (Cronbach's α, Based On Average Inter-Item Correlation) for Each Grade, as Well as the Average Across Grades

Grade	Dot response-time	Dot error rate	Numeral response-time	Numeral error rate
K	.937	.790	.857	.778
1	.943	.723	.929	.742
2	.943	.683	.949	.586
3	.943	.708	.950	.580
4	.951	.703	.960	.493
5	.970	.787	.977	.660
6	.959	.804	.980	.475
Average	.950	.742	.943	.616

Because of geometric imperative, within a given trial, all versions of a Dot-comparison task will allow for some nonnumerical parameters (such as area, perimeter, density, etc.) to covary with number. This problem is compounded by the fact that participants switch the parameters they rely upon from trial to trial (Gebuis & Reynvoet, 2012). Paradigms that rely solely on changing the congruency of parameters with number *between* trials may fail to adequately bias participants away from relying on nonnumerical parameters. In the current stimulus set, overall area and average individual dot-size were always incongruent with number (the array with fewer dots had greater overall area and larger average dot-size; individual dot-sizes varied randomly). In other words, the nonnumerical strategy available in our study was the more difficult one because relying on it would force children to focus on nonnumerical variables incongruent with the task goal (determine the numerically greater array). Therefore, our paradigm relied on the assumption that children would be more likely to rely on the relevant parameter (numerosity) that was, by definition, congruent with the task goal than on a parameter incongruent with the task goal (smaller overall area and/or individual dot-size). We saw this assumption as imperfect but less problematic than the assumption (demonstrated to be highly questionable by Gebuis & Reynvoet, 2012) that participants would not switch between the parameters across trials.

Ratio Effects

Ratio effects were determined by treating ratio $\left(R = \frac{n_{min}}{n_{max}}\right)$ as a continuous predictor of performance. For each child, individual trials were treated as separate observations. For each format, there were 64 observations for most children—with one or two fewer in the rare instances where a trial was rejected as an outlier.

Our primary focus was on response-times. Note that this is in contrast to the recent upsurge in popularity of "Weber fractions" (w), which focus exclusively on error-rates. We chose to focus on response-times for four reasons. First—for reasons outlined below and in keeping with previous studies (e.g., Holloway & Ansari, 2009)—our comparison paradigms were optimized for response-time analysis. Second, the most influential studies that have observed decreases in ratio or distance effects over development (and that have concluded that representational precision increases over development) have relied upon response-times (Holloway & Ansari, 2009; Sekuler & Mierkiewicz, 1977). Third, without even considering ratio effects, reliability for the tasks themselves (Cronbach's α), while excellent for response-times, was only acceptable to poor in most grades for errors (see Table 2).

Fourth, and perhaps most crucially, one of our central goals was to examine commonalities and differences in ratio effects both across symbolic and nonsymbolic number comparison, as well as across a range of ages. Error-rates for symbolic comparisons tend to be very low, with many individuals making no errors whatsoever (indeed, this was a key reason why we biased both symbolic and nonsymbolic paradigms to favor variability in response times over error-rates). It is of course impossible to estimate ratio effects on the basis of error-rates for such individuals, unless one assumes that ratio has absolutely no impact on performance, which in the current context would be circular. Furthermore, the prevalence of individuals with no errors increases over development, which can fundamentally confound attempts to draw inferences about how the magnitude of ratio effects may differ between younger and older children. Indeed, in the current dataset, the proportion of children showing no errors in either the Dot or Numeral comparison task rose dramatically from Kindergarten to 6th grade: 13.0%, 12.6%, 18.1%, 19.2%, 26.5%, 30.2%, and 33.3%, respectively. For these reasons, we thus focus our Results and Discussion primarily on response-times. For completeness, error-rates are treated in Supplementary Material; however, even there, we urge the reader to keep in mind the concerns raised here.

Our outcomes of interest were, for each child, the estimated regression coefficient associated with ratio (b_{RE}), the standard error (SE) of this estimate (s_{RE}), and the effect-size derived from the combination thereof (d_{RE}). The unstandardized coefficient can be thought of as the expected increase in response-time (in milliseconds for response-time, in percent-wrong for errors) for that subject from a trial with ratio of 0 and a trial with ratio of 1. Such ratios are somewhat nonsensical, so one might instead think of half the estimated coefficient to reflect the expected difference in response-times on a trial with ratio of .25 versus .75. For example, a participant with an estimated coefficient of 500 ms would be expected to take 250 ms longer on a trial comparing 15 to 20 (ratio = .75), relative to a trial comparing 5 to 20 (ratio = .25).

Ratio effects (b_{RE}). Most studies assess only this statistic for individual participants. It is typically calculated by subtracting performance on the easier condition from the harder condition, or by computing the slope of the relation between ratio and performance (as we have done here). The magnitude of this value is then taken as a measure of a given participant's ratio effect. Here, we calculated b_{RE} as the unstandardized coefficient (see above) of the relation between ratio and performance.

Variability of ratio effects (s_{RE}). b_{RE} is a point estimate of the relation between ratio and performance. Any such point estimate will be associated with a degree of error—which in this case is the variability of the estimated coefficient. For each subject, we also computed the (estimated) SE of the coefficient, which we write here as s_{RE}.

Effect-size of ratio effects (d_{RE}). In most statistical tests, one combines the estimate of the mean and variability of one's measurements of the mean to estimate the "true" size of the difference between those means. This estimate (often a t-statistic) is then commonly expressed as an effect-size, to give a relatively context-free metric of the true magnitude of the effect. Here, we computed Cohen's effect-size d as a function of b_{RE}, s_{RE}, and the number of valid trials n for that subject: $d_{RE} = \frac{2t}{\sqrt{n-2}}$, where t is the t-statistic associated with b_{RE} $\left(\frac{b_{RE}}{S_{RE}}\right)$. With b_{RE}, s_{RE}, and d_{RE}, we were able to compute, for each subject individually, their average ratio effect, variability of the ratio effect, and the ratio effect-*size*, respectively. Across Grades K–6, then, we were able to assess cross-sectional developmental trends in these measures separately.

Additional Measures

For the purpose of computing the relation between ratio effects and math ability, we report results from a mental arithmetic task and several other control measures described below.

Mental arithmetic (Arithmetic). The Arithmetic task was the standardized TempoTest Automatiseren (TTA) of basic arithmetic ability (De Vos, 2010). Children were administered two worksheets—one containing 50 addition and the other containing 50 subtraction problems. Children were instructed to calculate as many operations as possible within 2 min per worksheet. Scores were the total number of correctly answered problems across both worksheets. Reliability for this task is high (.92; Janssen et al., 2010).

Nonverbal intelligence (Ravens). The Ravens control measure comprised a battery of progressive matrices. This is a normed, untimed, visuospatial reasoning test for children (Raven et al., 1995). Children saw a colored pattern and were to select the missing piece from six choices. There were 36 test items. A child's score was the total number of correctly solved items. Van Bon (1986) reported reliabilities of .80 or higher for the Dutch version of this task.

Reading ability. The Reading measure was part of the normed Maastricht Dyslexia Differential Diagnosis battery (Blomert & Vaessen, 2009). It comprised three subtasks: high-frequency words, low-frequency words, or pseudowords. In each subtask, participants saw a series of up to five screens (advanced by the experimenter), each with 15 items (75 total items per task). Children were to read each item aloud as quickly and accurately as possible. An experimenter manually marked accuracy for each item. Scores were the total number of correctly read items in 30 s, summed across the three subtasks. Reported test–retest reliability for this task is high (.95; Blomert & Vaessen, 2009).

Basic stimulus-response processing. In this task, children saw four horizontally arranged boxes on the screen. On each of 20 trials, a fish appeared in one of the four boxes. Children's task was to press the corresponding key on the response box as quickly and accurately as possible. Reliability on this task was high: $\alpha = .944$.

Results

Developmental Trends in d_{RE}, s_{RE}, and d_{RE}

Developmental trends were assessed for each of the three measures (b_{RE}, s_{RE}, and d_{RE}) by entering each into a 2(Format: Dot, Numeral) $\times$ 7(Grade: K–6) analysis of variance (ANOVA). As noted in Method section above, Results focus on response-times (see Supplementary Material for error-rate results).

b_{RE} (average ratio effect). Results for b_{RE} (*average* ratio effect) are summarized in Figure 1a. The main effects of Format ($F(1, 1712) = 845.06$, $p < .001$, $d = 1.41$, $\eta_p^2 = .33$; Dots: $M = 1164.16$, $SE = 24.28$, Numerals: $M = 378.95$, $SE = 14.31$); and Grade ($F(6, 1712) = 35.69$, $p < .001$, $d = 0.71$, $\eta_p^2 = .11$) were highly significant. There was also a small (in terms of effect-size) interaction ($F(6, 1712) = 3.38$, $p = .003$, $d = 0.22$, $\eta_p^2 = .01$). Considering each Format separately, the strongest trend (in terms of polynomial contrast effect) was negative (indicating a decrease) and linear in both cases (Dots: $t_{1712} = -10.31$, $p < .001$, $d = -0.50$; Numerals: $t_{1712} = -10.32$, $p < .001$, $d = -0.50$). Average ratio effects were overall much larger for Dot than Numeral comparison; average ratio effects decreased as Grade increased; this decrease was similar for both formats, albeit slightly larger for Dots.

From the values in Figure 1a, it should also be clear that average ratio effects were highly significant in all grades and both formats (indeed, all $ps < 1E\text{-}10$). From this—as have many previous authors—one might conclude that ratio plays a dominant role in shaping numerical comparison behavior regardless of format, and that ratio effects decrease over development. Note, however, that this interpretation relies on computation of only average ratio effects in each child, and computes variability thereof *only at the group level*. In the next sections, we instead compute the variability of ratio effects at the individual subject level—that is, at the same level at which average ratio effects are computed (both traditionally in previous studies as well as in the current study). We then test whether the traditional inferences about the influence and developmental trajectory of ratio effects still obtain.

s_{RE} (variability of ratio effects). Results for s_{RE} are summarized in Figure 1b. The effect of Format was significant but modest in size ($F(1, 1712) = 14.20$, $p < .001$, $d = 0.18$, $\eta_p^2 = .01$; Dots: $M = 434.68$, $SE = 7.32$, Numerals: $M = 393.93$, $SE = 9.68$). The effect of Grade ($F(6, 1712) = 310.11$, $p < .001$, $d = 2.09$, $\eta_p^2 = .52$) and the interaction term ($F(6, 1712) = 75.90$, $p < .001$, $d = 1.03$, $\eta_p^2 = .21$) were highly significant. Considering each Format separately, the strongest trend (in terms of polynomial contrast effect) was negative (indicating a decrease) and linear in both cases (Dots: $t_{1712} = -21.32$, $p < .001$, $d = -1.03$; Numerals: $t_{1712} = -42.61$, $p < .001$, $d = -2.06$). Variability in ratio effects was overall slightly higher for Dots; variability dramatically reduced for both formats in older children, and this decrease was substantially more rapid for Numerals. Specifically, variability was higher for Numerals in Kindergarteners ($p < .001$, $d = -1.35$), the two formats did not differ in Grade 1 ($p = .364$, $d = -0.12$) and Grade 2 ($p = .173$, $d = 0.18$), and was higher for Dots thereafter ($ps < .001$, $ds > 1$).

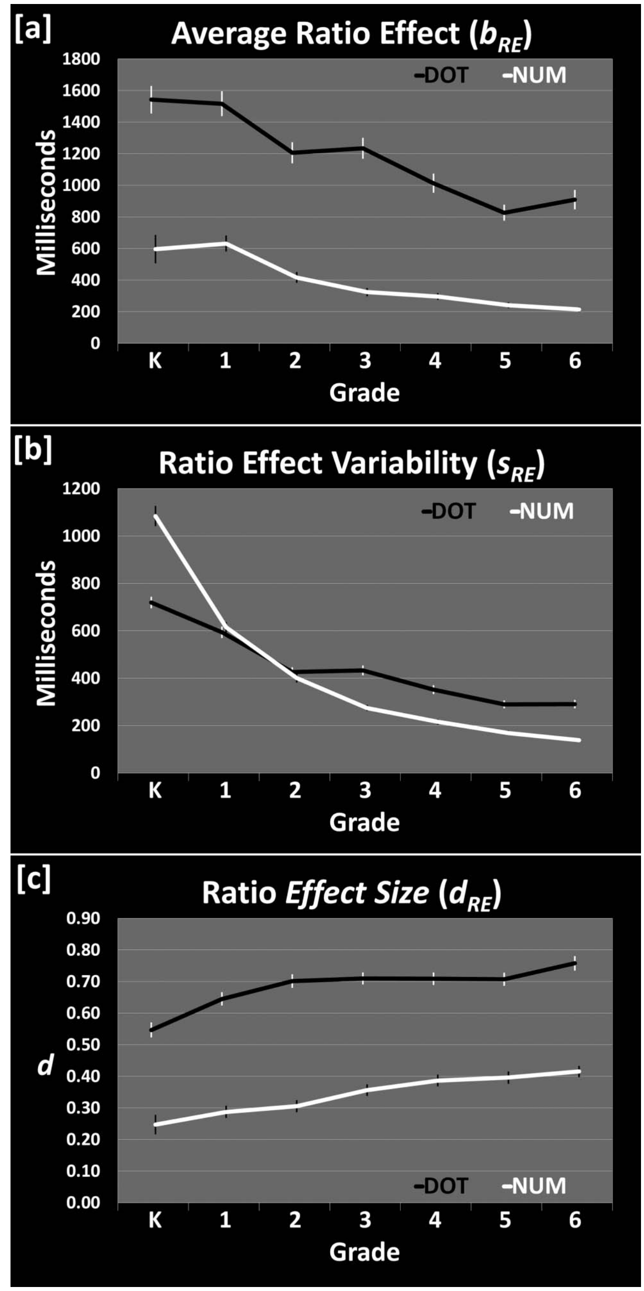

Figure 1. Figure 1 shows developmental trends in each of the measures of ratio effects: average ratio effect (b_{RE}), variability of the ratio effect (s_{RE}), and the effect-size of this effect (d_{RE}). Each measure was first computed for each child separately. Values were then averaged across all children in that grade. Error-bars: *SE*s of the mean.

d_{RE} (ratio effect-size). Results for d_{RE} are summarized in Figure 1c. The main effects of Format (Format: $F(1, 1712) = 1262.78$, $p < .001$, $d = 1.72$, $\eta_p^2 = .42$; Dots: $M = 0.686$, $SE = .007$, Numerals: $M = .346$, $SE = .007$) and Grade ($F(6, 1712) =$

20.86, $p < .001$, $d = 0.54$, $\eta_p^2 = .07$) were highly significant; the interaction term was not ($F(6, 1712) = 1.58$, $p = .150$, $d = 0.15$, $\eta_p^2 = .01$). Considering each Format separately, the strongest trend (in terms of polynomial contrast effect) was positive (indicating an increase) and linear in both cases (Dots: $t_{1712} = 7.55$, $p < .001$, $d = 0.37$; Numerals: $t_{1712} = 8.16$, $p < .001$, $d = 0.39$). Ratio effect-sizes were overall larger for Dots than Numerals, *increased* over development, and this increase was similar for both formats.

Summary. To summarize, average ratio effects (b_{RE}), and the variability of these ratio effects (s_{RE}) decreased over development. However, ratio effect-sizes (d_{RE}) increased. That is, the reliability of the influence of numerical ratio over performance (d_{RE}) was greater in older than in younger children, and this developmental difference was similar across formats. In other words, the greater developmental change with respect to ratio effects was in terms of their variability (i.e., the change in $s_{RE} > b_{RE}$). The parallel increases in d_{RE} for Dots and Numerals were underlain by different developmental trajectories for average and variability of ratio effects. While average ratio effects (b_{RE}) decreased more rapidly for Dots, variability thereof (s_{RE}) decreased significantly[2] more rapidly for Numerals. Finally, as will be taken up further in the next section, ratio effect-sizes were substantially higher for non-symbolic relative to symbolic number comparison.

Statistically Reliable Ratio Effects Within Individuals (Response-Times)

One reason for the popularity of ratio effects (as well as for distance effects and Weber-fractions) is that, at the population level, they are a very good predictor of performance (especially response-times, as in the present case). Indeed, if we take the average response-time on each trial across all 1,719 children, and then plot these averages against ratio, the correlations are reasonably high: Dots: $r_{1717} = .723$, Numerals: $r_{1717} = .544$. Note that this approach implicitly takes the view that there is some underlying principle or tendency that governs human behavior (ratio drives comparison performance), and individual variation around this central governing tendency is essentially noise. Recently, however, there has been a major upsurge in interest in this individual variation—treating it not as noise, but as a source of meaningful variation in itself. Researchers adopting such an "individual differences" approach have made several major discoveries in the field of numerical cognition of late, with special emphasis put on the relation between individual differences in ratio effects (or related measures such as distance, etc.) and individual differences in more complex numerical skills, such as mental arithmetic (e.g., Bonny & Lourenco, 2013; Bugden et al., 2012; De Smedt et al., 2009; Feigenson et al., 2013; Fuhs & McNeil, 2013; Halberda et al., 2008, 2012; Holloway & Ansari, 2009; Libertus et al., 2011; Lonnemann et al., 2011; Lyons & Beilock, 2011; Mazzocco et al., 2011; Piazza et al., 2010; Sasanguie et al., 2012).

On the other hand, when one is concerned with individual differences or variation around an assumed underlying principle, it is crucial to obtain evidence that this principle is indeed influencing a given individual's performance to begin with. More concretely, many researchers have interpreted small average ratio effects (b_{RE}) as evidence that a given individual is capable of representing numbers quite precisely. Another interpretation, however, is that ratio has no reliable influence over that person's performance to begin with—that is, ratio is unpredictive of the person's response-times. Notice that from an individual differences perspective, these two interpretations yield fundamentally different inferences with respect to underlying cognitive ability: does a small ratio effect indicate high numerical precision, or the absence of the influence of ratio altogether? Taking into account variability of a given person's ratio effect can help untangle these two possibilities. If a person has a small average ratio effect (b_{RE}) *and* shows low variability (s_{RE}) in the effect, one may conclude that ratio has a reliable influence over their performance, and hence the small ratio effect may indeed indicate greater numerical precision on their part. Conversely, if the same person shows relatively high variability with respect to ratio (s_{RE}), this would suggest that ratio has little influence over their performance, which in turn should limit one's inferences with respect to the meaningfulness of their small ratio effect (b_{RE}). In the current dataset, we have computed individuals' ratio *effect-sizes* (d_{RE}) which is largely driven by the ratio between b_{RE} and s_{RE}. In doing so, we saw in the previous section (see Figure 1) that, even though average ratio effects decreased with development, concomitant decreases in the variability of these effects yielded *increasing* ratio effect-sizes—indicating that, on average, ratio exerts more influence over performance over development, not less.

With respect to individual differences, for each format (Dots and Numerals) we can then ask what proportion of individuals showed a statistically reliable influence of ratio on their number comparison performance. If this proportion is high, then one might well draw meaningful conclusions about number representation from the magnitude of an individual's ratio effect. On the other hand, if the proportion is low, this would indicate that many individuals' small ratio effects are statistically indistinguishable from random noise—that should curb inferences regarding number representation accordingly.

To test this, we adopted a cut-off effect-size (d_{RE}) of 0.5. First, assuming 64 observations (i.e., the number of trials each participant completed in each format), a d of 0.5 corresponds to $t_{62} = 1.97$, $p = .053$. Second, general convention regards an effect-size of 0.5 as the lower-bound of a "medium" effect (Cohen, 1992). Third, if one presumes the existence of a true effect-size of 0.5, then one's power to detect this effect (at $p < .05$) with the current number of observations per child (64 trials) is high: .97. At each Grade and for each Format, we then computed the proportion of children showing an effect-size (d_{RE}) greater than or equal to 0.5. Results are plotted in Figure 2.

Relatively few Kindergarteners showed statistically reliable ratio effects: 57% for Dots and 26% for Numerals, indicating that ratio effects in many children at this age may often be indistinguishable from random noise. For Dots, the proportion of children showing statistically reliable ratio effects quickly climbed to over three-quarters of children in 2nd grade and to over four-fifths by 3rd grade, indicating that ratio effects may well capture meaningful numerical information in the majority of children. For Numerals on the other hand, the proportion of statistically reliable ratio effects peaked at just over a third (38%) in 6th graders. For

[2] The three-way interaction (Format × Grade × Measure: b_{RE}, s_{RE}) was highly significant: $F(6, 1712) = 19.29$, $p < .001$, $d = .52$, $\eta_p^2 = .21$.

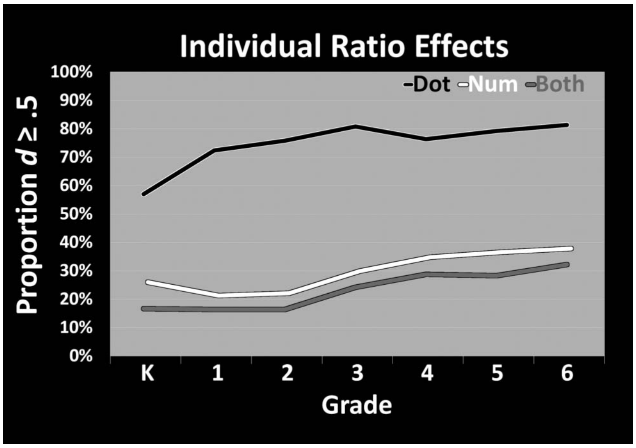

Figure 2. Figure 2 shows the proportion of children showing a statistically "meaningful" ratio effect-size ($d_{RE} \geq .5$) for symbolic (white line) and nonsymbolic (black line) comparisons. The gray line indicates the proportion of children showing a ratio effect-size ≥.5 for both formats.

Numerals, then, one might expect that ratio effects in the majority of even older children reflect little more than noise—indicating that ratio has only a small and relatively unreliable influence over Numeral comparison performance.

Ratio Effects Across Formats

There is considerable theoretical investment in the idea that symbolic and nonsymbolic numbers are closely linked early in development (Dehaene, 1997, 2008; Feigenson et al., 2004, 2013; Gallistel & Gelman, 2000; Piazza, 2010). Indeed, the fact that ratio effects are observed for both symbolic (Numerals) and nonsymbolic (Dots) number comparisons has been taken to indicate common underlying representation or basic processing across formats (e.g., Dehaene, 2008).

This view predicts that ratio effects across formats should be related. At first glance, Figure 2 would appear to confirm this prediction. In Figure 2, the gray line represents the proportion of children whom showed a statistically reliable ratio effect for both Dots and Numerals. The gray line does appear to closely follow the white line for Numerals. This might indicate that those children who show a statistically reliable ratio effect for one format do so also for the other format, which in turn would support the notion that the two formats are closely linked. A skeptic might suggest that the gray line is merely a reflection of the independent probabilities for Dots and Numerals (i.e., the product of the black and white lines). As it turns out, the gray line is very well predicted by simply multiplying (at each Grade) the black and white lines (R^2 = .973), indicating that the black and white lines are statistically independent of one another. Accordingly, if a child showed a statistically significant ratio effect in one format, in no Grade was this predictive of whether that child showed a statistically significant ratio effect in the other format (maximum R^2 = .011).

An alternative view is that the relative magnitude of ratio effects is valid only for those children who showed a statistically reliable ratio effect, and so correlations should be restricted to just those individuals. To test this, we restricted analysis to just those children with $d_{RE} > .5$ for both Dots and Numerals (n = 407).[3] We correlated each ratio effect measure across formats (i.e., Dot b_{RE} with Numeral b_{RE}), controlling for chronological age and basic processing speed (Ravens and Reading scores were not available for many Kindergarteners). Only s_{RE} showed a significant cross-format correlation (b_{RE}: r_{403} = .056, p = .264; s_{RE}: r_{403} = .155, p = .002; d_{RE}: r_{403} = .022, p = .659). (Note also that results were highly similar even if one assessed the entire sample.) In other words, the relative magnitudes (b_{RE}, d_{RE}) of children's ratio effects were unrelated across formats. Variability (s_{RE}) was weakly correlated across formats; however, this measure is closely related to overall mean performance (see the next section), which complicates one's interpretation of this result.

In summary, our results consistently failed to show a relation across formats in terms of ratio effects, indicating that what relation there may be between symbolic and nonsymbolic processing, it is unrelated to representational precision, or whatever ratio effects may be indexing in each respective format.

The Relation Between Ratio Effects and Arithmetic Achievement

There has been considerable recent interest in the relation between measures of basic numerical processing abilities (such as ratio effects, distance effects, and Weber-fractions) and math abilities (e.g., Bonny & Lourenco, 2013; Bugden et al., 2012; De Smedt et al., 2009; Feigenson et al., 2013; Fuhs & McNeil, 2013; Halberda et al., 2008, 2012; Holloway & Ansari, 2009; Libertus et al., 2011; Lonnemann et al., 2011; Lyons & Beilock, 2011; Mazzocco et al., 2011; Piazza et al., 2010; Sasanguie et al., 2012). Here, we assessed the relation between each of our measures of the ratio effect (b_{RE}, s_{RE}, and d_{RE}) and Arithmetic. Recent work has suggested that in some cases, simply taking average performance may be preferable to deriving ratio-related performance (Inglis & Gilmore, 2014). Indeed, results shown in Figure 2 indicate that, at least for Numerals, ratio effects may be capturing little more than statistical noise for the majority of subjects. Hence, in addition to the three ratio measures, we also assessed the relation between mean performance (M) and Arithmetic. For each measure, we also controlled for Ravens, Reading, and basic stimulus-response processing speed. Results are reported as partial correlations (r_p) and corresponding effect-sizes (d). Note that expected correlations are negative, such that smaller ratio effects (faster response-times for M) relate to better (higher) Arithmetic scores.

Table 3 summarizes the results. Note that nonsignificant correlations are greyed out (after correcting for multiple comparisons[4]). For both Dots and Numerals, M and s_{RE} were significantly related to mental arithmetic performance. For Dots, these relations should be considered against the background that they (a) were only obtained when using the full sample, (b) accounted for only about 1% of the total variance, and (c) were not significant in any of the individual grades. For Numerals, the correlations between M and

[3] Because of the limited sample size (less than a quarter of the original), several grades were left with only a few dozen qualifying individuals; hence, we collapse across grades here. Results were highly similar in each grade, however.

[4] Using the Dunn-Šidak method (Šidak, 1967), the critical threshold for 56 correlations was p = 9.2E-04.

Table 3
Partial Correlations (r_p) Between Overall Performance (in Terms of Response-Time: M) and Arithmetic Performance, as Well as That Between Each Ratio-Effect Measure (b_{RE}, s_{RE}, and d_{RE}) and Arithmetic

	Dots				Numerals			
Grade	M	b_{RE}	s_{RE}	d_{RE}	M	b_{RE}	s_{RE}	d_{RE}
All								
r_p	−0.129	−0.047	−0.089	0.034	−0.402	−0.106	−0.335	0.065
p	6.3E-07	7.0E-02	6.1E-04	1.9E-01	4.3E-59	4.0E-05	2.6E-40	1.2E-02
d	−0.259	−0.094	−0.178	0.069	−0.879	−0.214	−0.710	0.130
1								
r_p	−0.116	−0.049	−0.110	0.070	−0.379	−0.001	−0.303	0.107
p	8.2E-02	4.6E-01	9.8E-02	3.0E-01	4.1E-09	9.9E-01	3.5E-06	1.1E-01
d	−0.233	−0.098	−0.222	0.140	−0.818	−0.002	−0.635	0.215
2								
r_p	−0.136	−0.107	−0.152	−0.003	−0.392	−0.079	−0.348	0.123
p	4.3E-02	1.1E-01	2.3E-02	9.7E-01	1.3E-09	2.4E-01	9.7E-08	6.6E-02
d	−0.274	−0.216	−0.307	−0.005	−0.853	−0.159	−0.742	0.248
3								
r_p	−0.057	−0.027	−0.034	0.078	−0.304	−0.011	−0.237	0.052
p	3.6E-01	6.7E-01	5.8E-01	2.1E-01	5.3E-07	8.6E-01	1.0E-04	4.0E-01
d	−0.114	−0.053	−0.069	0.156	−0.638	−0.023	−0.489	0.105
4								
r_p	−0.059	0.028	−0.052	0.022	−0.350	−0.127	−0.331	0.058
p	3.3E-01	6.4E-01	3.9E-01	7.1E-01	2.5E-09	3.5E-02	1.8E-08	3.4E-01
d	−0.119	0.056	−0.104	0.045	−0.747	−0.256	−0.703	0.117
5								
r_p	−0.148	−0.078	−0.104	0.045	−0.437	−0.252	−0.384	−0.050
p	2.0E-02	2.2E-01	1.0E-01	4.8E-01	5.7E-13	5.9E-05	3.9E-10	4.3E-01
d	−0.299	−0.156	−0.209	0.090	−0.971	−0.521	−0.832	−0.100
6								
r_p	−0.146	0.023	−0.014	0.070	−0.468	−0.280	−0.380	−0.039
p	2.3E-02	7.2E-01	8.3E-01	2.8E-01	1.5E-14	1.0E-05	1.0E-09	5.5E-01
d	−0.296	0.045	−0.027	0.140	−1.058	−0.583	−0.821	−0.077

Note. Correlations were computed after controlling for Ravens and Reading scores, as well as basic processing speed. This was done for each grade separately as well as across the entire sample (all). Greyed-out values are nonsignificant after correcting for multiple (56) correlations (Šidak, 1967), requiring a corrected p-value $< 9.2\text{E-}04$. Other terms: p = p-value, d = effect-size.

Arithmetic and s_{RE} and Arithmetic were substantially larger (corresponding to effect-sizes of about .85 and .70, respectively), and obtained in all six grades independently. Although the overall relation between Numeral b_{RE} and Arithmetic was significant, this appears to have been driven by a single grade (5).

In summary, then, measures of ratio effects appear to have limited predictive capacity with respect to arithmetic performance. Neither b_{RE} nor d_{RE} was reliably correlated with Arithmetic. In fact, consistent with previous work, a simple calculation of mean performance (M) was the best predictor for both Dots and Numerals. This was followed closely by s_{RE}. However, as it turns out, M and s_{RE} are highly correlated with one another to begin with (Dots: $r_{p(1489)} = .826$, Numerals: $r_{p(1489)} = .786$); hence it is unsurprising that both measures correlate similarly with Arithmetic. Finally, consistent with prior work (Inglis & Gilmore, 2014), one may in the end opt simply for mean performance because (a) M was generally slightly more strongly correlated with Arithmetic than s_{RE}, and (b) M is substantially simpler to calculate than s_{RE}—with the upshot that one is effectively forgoing ratio-related measures of performance altogether.

Discussion

The ratio effect has long been considered a hallmark of numerical comparison tasks (e.g., Buckley & Gillman, 1974; Moyer & Landauer, 1967). Ratio effects—the slope of the degree to which ratio predicts performance—have long been interpreted as a measure of numerical representational precision, where a smaller ratio effect is thought to indicate a more precise underlying representation (Dehaene, 1997, 2008; Dehaene & Changeux, 1993; Feigenson et al., 2004; Halberda & Feigenson, 2008; Nieder & Dehaene, 2009; Verguts & Fias, 2004). Numerical comparisons made with symbolic stimuli tend to show smaller ratio effects than those made with nonsymbolic stimuli (e.g., Buckley & Gillman, 1974), and ratio effects are typically smaller in older relative to younger children (e.g., Halberda & Feigenson, 2008; Holloway & Ansari, 2009; Sekuler & Mierkiewicz, 1977). Accordingly, symbolic number representation has been interpreted as being more precise than nonsymbolic number representation, and representations in both formats are thought to become more precise over the course of development (Feigenson et al., 2004, 2013; Holloway & Ansari, 2009; Sekuler & Mierkiewicz, 1977). Similarly, correlations between ratio effects and other types of math processing are often interpreted to mean that more precise numerical representations predict better math ability (for a review, see Feigenson et al., 2013).

Previous work has largely focused on average effects across individuals, which tacitly assumes that individuals who fail to show a reliable effect are the exception—deviating from the more general underlying principle. Here, we test this assumption in a large, cross-sectional dataset (<1,700 children in Grades K–6) by computing not just each participant's average ratio effect (b_{RE}), but also the variability of each participant's ratio effect (s_{RE}), which in turn allowed us to compute the effect-size (and statistical significance) of each participant's ratio effect (d_{RE}). Our results lead us to reconsider several central assumptions regarding the interpretation of ratio effects—in particular, how ratio effects change across development, how and whether ratio effects index representational precision, and what ratio effects indicate across formats (i.e., symbolic and nonsymbolic).

Using these measures, we found that the majority of individuals showed statistically meaningful ratio effects for nonsymbolic comparisons. Indeed, effect-sizes were overall quite robust for nonsymbolic numbers, with d_{RE} about .69 on average (across all grades, and peaking at about .76 by Grade 6), which is well within the "medium" effect-size range (Cohen, 1992), and, in the current sample, corresponds to a significance level of $p = .008$ ($p = .004$ for Grade 6). Further, just over 75% of children showed a statistically significant ratio effect (with this figure reaching over 80% by Grade 6). Ratio effects thus exert a clear influence over performance on nonsymbolic comparison tasks—an effect that can robustly be measured in the large majority of children even at the individual level. However, when it comes to symbolic numbers, our results strongly undermine assumptions regarding the ubiquity of ratio effects for symbolic numbers: Only about a third of individuals showed a statistically meaningful symbolic ratio effect, calling into question how much stock should be put into this effect at the population level. Consistent with work on symbolic distance effects (van Opstal et al., 2008; Verguts & van Opstal, 2005), our results suggest the theoretical assumption that symbolic ratio effects index representational precision should be abandoned. Ratio effect-sizes were overall small for symbolic numbers: $d_{RE} = .34$ on average, which is traditionally a small effect and corresponds to about $p = .18$. Indeed, only 30% of children showed a statistically significant symbolic ratio effect overall—with barely a quarter doing so in Kindergarten (26%). In other words, if a given effect is exerting a barely detectable influence over performance, and this influence is detectable only in about a third of one's sample (i.e., the effect is indistinguishable from random noise in some 70% of one's sample), then the meaning of and weight given to one's interpretation of this effect, we believe, should be drastically curtailed.

In a similar vein, our results indicate that the sources of ratio effects for symbolic and nonsymbolic number comparisons are likely derived from independent sources. One might begin with the assumption that nonsymbolic ratio effects are indeed indicative of underlying nonsymbolic representational precision. A long-held view is that symbolic numbers point to these underlying nonsymbolic numerical representations; they just do so with more precision (Dehaene, 2008, p. 552; see also, Dehaene, 1997; Eger et al., 2009; Feigenson et al., 2004, 2013; Gallistel & Gelman, 2000; Hubbard et al., 2008; Libertus & Brannon, 2009; Lyons & Ansari, 2009; Nieder & Dehaene, 2009; Piazza et al., 2007; Verguts & Fias, 2004). If this were the case, then one would expect ratio effects for symbolic comparisons to be related to their nonsymbolic counterparts. Instead, we see that they are unrelated—a result that holds even when we consider only children who showed significant ratio effects in both formats. This is not only consistent with a growing body of evidence calling into question the link between symbolic and nonsymbolic number processing (Bulthé et al., 2014; Damarla & Just, 2013; Lyons et al., 2012, 2015), it also reaffirms the need to take greater care when interpreting the meaning of symbolic ratio and/or distance effects.

It is interesting that about 3 in 10 children *did* show a significant symbolic ratio effect. Why? We have already seen that, even in these children, their symbolic and nonsymbolic ratio effects are unrelated to one another, so the standard assumption that symbolic ratio effects indicate a connection with nonsymbolic ratio effects does not seem valid. We did see a weak correlation across formats for s_{RE}—a measure that is closely linked to overall mean performance (M). This suggests that symbolic ratio effects—at least those that can be detected—may derive from a more general cognitive source. Perhaps one might instead turn to the suggestion made by Verguts and van Opstal (2005) and van Opstal et al. (2008) regarding symbolic distance effects: They index little regarding representation precision, but instead are driven by factors such as relative word-frequency and response-selection. Symbolic ratio effects, by extension, might index these or still other cognitive factors (e.g., effects of unit notation; Moeller et al., 2012). One might also examine increasingly more difficult ratios—that is, those that would be impossible in a nonsymbolic context, such as 1,000,000 versus 1,000,001. In summary, there are many possible explanations; however, we must emphasize that explanations relying on a link between symbolic and nonsymbolic ratio effects and accompanying notions of representational precision are not supported by the current dataset.

The notion of representational precision also figures prominently in interpretations of ratio effects in the context of the development of numerical processing. The traditional view is that the ratio effect is an indicator of representational precision, and a decreasing ratio effect over development reflects an increase in said precision (Feigenson et al., 2004, 2013; Halberda & Feigenson, 2008; Holloway & Ansari, 2009; Sekuler & Mierkiewicz, 1977). What we show here is that *the influence of ratio on comparison performance* (in this case RTs[5]) in fact *increases* over development. That is, effect-sizes were in fact larger in older children (Figure 1c). On an arithmetic level, this simply means that the variability of ratio effects (s_{RE}) fell off more rapidly across ages than did average ratio effects (b_{RE}). More broadly, the increase in ratio *effect-sizes* (d_{RE}) over development puts one into an awkward conceptual corner. At face value, then, numerical representations are increasingly precise and this precision is increasingly influential over (one's ability to measure) comparison performance. On the other hand, it is unclear why numerical representations would contribute to numerical comparison performance to differing degrees at different points in (school-age) development. An alternative explanation is that we are simply dealing with a measurement issue—younger children's performance is more variable and so one's ability to detect subtle effects

[5] Effect-sizes for errors decreased across grades (see Supplementary Materials). On the other hand, both average ratio effects (b_{RE}) and ratio-effect variability (s_{RE}) actually *increased* across grades. This latter result is in contrast to previous studies (at least with respect to average ratio effects; e.g., Halberda & Feigenson, 2008). As such, we believe this is yet another reason (see Method for several other reasons) why inferences from the error-rate data *in the current study* are very likely invalid.

is substantially hampered. However, that is precisely the issue: what may be a statistically robust effect later in development may qualify as a "subtle" effect earlier in development. In particular, this view lends skepticism to statements regarding the magnitude of ratio effects early in development—especially if these statements do not take into account the underlying variability of the effect in question. Indeed, barely more than half (57%) of Kindergarteners showed a statistically robust ratio effect. The upshot is that the ratio effects of the other 43% are, strictly speaking, not reliably distinguishable from random noise. If nearly half of one's sample is failing to show a reliable effect, perhaps it is unwise to draw too strong a conclusion about the meaning of that effect—be it representational precision or some other interpretation such as input frequency, response selection, and so forth (e.g., Verguts & van Opstal, 2005; van Opstal et al., 2008). In summary, we show that the influence of ratio on performance during both symbolic and nonsymbolic number comparison is in fact greater in older children, which calls into question the assumption that developmental change in ratio effects is indicative of underlying change in representational precision.

The fact that many individuals do not show a significant or reliable ratio effect (especially in the case of symbols) is particularly salient in light of the recent popularity of relating individual differences in performance on basic numerical tasks with scores on more complex math processing (e.g., Bonny & Lourenco, 2013; Bugden et al., 2012; De Smedt et al., 2009; Feigenson et al., 2013; Fuhs & McNeil, 2013; Halberda et al., 2008, 2012; Holloway & Ansari, 2009; Libertus et al., 2011; Lonnemann et al., 2011; Lyons & Beilock, 2011; Mazzocco et al., 2011; Piazza et al., 2010; Sasanguie et al., 2012). When one is concerned with individual differences or variation around an assumed underlying principle, it is crucial to obtain evidence that this principle is indeed influencing a given individual's performance to begin with. With respect to ratio effects, the common assumption is that this principle is representational precision, and therefore, many have interpreted a small average ratio effect (b_{RE}) as evidence that a given individual is capable of representing numbers quite precisely. On the other hand, if a small mean effect is not accompanied by proportionally small variability in the effect, then, statistically speaking, one may question the presence of an effect to begin with. That is, the underlying principle—ratio drives performance as a function of representational precision—may not be applicable to the individual in question. In fact, we found that individual differences in ratio effects were unrelated to individual differences in mental arithmetic ability. By contrast, consistent with previous work (Inglis & Gilmore, 2014), overall average response times (M) *was* correlated with performance, suggesting that, when it comes to predicting individual variability in Arithmetic, one might well forgo ratio-related measures of performance altogether. Also consistent with previous work, this relation was especially robust for symbolic number comparisons (e.g., Lyons et al., 2014; for a review and meta-analysis, respectively, see De Smedt et al., 2013; Fazio et al., 2014). It is worth highlighting this latter result because it indicates that the task that showed the least amount of statistical validity in terms of ratio effects (symbolic comparisons) was nevertheless the strongest predictor of arithmetic performance. Crucially, however, this was the case only once one abandoned the notion of ratio effects altogether in favor of simple mean response times (M). This raises the striking possibility that the relation between basic numerical competencies and more complex math processing in fact has little to do with representational precision.

Along these lines, previous researchers have already begun to call into question the validity and meaning of symbolic distance effects (and thus, by extension, ratio effects as well). Maloney et al. (2010) showed that numerical distance effects for symbolic number comparisons were unreliable even for adult subjects within the same testing session (distance effects were uncorrelated across blocks). Our results help contextualize Maloney et al.'s: symbolic ratio/distance effects may be unreliable because, in the majority of individuals, these effects are indistinguishable from random noise. As noted previously, Verguts and van Opstal (2005; see also van Opstal et al., 2008) have proposed an alternative explanation (based on both empirical evidence and a computational model) of the symbolic distance effect. The model eschews representational precision, and instead generates the classic distance effect because response selection mechanisms are sensitive to the relative frequency with which it encounters different numbers (frequencies that were calculated to match those found in human lexical corpi). Perhaps, then, rather than representational precision, it is individuals' familiarity and fluency with which they can manipulate (especially symbolic) numbers that serves as the crucial link between basic numerical skills (such as number comparison) and more complex math abilities (such as mental arithmetic). Furthermore, both we and several previous studies (e.g., De Smedt et al., 2013; Fazio et al., 2014; Lyons et al., 2014) have shown that symbolic comparison performance is a stronger and more robust predictor of arithmetic processing. Taken together, both our data and the evidence reviewed above support the notion that sophisticated math skills have little if anything to do with the precision with which one can represent nonsymbolic numbers. To be clear, we are not claiming that nonsymbolic number processing is entirely irrelevant for other kinds of math processing. Rather, our statement pertains to the notion of "representational precision" per se (for consistent recent evidence, see, e.g., Park & Brannon, 2014).

Turning to methodological considerations, it is important to note that, with 64 trials,[6] our power to obtain statistical significance for the average effect-size seen for symbolic numbers (.35) was reasonably high: .78. Hence, one might ask: why not simply establish a cut-off at .35, instead of the stricter .5—would not a simple shift of the arbitrary cut-off change the proportion of children showing a significant effect? Similarly, could one not simply increase the number of trials to increase power and so make a statistically significant result more likely? On both counts, the answer is of course yes, it would. However, we believe to do so would be to miss the broader point. First, because .35 was the *average* effect-size, one would still expect that roughly half of children would fail to pass even this more liberal threshold. Second, while increasing

[6] Only 32 trials of the symbolic task were completed by Kindergarteners. Note, however, that Figures 1c and 2c show that results for Kindergarteners fit well within the overall trend in terms of effect-size (all trends across Grades remained consistent even when only Grades 1–6 were considered). Indeed, if anything, a slightly higher proportion of Kindergarteners (26%) than 1st or 2nd graders (21% and 22%) passed the $d = .5$ threshold for the symbolic task. Kindergarteners completed 64 trials in the nonsymbolic comparison task.

the number of trials will decrease *p* values, one's estimate of the underlying effect-size should remain relatively unchanged, or even decrease (Button et al., 2013; to say nothing of the methodological difficulty that imposing a large number of trials on young children would present). Hence the real question: what is a meaningful effect-size? One could theoretically drop the effect-size threshold asymptotically toward zero indefinitely, so there is no purely mathematical or universal answer to this question. On the other hand, if one's goal is to identify and measure an indicator of a crucial aspect of numerical processing—representational precision—then we believe that a relatively small effect-size of .35 (or even smaller for many children, especially those in the early grades) makes ratio effects for symbolic numbers an unlikely candidate in this respect.

One might also object to this conclusion on other methodological grounds. For instance, the tasks themselves might be unreliable, or perhaps children's performance therein might be too variable to derive meaningful conclusions from subtler modulators of performance. In response, overall reliability for both symbolic and nonsymbolic comparison tasks was excellent for response-times ($\alpha s > .9$ in all cases but one, and $\alpha s > .94$ on average; see Table 2).[7] Another potential objection is that a subset of nonsymbolic comparison trials involved quantities entirely within the subitizing range (≤ 4). For the nonsymbolic task, ratio effects were significant for the large majority of children, so if anything we might be underestimating this result (note also that removing these trials did not substantially change the results, including the lack of correlation between symbolic and nonsymbolic ratio effects). Moreover, one might also consider that the range of ratios used here might be too narrow to detect meaningful modulation of performance by numerical ratio. As noted above, for the nonsymbolic task, ratio effects were significant for the large majority of children, so this objection would need to apply more specifically to the symbolic results. First, the ratios and quantities used here (.25 to .77) are quite similar to those used throughout the literature. This means that symbolic ratio effects in such cases should be viewed with skepticism, which is what we are arguing for in any event. Second, even if one were to find more reliable ratio effects in a symbolic number comparison task using ratios much closer to 1, it is worth noting that such ratios would begin to exceed the perceptual limits of nonsymbolic number comparison even in adults (usually around .85 to .90; Halberda & Feigenson, 2008; Pica et al., 2004). The upshot here is that one would need different ratios to detect ratio effects across the two formats, but such a turn would merely reinforce one of our other central claims: the underlying processes that generate ratio effects in symbolic and nonsymbolic comparisons are less related than has been previously assumed (see below for further discussion of this issue, especially with respect to the notion of representational precision). In summary, then, we retain the overall conclusion that placing too much theoretical stock in the interpretation of ratio effects in symbolic number comparison tasks is likely to be a mistake.

A final but important point to consider is that the current analyses and results focus almost exclusively on response times. We have already argued extensively (see Method and Supplementary Materials) why interpretation of error-rates in the current dataset is substantially compromised, so we will not repeat those arguments here. However, there are certainly cases where one would want to examine error-rates and be entirely justified in doing so. Indeed, especially for nonsymbolic numbers, a popular measure of comparison performance is the Weber-fraction (w), which relies exclusively on error-rates. First, it is worth noting that, strictly by virtue of the math involved, w is inextricably tied to the notion of ratio. In essence, one can see w as a modulating factor of the relation between ratio and performance—that is, *a relation between ratio and performance is a necessary assumption for estimating* w.[8] That said, given the current popularity of w in the literature, it remains to be seen whether our observations concerning the impact (or lack thereof, especially for symbolic numbers) of ratio effects on response times obtain or differ in a setting better optimized for capturing error data.

In conclusion, our results call into question the traditional interpretation of ratio effects as reflective of representational precision. For symbolic numbers in particular, the existence and broader meaning of ratio effects needs to be re-examined in a major way. Only about a third of individuals showed a statistically meaningful symbolic ratio effect, calling into question how much stock should be put into this effect at the population level. Our results also contradict the standard interpretation that developmental change in ratio effects is driven by change in representational precision for either symbolic or nonsymbolic numbers. We are not denying the possibility that numerical representations do not become more precise in older children; rather, we are saying that changes in ratio effects do not index this change. Furthermore, our results indicate that the sources of ratio effects for symbolic and nonsymbolic number comparisons are derived from independent sources. This is not only consistent with a growing body of evidence calling into question the link between symbolic and nonsymbolic number processing more broadly. Finally, our results show that the link between number comparison tasks and individual differences in arithmetic is both largely independent of ratio effects, and hence unlikely to be driven by representational precision. Instead, we suggest that greater attention be given to more general processing mechanisms, such as response selection, response variability, and various cognitive control mechanisms. Taken together, the relatively simple methodological point we make here—computing the statistical magnitude, or effect-size of ratio effects at the individual instead of the sample level—leads us to call into question several major assumptions in the field of numerical cognition, and to shift focus to alternative interpretations that have recently gained support elsewhere in the literature.

[7] Admittedly, reliabilities were notably lower for errors; but here again, we must note that the current tasks were optimized for measuring response-times, and reiterate our own skepticism when it comes to overly interpreting the error data presented in the Supplementary section. As such, our central conclusions focus on response-times, which, as noted, showed excellent task-reliability levels for both comparison tasks.

[8] When estimating w one fits the equation, $\widehat{ER} = \frac{1}{2}\mathrm{erfc}\left(\left(\frac{1}{w}\right)\frac{R-1}{\sqrt{2(R^2+1)}}\right)$, where $\widehat{ER}$ is the estimated error-rate (i.e., probability of making an error), R is the ratio: $\frac{\max(n_1, n_2)}{\min(n_1, n_2)}$. If one removes w from the equation and simply evaluates $\widehat{ER} = \frac{1}{2}\mathrm{erfc}\left(\frac{R-1}{\sqrt{2(R^2+1)}}\right)$, the result will be correlated with R^{-1} at .99. In other words, the equation sans w is essentially an estimate of the relation between ratio and error-rates. It is in this way (a) that calculating w assumes a relation between ratio and error-rates, and (b) w serves essentially as a modulating factor of this assumed relation.

References

Blomert, L., & Vaessen, A. (2009). *Differentiaal Diagnostiek van Dyslexie: Cognitieve analyse van lezen en spellen* [*Dyslexia Differential Diagnosis: Cognitive analysis of reading and spelling*]. Amsterdam: Boom Test Publishers.

Bonny, J. W., & Lourenco, S. F. (2013). The approximate number system and its relation to early math achievement: Evidence from the preschool years. *Journal of Experimental Child Psychology, 114,* 375–388. http://dx.doi.org/10.1016/j.jecp.2012.09.015

Buckley, P. B., & Gillman, C. B. (1974). Comparisons of digits and dot patterns. *Journal of Experimental Psychology, 103,* 1131–1136. http://dx.doi.org/10.1037/h0037361

Bugden, S., Price, G. R., McLean, D. A., & Ansari, D. (2012). The role of the left intraparietal sulcus in the relationship between symbolic number processing and children's arithmetic competence. *Developmental Cognitive Neuroscience, 2,* 448–457. http://dx.doi.org/10.1016/j.dcn.2012.04.001

Bulthé, J., De Smedt, B., & Op de Beeck, H. P. (2014). Format-dependent representations of symbolic and non-symbolic numbers in the human cortex as revealed by multi-voxel pattern analyses. *NeuroImage, 87,* 311–322. http://dx.doi.org/10.1016/j.neuroimage.2013.10.049

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14,* 365–376. http://dx.doi.org/10.1038/nrn3475

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159. http://dx.doi.org/10.1037/0033-2909.112.1.155

Damarla, S. R., & Just, M. A. (2013). Decoding the representation of numerical values from brain activation patterns. *Human Brain Mapping, 34,* 2624–2634. http://dx.doi.org/10.1002/hbm.22087

Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York, NY: Oxford University Press.

Dehaene, S. (2008). Symbols and quantities in parietal cortex: Elements of a mathematical theory of number representation and manipulation. In P. Haggard & Y. Rossetti (Eds.), *Sensorimotor foundations of higher cognition (attention and performance)* (pp. 527–574). New York, NY: Oxford University Press.

Dehaene, S., & Changeux, J. P. (1993). Development of elementary numerical abilities: A neuronal model. *Journal of Cognitive Neuroscience, 5,* 390–407. http://dx.doi.org/10.1162/jocn.1993.5.4.390

Dehaene, S., & Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron, 56,* 384–398. http://dx.doi.org/10.1016/j.neuron.2007.10.004

De Smedt, B., Noël, M. P., Gilmore, C., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends in Neurosciences and Education, 2,* 48–55. http://dx.doi.org/10.1016/j.tine.2013.06.001

De Smedt, B., Verschaffel, L., & Ghesquière, P. (2009). The predictive value of numerical magnitude comparison for individual differences in mathematics achievement. *Journal of Experimental Child Psychology, 103,* 469–479. http://dx.doi.org/10.1016/j.jecp.2009.01.010

De Vos, T. (2010). *Tempo test automatiseren*. Amsterdam: Boom Test Publishers.

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,* 295. http://dx.doi.org/10.3389/fpsyg.2015.00295

Eger, E., Michel, V., Thirion, B., Amadon, A., Dehaene, S., & Kleinschmidt, A. (2009). Deciphering cortical number coding from human brain activity patterns. *Current Biology, 19,* 1608–1615.

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. http://dx.doi.org/10.1016/j.jecp.2014.01.013

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

Feigenson, L., Libertus, M. E., & Halberda, J. (2013). Links between the intuitive sense of number and formal mathematics ability. *Child Development Perspectives, 7,* 74–79. http://dx.doi.org/10.1111/cdep.12019

Fuhs, M. W., & McNeil, N. M. (2013). ANS acuity and mathematics ability in preschoolers from low-income homes: Contributions of inhibitory control. *Developmental Science, 16,* 136–148. http://dx.doi.org/10.1111/desc.12013

Gallistel, C. R., & Gelman, I. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in Cognitive Sciences, 4,* 59–65. http://dx.doi.org/10.1016/S1364-6613(99)01424-2

Gebuis, T., & Reynvoet, B. (2012). The interplay between nonsymbolic number and its continuous visual properties. *Journal of Experimental Psychology: General, 141,* 642–648. http://dx.doi.org/10.1037/a0026218

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,* 1457–1465. http://dx.doi.org/10.1037/a0012682

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,* 11116–11120. http://dx.doi.org/10.1073/pnas.1200196109

Holloway, I. D., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children's mathematics achievement. *Journal of Experimental Child Psychology, 103,* 17–29. http://dx.doi.org/10.1016/j.jecp.2008.04.001

Hubbard, E. M., Diester, I., Cantlon, J. F., Ansari, D., Opstal, Fv., & Troiani, V. (2008). The evolution of numerical cognition: From number neurons to linguistic quantifiers. *Journal of Neuroscience, 19,* 11819–11824.

Inglis, M., & Gilmore, C. (2014). Indexing the approximate number system. *Acta Psychologica, 145,* 147–155. http://dx.doi.org/10.1016/j.actpsy.2013.11.009

Janssen, J., Verhelst, N., Engelen, R., & Scheltens, F. (2010). *Wetenschappelijke verantwoording van de toetsen LOVS rekenen-wiskunde voor groep 3 tot en met 8* [Scientific justification of the mathematical test for grades 1 to 6]. Arnhem, The Netherlands: Cito.

Libertus, M. E., & Brannon, E. M. (2009). Behavioral and neural basis of number sense in infancy. *Current Directions in Psychological Science, 18,* 346–351.

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. http://dx.doi.org/10.1111/j.1467-7687.2011.01080.x

Lonnemann, J., Linkersdörfer, J., Hasselhorn, M., & Lindberg, S. (2011). Symbolic and non-symbolic distance effects in children and their connection with arithmetic skills. *Journal of Neurolinguistics, 24,* 583–591. http://dx.doi.org/10.1016/j.jneuroling.2011.02.004

Lyons, I. M., & Ansari, D, (2009). The cerebral basis of mapping nonsymbolic numerical quantities onto abstract symbols: An fMRI training study. *Journal of Cognitive Neuroscience, 21,* 1720–1735.

Lyons, I. M., Ansari, D., & Beilock, S. L. (2012). Symbolic estrangement: Evidence against a strong association between numerical symbols and the quantities they represent. *Journal of Experimental Psychology: General, 141,* 635–641. http://dx.doi.org/10.1037/a0027248

Lyons, I. M., Ansari, D., & Beilock, S. L. (2015). Qualitatively different coding of symbolic and nonsymbolic numbers in the human brain.

Human Brain Mapping, 36, 475–488. http://dx.doi.org/10.1002/hbm.22641

Lyons, I. M., & Beilock, S. L. (2011). Numerical ordering ability mediates the relation between number-sense and arithmetic competence. *Cognition, 121,* 256–261. http://dx.doi.org/10.1016/j.cognition.2011.07.009

Lyons, I. M., Price, G. R., Vaessen, A., Blomert, L., & Ansari, D. (2014). Numerical predictors of arithmetic success in grades 1–6. *Developmental Science, 17,* 714–726. http://dx.doi.org/10.1111/desc.12152

Maloney, E. A., Risko, E. F., 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. http://dx.doi.org/10.1016/j.actpsy.2010.01.006

Mazzocco, M. M., Feigenson, L., & Halberda, J. (2011). Preschoolers' precision of the approximate number system predicts later school mathematics performance. *PLoS ONE, 6,* e23749.

Moeller, K., Klein, E., Nuerk, H. C., & Cohen Kadosh, R. (2012). A unitary or multiple representations of numerical magnitude? The case of structure in symbolic and non-symbolic quantities. *Frontiers in Psychology, 3,* 191. http://dx.doi.org/10.3389/fpsyg.2012.00191

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature, 215,* 1519–1520. http://dx.doi.org/10.1038/2151519a0

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

Nuerk, H. C., Kaufmann, L., Zoppoth, S., & Willmes, K. (2004). On the development of the mental number line: More, less, or never holistic with increasing age? *Developmental Psychology, 40,* 1199–1211. http://dx.doi.org/10.1037/0012-1649.40.6.1199

Park, J., & Brannon, E. M. (2014). Improving math with number sense training: An investigation of its underlying mechanism. *Cognition, 133,* 188–200. http://dx.doi.org/10.1016/j.cognition.2014.06.011

Piazza, M. (2010). Neurocognitive start-up tools for symbolic number representations. *Trends in Cognitive Sciences, 14,* 542–551. http://dx.doi.org/10.1016/j.tics.2010.09.008

Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., . . . Zorzi, M. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition, 116,* 33–41. http://dx.doi.org/10.1016/j.cognition.2010.03.012

Piazza, M., Pinel, P., Le Bihan, D., & Dehaene, S. (2007). A magnitude code common to numerosities and number symbols in human intraparietal cortex. *Neuron, 53,* 293–305.

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

Raven, J., Court, J. H., & Raven, J. C. (1995). *Coloured progressive matrices*. Oxford, United Kingdom: Oxford Psychologists Press.

Sasanguie, D., Vanden Bussche, E., & Reynvoet, B. (2012). Predictors for mathematics achievement? Evidence from a longitudinal study. *Mind, Brain, and Education, 6,* 119–128. http://dx.doi.org/10.1111/j.1751-228X.2012.01147.x

Sekuler, R., & Mierkiewicz, D. (1977). Children's judgments of numerical inequality. *Child Development, 48,* 630–633.

Šidak, Z. K. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association, 62,* 622–633.

Tzelgov, J., Zohar-Shai, B., & Nuerk, H. C. (2013). On defining quantifying and measuring the SNARC effect. *Frontiers in Psychology, 4,* 302. http://dx.doi.org/10.3389/fpsyg.2013.00302

Van Bon, W. H. J. (1986). *Raven's colored progressive matrices: Nederlandse normen en enige andere uitkomsten van onderzoek* [Raven's Colored Progressive Matrices: Dutch norms and other results.]. Lisse, Netherlands: Swets Test Services.

Van Opstal, F., Gevers, W., De Moor, W., & Verguts, T. (2008). Dissecting the symbolic distance effect: Comparison and priming effects in numerical and nonnumerical orders. *Psychonomic Bulletin & Review, 15,* 419–425. http://dx.doi.org/10.3758/PBR.15.2.419

Verguts, T., & Fias, W. (2004). Representation of number in animals and humans: A neural model. *Journal of Cognitive Neuroscience, 16,* 1493–1504. http://dx.doi.org/10.1162/0898929042568497

Verguts, T., & Van Opstal, F. (2005). Dissociation of the distance effect and size effect in one-digit numbers. *Psychonomic Bulletin & Review, 12,* 925–930. http://dx.doi.org/10.3758/BF03196787

Appendix

Shows Quantities Used for All 64 Trials

Left	Right	Ratio	Left	Right	Ratio
4	1	.250	33	11	.333
8	2	.250	39	13	.333
2	8	.250	11	33	.333
1	4	.250	30	10	.333
2	6	.333	36	12	.333
3	9	.333	12	36	.333
9	3	.333	15	45	.333
6	2	.333	13	39	.333
4	2	.500	11	22	.500
3	6	.500	14	28	.500
1	2	.500	22	11	.500
8	4	.500	13	26	.500
6	3	.500	24	12	.500
2	4	.500	12	24	.500
4	8	.500	26	13	.500
2	1	.500	20	10	.500
3	5	.600	17	11	.647
5	3	.600	19	29	.655
2	3	.667	29	19	.655
6	9	.667	21	32	.656
3	2	.667	15	10	.667
9	6	.667	12	18	.667
6	4	.667	18	12	.667
4	6	.667	20	30	.667
5	7	.714	11	15	.733
7	5	.714	15	11	.733
6	8	.750	12	16	.750
4	3	.750	16	12	.750
3	4	.750	17	13	.765
8	6	.750	13	17	.765
9	7	.778	10	13	.769
4	5	.800	13	10	.769

Note. Trials were the same for both symbolic and nonsymbolic comparison tasks. Trial presentation order was randomized.

Received January 12, 2015
Revision received May 6, 2015
Accepted June 11, 2015 ■