Numerical Order Processing in Children: From Reversing the Distance-Effect to Predicting Arithmetic

Ian M. Lyons[1] and Daniel Ansari[1]

ABSTRACT— Recent work has demonstrated that how we process the relative order—ordinality—of numbers may be key to understanding how we represent numbers symbolically, and has proven to be a robust predictor of more sophisticated math skills in both children and adults. However, it remains unclear whether numerical ordinality is primarily a by-product of other numerical processes, such as familiarity with overlearned count sequence, or is in fact a fundamental property of symbolic number processing. In a sample of nearly 1,500 children, we show that the reversed distance effect—a hallmark of symbolic ordinal processing—obtains in children as young as first grade, and is larger for *less* familiar sets of numbers. Furthermore, we show that the children's efficiency in evaluating the simplest ordered sequences (e.g., 2-3-4, 6-7-8) captures more unique variance in mental arithmetic than any other type of numerical sequence, and that this result cannot be accounted for by counting ability. Indeed, performance on just five such trials captured more unique mental arithmetic variance than any of several other numerical tasks assessed here. In sum, our results are consistent with the notion that ordinality is a fundamental property of how children process numerical symbols, that this property helps underpin more complex math processing, and that it shapes numerical processing even at the earliest stages of elementary education.

In the field of numerical cognition, relative order—ordinality—has until recently been a largely overlooked property of numbers. This is true of efforts to understand how the brain represents basic numerical representations, as well as research into how these basic processes underpin more sophisticated mathematical thinking. In the past few years, there has been a steady uptick in work on how we process the relative order of numbers (e.g., Brannon, 2002; Colomé & Noël, 2012; Delazer & Butterworth, 1997; Fias, Lammertyn, Caessens, & Orban, 2007; Franklin & Jonides, 2009; Franklin, Jonides, & Smith, 2009; Jou, 2003; Knops & Willmes, 2014; LeFevre & Bisanz, 1986; LeFevre, Kulak, & Bisanz, 1991; Lyons & Beilock, 2009, 2011, 2013; Lyons, Price, Vaessen, Blomert, & Ansari, 2014; Rubinsten & Sury, 2011; Turconi, Campbell, & Seron, 2006; Turconi & Seron, 2002; Zorzi, Di Bono, & Fias, 2011). Intriguingly, this work suggests that ordinality may be a key to understanding how we represent numbers symbolically (e.g., as Indo-Arabic numerals; Delazer & Butterworth, 1997; Lyons & Beilock, 2009, 2013; Turconi & Seron, 2002). Moreover, how efficiently an individual can assess the relative numerical order of a set of numbers has proven to be a strong and robust predictor of more sophisticated math skills, such as mental arithmetic, in both adults (Lyons & Beilock, 2011) and children (Lyons et al., 2014). For instance, Lyons and Beilock (2011) showed that adults' performance on a simple numerical-ordering task predicted nearly half the variance in their performance on a complex mental arithmetic task. This result remained robust even after controlling for several potentially confounding factors, such as working memory, non-numerical ordering, and several other numerical tasks. Lyons et al. (2014) showed in a large sample of Dutch children that by Grade 6 numerical-ordering performance was a better predictor of mental arithmetic performance than seven other numerical tasks (and remained so even after controlling for non-numerical factors as well, such as reading ability and nonverbal intelligence).

[1]Department of Psychology, University of Western Ontario

Address correspondence to Daniel Ansari, Department of Psychology, The University of Western Ontario, Westminster Hall, Room 325, London, Ontario N6A 3K7, Canada; e-mail: daniel.ansari@uwo.ca

Such a robust and unique predictor of a crucial developmental skill—arithmetic—is likely to be of keen interest to researchers and educators alike. However, there is currently a gap in our understanding of precisely *why* the ability to judge the relative order of a few numerals is so strongly related to more sophisticated math processing. One approach is to look for other known cognitive factors that might explain this relationship, such as working memory, reading ability, other numerical factors, and so on. As noted above though, many such factors have already been shown to fall short of fully accounting for the relation between ordering and arithmetic. Although one could theoretically never exhaust the infinite set of possible external factors, work to date does suggest that there may be something unique to the ordinality of numbers, and so perhaps one might instead look *within* the ordinal processing task itself.

In other words, we can ask whether there are subcomponents of the task we use to measure numerical-ordering proficiency which capture more or less variance in arithmetic ability. By knowing which components are most predictive, we can thus refine our search for the underlying mechanisms that lie closest to the heart of our ultimate goal—understanding how and which basic numerical skills underpin more complex ones. Note that this approach is not without precedent. For instance, Gilmore et al. (2013) decomposed a simple nonsymbolic quantity comparison task (which of two dot arrays contains more dots) into trials where the non-numerical, continuous parameters of the arrays were either congruent or incongruent with the goal of the task. The authors found it was the incongruent trials that largely accounted for the widely reported relation between performance on the comparison task and mental arithmetic. This in turn led the authors to consider cognitive control as a key factor in explaining this relation—an hypothesis that was confirmed by subsequent analysis. Here, our intention is similar: by decomposing the ordering task into subsets of trials, we can potentially begin to dig deeper into understanding the relation between numerical ordering and arithmetic.

This approach is not without its pitfalls, however—one of which is collinearity. Performance on subconditions within a task can often be highly correlated. Multicollinearity reduces the amount of information that may contribute to a given regression model, which can be thought of as effectively reducing the sample size (Baguley, 2012). In this study, we overcome this limitation by assessing numerical ordinal and mental arithmetic processing in a large sample of nearly 1,500 Dutch children (over 200 children in each grade, 1–6).

A second concern is theoretical: how should subsets of trials be identified—that is, which subconditions should be considered? A classic result in the field of numerical cognition is that when comparing two numbers, performance worsens (longer response times, higher error rates) as the numerical distance between the two numbers decreases (e.g., performance when comparing {4, 5} is worse than when comparing {3, 5}; Buckley & Gillman, 1974; Moyer & Landauer, 1967). When assessing the relative order of numbers, however, this canonical effect is sometimes reversed: performance worsens as the numerical distance between numbers *increases* (e.g., performance is *better* when assessing {4, 5, 6} relative to {3, 5, 7}; Franklin & Jonides, 2009; Franklin et al., 2009; Lyons & Beilock, 2013; Turconi et al., 2006). Interestingly, this effect is only seen when the numbers in question are in the correct order (e.g., increasing, from left-to-right, as in 1-2-3, but not 2-1-3). Furthermore, this result is specific to symbolic numbers (i.e., it is not observed for ordered dot arrays; Lyons & Beilock, 2013). This latter point is important because here we are concerned specifically with the previously reported strong relation between *symbolic* number-ordering ability and mental arithmetic. More broadly, the reversed distance effect appears to distinguish symbolic ordinal processing from other types of numerical processing, which in turns suggests a clear framework for decomposing the ordinal task by distance and whether or not the items are in the correct order.

Note that previous work showing reversed distance effects has primarily focused on highly literate adults, whose number-symbol systems are likely relatively mature and overlearned. It remains something of a question, then, whether children toward the beginning of formal education and experience regularly using number symbols in math contexts, such as arithmetic, will also show a reversal of the distance effect. In other words, we face both the obligation and the opportunity to examine whether reversed distance effects obtain in elementary-age children. If so, this would suggest that how we process ordinality in symbolic numbers is distinguished from other types of numerical processing from the earliest stages when mastery over these symbols has only just begun.

Perhaps the current leading explanation for the reversed distance effect is that it is a by-product of familiarity with a recited sequence (e.g., the counting or alphabetical sequence: Bourassa, 2014; Franklin et al., 2009; LeFevre & Bisanz, 1986; Lovelace & Snodgrass, 1971; Lyons & Beilock, 2013; Turconi et al., 2006). That is, one directly retrieves and matches an ordered sequence with a subset of the count list: for instance, 1-2-3 and 5-6-7 are more familiar and thus more rapidly identified as "in-order" than 1-3-5 or 4-6-8. This view suggests that the reversed distance effect should be reduced or even eliminated for less familiar number sequences. Reduced familiarity should slow the retrieval process, and may even lead one to revert to the more cumbersome process of directly comparing each pair of numbers (which in turn would predict more canonical distance effects; Turconi et al., 2006). Here, we examined ordinal processing for both single- and double-digit number sequences. Double-digit sequences (e.g., 29-30-31), being less familiar, should thus show a reduced reversal of or even canonical distance effect (note that double-digit sequences in the current data set also crossed decades to prevent children from simply ignoring the tens digit; Franklin et al., 2009). Alternatively, if knowing their correct relative order is a fundamental property by which we process number symbols (and not just an incidental by-product of rehearsing familiar chunks of the count list), then one would expect the fallback or default mode of *ordinal* processing to be not comparison, but ordinality itself. In this view, one would expect reversed distance effects to be *larger* for double- relative to single-digit trials.

To summarize, our central aims in this study are twofold. (1) We examine the nature of symbolic ordinal processing of numbers in children at the beginning of their formal education. We do so by testing for the presence of reversed distance effects in children as young as 6 years, and whether or not these effects are attenuated or accentuated when processing less familiar sequences. (2) We examine which subcomponents of ordinal processing capture the most unique variance with respect to more complex math processing—in this case, mental arithmetic. Doing so will help to pinpoint the crucial cognitive factors that link ordinal and other types of mental math, and may lead to more targeted recommendations for researchers and teachers alike in designing math interventions and curricula.

METHODS

Participants

The data collection protocol was approved by the ethics review board at Maastricht University. A total of 1,512 Dutch children in Grades 1–6 completed the ordering task. Chance performance is difficult to interpret, so we removed children who performed at chance on any of the tasks for which chance could be defined (>49% error rates). This removed 24 children from the analysis (1.59%); from each grade, respectively, 16 (6.83%), 2 (0.88%), 2 (0.75%), 4 (1.43%), 0, and 0. The overall final sample size was $N = 1{,}488$ (775 female); Grade 1 $n = 218$ (100 female), Grade 2 $n = 226$ (117 female), Grade 3 $n = 266$ (140 female), Grade 4 $n = 276$ (148 female), Grade 5 $n = 257$ (137 female), Grade 6 $n = 245$ (133 female).

Children in all grades completed the ordering task using single-digit numbers. Stimuli for children in Grades 2–6 also included double-digit numbers (see task description below). Hence, for analyses involving two-digit numbers, the sample was restricted to Grades 2–6 only.

In addition, for analyses examining arithmetic achievement, we were unable to collect data on the arithmetic task from eight additional children (1, 1, 0, 1, 4, 1 children in Grades 1–6, respectively), so these analyses proceeded with an N of 1,480. It is important to note that the data reported here are part of a larger data set, 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 Lyons et al. Furthermore, the sample *n*s do 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 thus resulted in omission of a few more participants in each grade in that article.

Procedure

Children were from seven different primary schools in the Netherlands. Parents denied consent by returning an enclosed nonconsent 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 1 and 2 and Grades 5 and 6, and in two sessions separated by no more than 5 days for Grades 3 and 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.

Ordering Task

In the ordering task, children saw three numbers presented horizontally as Indo-Arabic numerals. Half the time, the three numbers were all in numerically increasing order (left–right). In the remaining trials, numbers were either in decreasing or mixed order. Children were instructed to push a button with their left hand if the numbers were all increasing (*in-order*) or a button with their right hand if they were not (*not-in-order*). Stimuli remained on the screen until the child responded. There were 28 single-digit trials and 28 double-digit trials (note that children in Grade 1 completed only the single-digit trials). The distances between numbers were roughly evenly divided across trials into distances of 1–3, where absolute distance was always symmetrical around the median number and distance for a given trial was calculated as (max−min)/2. Hence, for a given cell (order × distance × digits), there were four or five trials. Double-digit trials all crossed a decade unit (single-digit trials, by definition, did not). A complete list of all trials can be found in Table A1. Trial order was randomized. Reliability (computed for combined performance; see below) on this task was high, for both one-digit (Cronbach's $\alpha = .938$) and two-digit trials (Cronbach's $\alpha = .960$).

Performance was computed as a composite of response times and error rates (as in Lyons et al., 2014). Combining measures provides a more complete picture of overall performance in a given task; it halves the number of statistical tests

Table 1
Mean Performance on the Ordering Task for *Single-Digit* Trials, Broken Down by Grade and Subcondition

	Subcondition					
Grade	*i1*	*i2*	*i3*	*n1*	*n2*	*n3*
1	4,995	4,928	5,870	5,219	4,934	4,650
	(162)	(181)	(248)	(196)	(172)	(149)
2	3,237	3,379	3,762	3,532	3,541	3,478
	(94)	(114)	(144)	(113)	(127)	(102)
3	2,522	2,685	2,999	2,839	2,870	2,935
	(64)	(75)	(94)	(83)	(91)	(87)
4	2,289	2,361	2,522	2,444	2,503	2,571
	(72)	(63)	(76)	(72)	(80)	(82)
5	1,982	2,005	2,187	2,117	2,173	2,244
	(66)	(69)	(76)	(72)	(87)	(91)
6	1,577	1,665	1,830	1,831	1,880	1,818
	(41)	(41)	(57)	(60)	(61)	(61)
Overall	2,701	2,773	3,114	2,929	2,923	2,898
	(45)	(47)	(60)	(51)	(50)	(45)

Note: Overall is the average across grades (see also Figure 1a). Values in parentheses are standard errors of the mean.
i = in-order, n = not-in-order; numbers indicate numerical distance.

needed for the analysis as a whole, thus reducing the risk of false positives, and it implicitly controls for any variation in speed-accuracy trade-offs across tasks. Measures were combined according to the formula: $P = RT(1 + 2ER)$, where a higher value indicates worse performance. In essence, one can interpret this measure as reaction times (ms) after they have been penalized for inaccurate performance. The scale runs linearly between a child's actual average response time (where $P = RT$) for perfectly accurate performance (0% errors) and double that value ($P = 2RT$) for chance performance (50% errors). This method is identical to that used in Lyons et al. (2014).

Mental 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 problems 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, Verhelst, Engelen, & Scheltens 2010).

Additional Measures

For analyses where we attempted to reproduce the final model in Lyons et al. (2014), we included additional measures from the final model of that article. These were six additional numerical tasks: numeral comparison, dot comparison, object matching, counting, number-line estimation, and dot-quantity estimation; two non-numerical tasks: nonverbal intelligence and reading scores; and age. Owing to space limitations, descriptions of these are not listed here, but can be found in Appendix 2 (see also Lyons et al., 2014). Finally, note that, in keeping with Lyons et al. (2014), for those analyses, we removed an additional 89 participants ($N = 1{,}391$). Those additional participants were either missing data or performed at chance on one or more of the additional measures noted above.

RESULTS

(Reverse) Distance Effects

We begin by considering single- and double-digit ordering trials separately because double-digit data were only available for Grades 2–6. We explicitly contrast single- and double-digit effects at the end of this section.

Single-Digit Trials

Our primary interest in this analysis was the potential reversal of distance effects. As such, our main focus was the factor Distance and its modulation by Order. Data were analyzed in a 2(Order: in-order, not-in-order; within-participants) × 3(Distance: 1–3; within-participants) × 6(Grade: 1–6; between-participants) analysis of variance (ANOVA). A complete table of means and standard errors can be found in Table 1. Figure 1a shows Order × Distance, averaged across grades. All main effects and interactions were highly significant ($ps < 6.8\text{E-}9$, $ds > .25$), except for the main effect of Order, $F(1, 1{,}482) = 2.63$, $p = .105$, $d = .08$, and the Distance × Grade

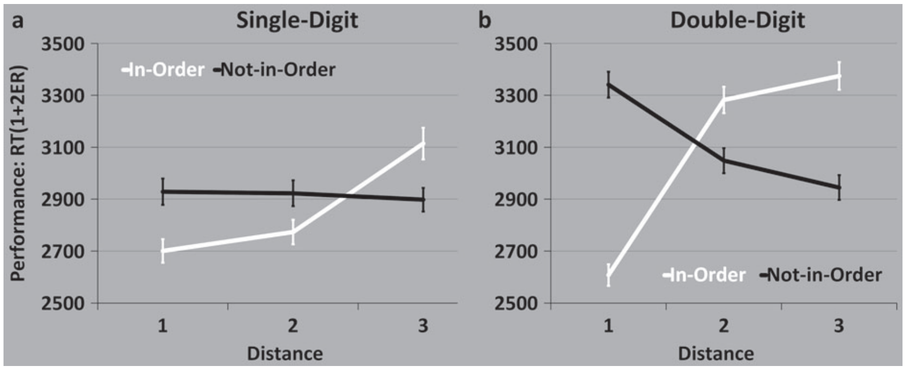

Fig. 1. (Reversed) distance effects for (a) single-digit and (b) double-digit trials. Data are collapsed across all grades (1–6 for single digit; 2–6 for double digits; see text and Tables 1 and 2 for grade-wise details).

interaction, $F(10, 2{,}964) = 1.43$, $p = .160$, $d = .14$. Crucially, our primary effect of interest, the Order × Distance interaction, was highly significant: $F(2, 2{,}964) = 32.72$, $p = 8.8\text{E-}15$, $d = .30$.

Examining Distance × Grade ANOVAs separately for each order, we found a main effect of Distance only for in-order trials. In-order: $F(2, 2{,}964) = 54.07$, $p = 8.6\text{E-}24$, $d = .38$; not-in-order: $F(2, 2{,}964) = 0.71$, $p = .494$, $d = .04$. The in-order effect of Distance was well fit by a linear contrast: $F(1, 1{,}482) = 80.36$, $p = 9.2\text{E-}19$, $d = .47$ (quadratic effect, $F = 17.05$). Figure 1a clearly shows that this was driven by a reversal of the distance effect, with performance worsening as distance increased. It is important to know, however, whether this effect was driven by just one or two grades (e.g., only older children), or whether it is a consistent effect which is present in all grades, including children as young as first graders. This question is especially pertinent because the in-order effect of Distance was qualified by an interaction with grade, $F(5, 1{,}482) = 4.56$, $p = 3.9\text{E-}4$, $d = .25$. A glance at Table 1 suggests an effect consistent across grades, so we next verified that a reversed distance effect (specifically, the linear contrast effect over Distance) obtained for each grade separately. It did: all Fs > 9.46, all ps < .002,[1] all ds ≥ .38 (all linear effects > quadratic effects). The interaction with grade appears to have been driven by larger effects of Distance in Grades 3 and 6 (ds = .75), compared with ds of .38–.52 for the other grades. Also of interest is that the reversed distance effect in the youngest children (Grade 1, $d = .49$) was comparable to that of the other grades.

To summarize the above results, symbolic ordinal processing of single-digit numbers in children appears to be quite similar to that seen in adults. Specifically, when stimuli are in the correct numerical order, there is a clear reversal of the distance effect. Note that this effect does not appear to be driven solely by trials with a distance of 1, but instead shows a consistent linear trend, at least up to distance 3 (all distances were significantly different from one another). Finally, reversed distance effects appeared to be relatively stable across grades—easily obtaining significance in all grades measured here, even in Grade 1.

Double-Digit Trials

As with single-digit trials, our primary interest in this analysis was the potential reversal of distance effects. As such, our main focus was the factor Distance and its modulation by Order. Data were analyzed in a 2(Order: in-order, not-in-order; within-participants) × 3(Distance: 1–3; within-participants) × 5(Grade: 2–6; between-participants) ANOVA. A complete table of means and standard errors can be found in Table 2. Figure 1b shows Order × Distance, averaged across grades. Consistent with single-digit trials, all main effects and interactions were highly significant (ps < 1.8E-5, ds > .24), excepting the main effect of Order, $F(1, 1{,}265) = 0.64$, $p = .424$, $d = .04$, and the Order × Grade, $F(4, 1{,}265) = 1.64$, $p = .163$, $d = .14$, and Distance × Grade, $F(8, 2{,}530 = 1.41$, $p = .187$, $d = .13$, interactions. Crucially, our primary effect of interest, the Order × Distance interaction, was highly significant: $F(2, 2{,}530) = 246.30$, $p = 1.8\text{E-}98$, $d = .88$.

Examining Distance × Grade ANOVAs separately for each order, we found main effects of Distance for both in-order and not-in-order trials. In-order: $F(2, 2{,}530) = 222.40$, $p = 1.1\text{E-}89$, $d = .84$; not-in-order: $F(2, 2{,}530) = 49.14$, $p = 1.2\text{E-}21$, $d = .39$. In both cases the effect of Distance was well fit by a linear contrast. In-order: $F(1, 1{,}265) = 371.96$, $p = 7.3\text{E-}73$, $d = 1.08$ (quadratic effect: $F = 70.91$); not-in-order: $F(1, 1{,}265) = 96.21$, $p = 6.1\text{E-}22$, $d = .55$ (quadratic effect: $F = 6.91$). Figure 1b clearly shows that this was driven by a reversal of the distance effect for in-order trials (performance worsened as distance increased) and a canonical distance effect for not-in-order trials (performance improved as distance increased). The

Table 2
Mean Performance on the Ordering Task for *Double-Digit* Trials, Broken Down by Grade and Subcondition

	Subcondition					
Grade	i1	i2	i3	n1	n2	n3
2	3,729	4,643	4,924	4,738	4,236	4,181
	(129)	(155)	(172)	(155)	(136)	(148)
3	2,927	3,754	3,859	3,665	3,340	3,305
	(79)	(105)	(107)	(101)	(109)	(95)
4	2,566	3,161	3,255	3,246	3,034	2,859
	(83)	(84)	(92)	(87)	(95)	(84)
5	2,146	2,708	2,695	2,780	2,540	2,431
	(62)	(91)	(77)	(82)	(83)	(82)
6	1,757	2,253	2,267	2,395	2,186	2,049
	(51)	(55)	(61)	(67)	(59)	(63)
Overall	2,608	3,282	3,375	3,341	3,048	2,945
	(41)	(50)	(53)	(50)	(48)	(47)

Note: Overall is the average across grades (see also Figure 1b). Values in parentheses are standard errors of the mean.
i = in-order, n = not-in-order; numbers indicate numerical distance.

in-order effect of Distance was qualified by an interaction with Grade, $F(4, 1{,}265) = 9.61$, $p = 1.2\text{E-}7$, $d = .35$; the not-in-order effect of Distance was not, $F < 1$. As with single-digit above, we next verified that a reversed distance effect (specifically, the linear contrast effect over Distance for in-order trials) obtained for each grade separately. It did: all Fs > 77.66, all ps < 2.0E-16,[2] all ds ≥ 1.06 (all linear effects > quadratic effects). The interaction with Grade appears to have been driven by slightly larger effects of Distance in Grades 3 and 6 (ds = 1.26–1.29), compared with ds of 1.06–1.10 for the other grades. Also of interest is that the reversed distance effect in the youngest children (Grade 2, $d = 1.06$) was comparable to that of the other grades.

To summarize the above results, symbolic ordinal processing of double-digit numbers in children appears to be similar to that seen in single-digit numbers. Specifically, when stimuli are in the correct numerical order, there is a clear reversal of the distance effect. As with single-digit numbers, this effect exhibited a linear trend (all distances were significantly different from one another), and was relatively stable across grades. We next examine whether the magnitude of reversed distance effects was greater or lesser for double- relative to single-digit trials.

Comparing Single- and Double-Digit Trials

One of our key questions concerns what happens to the reversal of the distance effect when the numbers involved become less familiar. If the reversal of distance effects in ordinal tasks is largely driven by familiarity, then we would expect the magnitude of this effect to be reduced in double-digit trials, as these numbers are far less frequent and thus also likely to be less familiar, especially to elementary-age children (Dehaene & Mehler, 1992). Conversely, if ordinal relations are a fundamental part of how we process and understand symbolic numbers, then one might expect children to fall back on this more basic process as items become less familiar. This latter account predicts that double-digit trials should show a larger reversal of the distance effect.

For simplicity, because the current hypotheses specifically concern reversed distance effects, and these effects are seen only in the in-order trials, we limited the analysis below to in-order trials (results are highly similar whether one includes not-in-order trials in the model or not). We thus computed a 2(Digits: single, double) × 3(Distance: 1–3) × 5(Grade: 2–6) ANOVA. All effects were highly significant (ps < 4.6E-8, ds > .28), excepting the three-way Digits × Distance × Grade interaction, $F(8, 2{,}530) = 1.08$, $p = .374$, $d = .12$. Crucially, our primary effect of interest, the Digits × Distance interaction, was highly significant: $F(2, 2{,}530) = 66.63$, $p = 6.3\text{E-}29$, $d = .46$. An examination of Figure 1 shows that the decrease in performance (recall that a greater value indicates worse performance) from distance 1 to distance 3 in double digits ($M = 767$, $SE = 41$) was nearly twice that seen in single digit ($M = 413$, $SE = 48$). Note also that this cannot be explained as simply due to overall worse performance on double-digit trials, as this result was even stronger after normalizing distance scores, respectively, in terms of average single- or double-digit increasing performance [(i3–i1)/mean(i1,i2,i3)]: double digits, $M = .240$, $SE = .010$; single digit, $M = .114$, $SE = .011$. In sum, the data clearly show that reversed distance effects are significantly larger for double- relative to single-digit trials, which is consistent with the ordinality hypothesis above. Note also that the lack of a three-way interaction indicates this result was not modulated by—that is, was consistent across—different grades.

Predicting Arithmetic

Distance effects (next section) were computed for each participant by correlating[3] performance with distance across trials (this was performed separately for in-order and not-in-order trials, as well as single-digit and double-digit trials). Note that a positive value indicates a reversal of the distance effect and a negative value indicates a canonical distance effect. For all other regression models (remaining sections), because the dependent measure (Arithmetic) was scored with a higher value indicating better performance, for ease of interpretation, performance scores were multiplied by −1 before being entered into regression analyses. This was so that a positive relation meant better performance on a given measure was related to better arithmetic performance.

(Reversed) Distance Effects

We first examined whether reversed distance effects (1) are predictive of mental arithmetic performance and (2) can help to explain the strong relation between ordering and arithmetic (Lyons & Beilock, 2011; Lyons et al., 2014). To begin with, better overall performance on both single- and double-digit ordering trials was highly correlated with mental arithmetic (single digit: $r_{1,478} = .686$, $p = 2.0\text{E-}206$; double digits: $r_{1,261} = .635$, $p = 8.3\text{E-}144$), capturing over 40% of the variance in each case.

For single-digit trials, both in-order and not-in-order distance effects were positively correlated with arithmetic performance (greater reversal of the distance effect was related to higher arithmetic scores; in-order: $r_{1,478} = .108$, $p = 3.0\text{E-}5$; not-in-order: $r_{1,478} = .084$, $p = 1.2\text{E-}3$). For double-digit trials, this was the case only for in-order trials (in-order: $r_{1,261} = .113$, $p = 5.9\text{E-}5$; not-in-order: $r_{1,261} = -.034$, $p = .228$). In all cases, however, correlations were substantially smaller than those seen for overall ordering performance, accounting at most for only about 1.3% of arithmetic variance. Moreover, in no case did reversed distance effects mediate the relation between overall ordering and arithmetic performance (all ps > .10).

In other words, computing reversed distance effects based on a combination or derivation of subconditions proved of little use in explaining the relation between numerical ordering and arithmetic performance. This is perhaps not surprising as recent work has demonstrated that such derived measures—e.g., distance effects, ratio effects, Weber fractions, and so on—in basic numerical tasks are typically less reliable predictors of math achievement than is simple mean performance on such tasks (Dietrich, Huber, & Nuerk, 2015; Inglis & Gilmore, 2014). For current purposes, the poor predictive power of reversed distance effects simply underscores the need for a different tack. In the following sections, we instead decompose ordering performance into its various subconditions and examine which of these captures the most unique variance in mental arithmetic scores.

Single-Digit Trials

Single-digit ordering trials were divided into six subconditions based on whether stimuli were in-order (i), not-in-order (n) and distance (1–3) (subconditions are written hereafter as i1, i2, i3, n1, n2, n3). Average performance for each participant was computed across the four or five trials in each subcondition. Correlations with arithmetic performance for each subcondition were .641, .575, .544, .562, .535, .534 (respectively, all dfs = 1,478, ps ≤ 6.0E-110), indicating consistently high correlations for all subconditions, near to that seen for the ordering task overall (.686) and capturing about 29%–41% of arithmetic variance. The correlation for i1 was significantly greater than that for the other five subconditions (all ps < 1.0E-4), suggesting that this subcondition might be especially closely related to arithmetic processing. On the other hand, these differences are rather subtle (.066–.107), and obtain significance here primarily due to the particularly large sample. Moreover, it may be of greater interest to examine *unique* contributions made by each subcondition to arithmetic variance. That is, we expect all of the subconditions to capture some degree of common arithmetic variance; however, we want to know whether any of the individual subconditions captures substantially more additional variance beyond this common core.

Table 3
Single-Digit Regression Model Results

Predictor	*F*	*p*
i1	87.36	3.3E-20
i2	10.72	1.1E-03
i3	0.05	8.3E-01
n1	7.81	5.3E-03
n2	0.02	8.9E-01
n3	0.04	8.5E-01
i1 × Grade	11.95	2.4E-11
i2 × Grade	2.64	2.2E-02
i3 × Grade	2.77	1.7E-02
n1 × Grade	2.58	2.5E-02
n2 × Grade	2.85	1.5E-02
n3 × Grade	0.94	4.5E-01
Grade	126.05	8.0E-111
Intercept	5910.00	<1.0E-999

Note: Overall model fit: $R^2 = .730$, adjusted $R^2 = .722$. For grade and all interaction terms, numerator $df = 5$; for all other predictors, $df = 1$. Error (denominator) $df = 1{,}438$.

To test this, all six subconditions were entered as simultaneous predictors of arithmetic performance, along with Grade (six levels, 1–6), and the interaction with Grade for each subcondition. These interaction terms effectively tested for developmental effects: they tested whether the slope of the relation between a given subcondition and arithmetic varied as a function of grade. Model results can be found in Table 3. Collinearity was within acceptable limits (all variance inflation factors [VIFs] ≤ 2.54, where a VIF > 10 is typically viewed as problematic; e.g., Baguley, 2012; Neter, Wasserman, & Kutner, 1989). The i1 (in-order, distance 1) subcondition clearly showed the strongest overall unique predictive value. It also showed the strongest interaction effect, although all subconditions, save n3, showed a small but significant interaction effect.

It is important to note that we are talking about unique variance—that is, variance captured by the i1 subcondition *over and above* that captured by the other subconditions. Indeed, if we simply enter the six subconditions as predictors (for simplicity, leaving out Grade and interactions with Grade), we find the R^2 for the overall model is .488, or only a

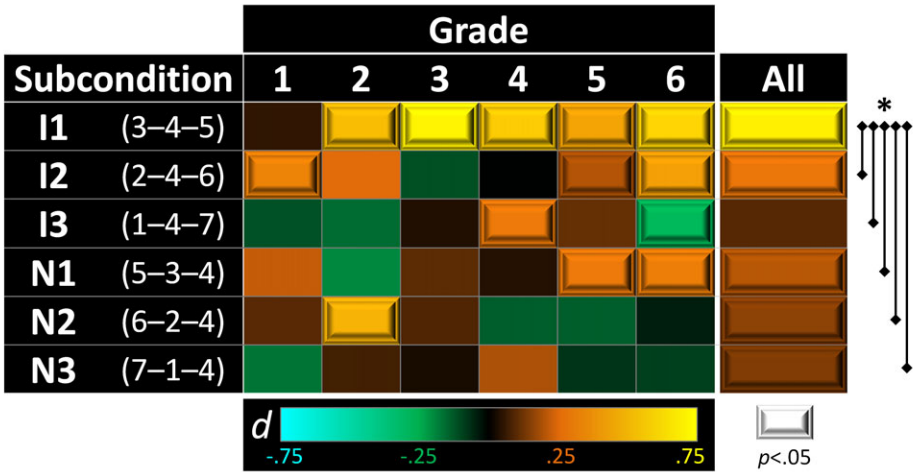

Fig. 2. Unique contributions (in terms of predicting arithmetic performance) of each subcondition at each grade for *single-digit* trials. Numbers in parentheses are example trials for each subcondition. Owing to varying degrees of freedom across grades, contributions are converted to effect sizes (*d*). Note that *d* of about ±.25 corresponds to roughly $p = .05$ for individual grades and *d* of about ±.10 corresponds to $p = .05$ for "All." All: unique contributions when computed across all grades. *Comparison of unique effects (computed based on "All"): i1 versus each of the other conditions, where * indicates $p < 3.0\text{E-}7$.

bit higher than the raw correlation for the i1 condition from above: $.641^2 = .411$, and it is quite similar to what we see if we simply average over all single-digit trials, ignoring subcondition: $R^2 = .471$. This is important when interpreting the results in this section (Table 3, Figure 2): there is variance common to all subconditions that is predictive of arithmetic variance. What these analyses show is the contribution each subcondition makes beyond this common variance, and in particular, that it is the i1 condition that accounts for most such unique variance.

Given the strong interaction effect for i1, Figure 2 visualizes the unique contributions of each subcondition at each grade (a separate multiple regression model was run at each grade), along with the overall effect (i.e., collapsing across all grades). Given the varying degrees of freedom across grades, these were converted to effect sizes (*d*s). In terms of overall effects, as with the zero-order correlations above, the unique predictive value of the i1 condition was significantly greater than all other subconditions. Moreover, we can see this predictive value rose sharply between Grades 1 and 2, and remained consistently higher than the other subconditions thereafter.

To summarize, single digit, in-order trials at distance 1 (i1, e.g., 1-2-3, 6-7-8) clearly captured the most unique variance in arithmetic processing.

Double-Digit Trials

Double digit-ordering trials were subdivided and computed in the same manner as single-digit trials above. Correlations with arithmetic performance for each subcondition were .541, .554, .554, .508, .432, .523 (respectively, all $df\text{s} = 1{,}261$, $p\text{s} \leq 1.4\text{E-}58$), indicating consistently high correlations for all subconditions, near to that seen for the ordering task overall (.635), and capturing about 19%–31% of arithmetic variance. No single condition appeared to stand out above the rest, although in-order subconditions did appear to show slightly stronger correlations than not-in-order subconditions. Collapsing across distance, the correlation for in-order trials (.625) was significantly greater ($p = .003$) than that for not-in-order trials (.569).

Turning to unique contributions, these were computed in the same manner as single-digit trials above. Collinearity was acceptable (all VIFs ≤ 2.60). Model results can be found in Table 4. All subconditions showed significant main effects, with small but significant interaction effects for i1 and n1. Figure 3 breaks unique contributions down by grade, although it is difficult to discern an overarching pattern, as is perhaps indicated by the relatively weak interaction effects. Looking instead at the overall results, we can see that, consistent with the zero-order correlations above, it appears that whether or not trials are in order is the most consistent effect. As a post-hoc test of this intuition, we can simply collapse across distance and enter in-order and not-in-order performances as competing predictors, compute the partial correlations for each (in-order: $r_{p(1,\,260)} = .352$, $p = 3.5\text{E-}38$; not-in-order: $r_{p(1,\,260)} = .168$, $p = 1.8\text{E-}9$), and test whether the former correlation is greater than the latter (it is: $p = 6.6\text{E-}7$).

To summarize, for double digits, there did not appear to be any one subcondition that captured overwhelmingly more unique arithmetic variance than the others (i.e., as

Table 4
Double-Digit Regression Model Results

Predictor	*F*	*p*
i1	20.31	7.2E-06
i2	18.25	2.1E-05
i3	15.22	1.0E-04
n1	7.20	7.4E-03
n2	6.41	1.1E-02
n3	9.36	2.3E-03
i1 × Grade	2.77	2.6E-02
i2 × Grade	0.24	9.2E-01
i3 × Grade	1.20	3.1E-01
n1 × Grade	2.39	4.9E-02
n2 × Grade	1.71	1.5E-01
n3 × Grade	0.57	6.8E-01
Grade	48.52	6.4E-38
Intercept	6010.03	<1.0E-999

Note: Overall model fit: $R^2 = .557$, adjusted $R^2 = .545$. For Grade and all interaction terms, numerator $df = 4$; for all other predictors, $df = 1$. Error (denominator) $df = 1{,}228$.

Table 5
Model Results When Both Single- and Double-Digit Subconditions Are Entered Into the Model Together

Predictor	*F*	*p*
i1, 1-digit	53.38	4.9E-13
i2, 1-digit	0.03	8.6E-01
i3, 1-digit	0.28	6.0E-01
n1, 1-digit	0.03	8.7E-01
n2, 1-digit	0.17	6.8E-01
n3, 1-digit	0.03	8.7E-01
i1, 2-digit	6.25	1.3E-02
i2, 2-digit	13.00	3.2E-04
i3, 2-digit	3.17	7.5E-02
n1, 2-digit	2.43	1.2E-01
n2, 2-digit	4.89	2.7E-02
n3, 2-digit	17.13	3.7E-05
Grade	72.26	4.2E-55
Intercept	5975.34	<1.0E-999

Note: Overall model fit: $R^2 = .548$, adjusted $R^2 = .542$. For grade, $df = 4$; for all other predictors, $df = 1$. Error (denominator) $df = 1{,}246$.

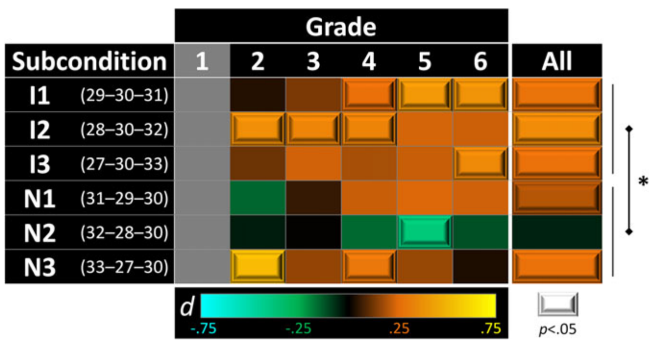

Fig. 3. Unique contributions (in terms of predicting arithmetic performance) of each subcondition at each grade for *double-digit* trials. Numbers in parentheses are example trials for each subcondition. Owing to varying degrees of freedom across grades, contributions are converted to effect sizes (d). Note that d of about ±.25 corresponds to roughly $p = .05$ for individual grades and d of about ±.11 corresponds to $p = .05$ for "All." All: unique contributions when computed across all grades. *Comparison of unique effects (computed based on "All"): None of the individual subconditions were significantly different from one another, with the exception that n2 was significantly less than all other subconditions. Instead, in-order (I) subconditions were overall greater than not-in-order (N) subconditions, where * indicates p =6.6E-7.

was the case for i1 single-digit trials). Instead, we see that in-order trials were overall (i.e., regardless of distance) better predictors than not-in-order trials.

Comparing Single- and Double-Digit Trials

Here, we were concerned with whether any (and which) of the subconditions, across single- and double-digit trials predicted the most unique arithmetic variance. The most direct approach is to simply enter all 12 subconditions (the 6 subconditions for each single- and double-digit trials) into a single regression model. Collinearity was acceptable (all VIFs ≤ 2.78). Results (Table 5) show that the i1 subcondition was the only single-digit subcondition to account for unique variance, whereas all of the double-digit subconditions did, with the exception of n2. On the other hand, the single-digit i1 condition was the strongest overall predictor, and was in fact statistically greater than all other subconditions (all $ps < .05$). Results were consistent across individual grades: the single-digit i1 subcondition accounted for significant ($p < .05$) unique variance in four of the five grades (excepting Grade 5); double-digit n3 obtained just two of five grades (Grades 2 and 4); no other subcondition obtained in more than one grade. In sum, considering the sample as a whole, although more of the double-digit subconditions captured unique variance, the most such variance was in fact captured by single-digit i1 trials (e.g., 1-2-3, 6-7-8). Moreover, this subcondition was the most consistent predictor of arithmetic across individual grades.

Reconstructing Regression Results from Lyons et al. (2014)

The single-digit i1 subcondition was the subcondition most strongly related to arithmetic—in terms of both raw correlations and amount of unique variance captured. Thus, we next examined whether this subcondition, alone, could reproduce one of the central results from Lyons et al. (2014). Overall performance on the same numerical-ordering task under scrutiny here was more predictive of arithmetic scores than any of the other numerical tasks (see Table 5 of that article). To do so, we took the final model from Lyons et al.

Table 6
Comparing Model Results From Lyons et al. (2014) and the Same Model With the Single-Digit i1 Subcondition Substituted for NumOrd

Predictor	*F*	*p*	*Predictor*	*F*	*p*
NumOrd	63.53	*3.3E-15*	*i1, 1-digit*	60.64	*1.4E-14*
NumComp	46.53	1.4E-11	NumComp	48.67	4.7E-12
DotComp	1.85	1.7E-01	DotComp	1.90	1.7E-01
ObjMatch	34.06	6.7E-09	ObjMatch	36.50	2.0E-09
Counting	17.67	2.8E-05	Counting	21.16	4.6E-06
DotEst	12.68	3.8E-04	DotEst	12.31	4.6E-04
NumLine	42.03	1.3E-10	NumLine	42.92	8.1E-11
Ravens	5.77	1.6E-02	Ravens	4.94	2.6E-02
Reading	37.65	1.1E-09	Reading	41.15	1.9E-10
Age	14.59	1.4E-04	Age	14.81	1.2E-04
NumOrd × Grade	9.80	3.3E-09	NumOrd × Grade	10.65	4.7E-10
ObjMatch × Grade	4.00	1.3E-03	ObjMatch × Grade	4.59	3.7E-04
Counting × Grade	2.35	3.9E-02	Counting × Grade	3.14	7.9E-03
NumLine × Grade	4.72	2.8E-04	NumLine × Grade	4.78	2.4E-04
Grade	55.25	2.4E-52	Grade	54.05	2.8E-51
Intercept	461.35	2.6E-88	Intercept	450.78	1.4E-86

Note: The left-hand side of the table shows final model results from Lyons et al. (2014), where NumOrd indicates performance on the numeral-ordering task, averaged across all trials (including double-digit, for Grades 2–6). The right side shows the same model with the single-digit i1 subcondition substituted for NumOrd. Overall model fit on the left: $R^2 = .8099$, adjusted $R^2 = .8050$. Overall model fit on the right: $R^2 = .8093$, adjusted $R^2 = .8043$. Both models: Grade and all interaction terms, numerator $df = 5$; for all other predictors, $df = 1$. Error (denominator) $df = 1{,}355$.

(2014), Table 5)[4] and reran it substituting the single-digit i1 subcondition for overall ordering performance ("NumOrd," in that article).

The left side of Table 6 shows the final model results from Lyons et al. (2014); the right side of Table 6 shows the same model with the single-digit i1 subcondition substituted for overall numeral-ordering performance (NumOrd). The two models are remarkably similar. Overall model fit was reduced by less than 0.1% of the variance. Moreover, the unique *F* value associated with ordering was reduced from 63.53 to 60.64 (or a corresponding decrease in partial-*r* values of less than .005).[5] In other words, average performance on the 5 single-digit i1 trials was very nearly as good a predictor of arithmetic performance as ordering performance averaged across all 56 trials (or 28 in Grade 1). Indeed, it is worth noting that in Lyons et al. (2014), children's ordering scores in Grades 2–6 were averaged over both single- and double-digit trials, which in turn makes it all the more remarkable that the single-digit i1 trials almost completely reproduces the original result (note also that results were consistent if restricted to just Grades 2–6).

In sum, single-digit i1 trials are not just uniquely predictive of arithmetic performance over and above other ordering subconditions—as was demonstrated in the preceding sections. In addition, this subcondition was able to largely reproduce the Lyons et al. (2014) results, which pit it against several other numerical and non-numerical tasks. This remarkable result points to the high diagnostic potential that can be achieved from just a handful of trials on a simple numeral-ordering task.

DISCUSSION

That ordinality plays a key role in understanding (especially symbolic) numerical cognition is becoming increasingly apparent. Whether numerical ordinality is primarily a by-product of other numerical processes (such as familiarity with overlearned count sequences) or is in fact a fundamental property of symbolic numerical processing remains unclear, however. To address this question, we examined the nature of symbolic ordinal processing of numbers in children at the beginning of their formal education.

Lyons and Beilock (2013) demonstrated that the nature of ordinal processing in symbolic numbers is fundamentally different than in nonsymbolic magnitudes (e.g., arrays of dots)—at least in literate adults. Moreover, one hallmark of the ordinal nature of number symbols is the reversed distance effect. Here, we tested for the presence of reversed distance effects in a large cohort of children in Grades 1–6. Overall, children showed robust and consistent reversed distance effects for ordered (in-order) sequences, and the magnitude of this effect was similar for first graders as in older children. To our knowledge, this is the youngest group of children in which reversed distance effects have been demonstrated. Crucially, this hallmark of distinctly symbolic ordinal processing appears to be present even in children who have only just begun their formal education. Although further work is needed to examine ordinal processing in number symbols in children who are first acquiring the use of number symbols, the current data suggest that the influence of ordinality may be present from the very outset of children's quest to understand and manipulate symbolic representations of number. Consistent with this notion, Lyons and Beilock (2009) trained adult participants to associate large, approximate quantities with a novel set of arbitrary visual shapes, and then tested participants' ability to use these new "symbols" in various numerical contexts. Results showed that it was precisely those participants who reported relying directly on an ordinal strategy (e.g., relative dot quantities were explicitly used to determine the novel symbols' relative order, which was then memorized) that proved most proficient at using the symbols in numerical contexts, including both numerical ordering and numerical comparison (i.e., which novel symbol indicated the greater quantity of dots). Taken together, it would appear that ordinality plays a fundamental role in how we process and understand number symbols—both as adults and, as we show here, in the early stages of formal education.

That said, it may be that reversed distance effects are not so much indicative of a fundamental property of number symbols but are more just a by-product of other numerical processes. For instance, one account of reversed distance effects is that they are essentially a consequence of our familiarity with the count sequence (Bourassa, 2014; Franklin et al., 2009; LeFevre & Bisanz, 1986; Lyons & Beilock, 2013; Turconi et al., 2006). However, the current data do not support this view. Instead, we find that the reversal of the distance effect is greater for *less* familiar number sequences. That is, single-digit numbers, occurring near the beginning of the standard count sequence and more frequently in general (Dehaene & Mehler, 1992), should be rehearsed more often than large numbers (Rundus, 1971), which in turn should increase one's overall familiarity with them (relative to larger numbers). And yet, we found that reversed distance effects for double-digit sequences ($d = 1.05$) were more than twice as large as those seen for single-digit sequences ($d = .47$). As numbers become *less* familiar, one is *more* likely to fall back on ordinal processing. Finally, counting ability was unrelated to reversed distance effects ($r = .036$, $p = .162$). Taken together, these results are consistent with the notion that ordinality is a fundamental, and not simply a derived property of symbolic number processing.

Consistent with this idea, recent work has demonstrated that numerical ordering ability is a strong and robust predictor of more sophisticated math skills, such as mental arithmetic in both adults (Lyons & Beilock, 2011) and children (Lyons et al., 2014). Here, we broke this ordering task into its constituent components and examined which of these captured the most unique variance with respect to mental arithmetic. Results showed that it was performance in the single-digit i1 subcondition (e.g., 2-3-4, 6-7-8) that was both the best raw (zero-order) correlate of mental arithmetic, and captured the most unique arithmetic variance—that is, over and above that captured by the other subconditions. Note also that Lyons et al. (2014) showed a developmental effect for the ordering task, wherein the unique arithmetic variance it captured rose from nonsignificant in Grade 1 to highly significant in later grades. Figure 2 in this article shows a similar pattern for the single-digit i1 subcondition, suggesting that these trials may have been driving not only the overall result but also the developmental change as well.

Here again, one might be tempted to attribute the predictive success of the i1 subcondition to the fact that these items appear to strongly evoke notions of counting. However, the contribution of the single-digit i1 subcondition remained robust even after controlling for children's counting ability. Indeed, the unique predictive capacity of the single-digit i1 subcondition outstripped counting (Table 6), the other ordering subconditions (Figure 2 and Tables 3 and 5), and all of the other numerical and non-numerical tasks considered here (Table 6; see also Lyons et al., 2014). In other words, children's proficiency at recognizing the relative order of arguably the most basic and quintessentially "ordinal" sets of numbers was both strongly and uniquely related to their mental arithmetic scores. This is precisely what one would expect under the assumption that ordinality is a fundamental property of number symbols that in turn forms a fundamental building block from which more sophisticated math abilities are constructed.

Another potential counter to the notion that ordinality is a fundamental component of number symbols is that children may simply be more familiar with specific sets of numbers, and our single-digit i1 trials just happen to comprise such sets. Here, it is important to note that the not-in-order versions of these items include the same numbers (see Table A1 in Appendix 1), so familiarity with the numbers themselves, or even specific sets of numbers could not explain the unique arithmetic variance captured by the single-digit i1 condition. Instead, a familiarity account would have to limit itself to familiarity *with that specific set of numbers in that specific order*. In other words, at this point, ordinality is effectively an assumed component of the explanation. Furthermore, the arithmetic problems of course do not include just these triplets of numbers, so it is difficult to imagine the underlying mechanism that explains the ordering–arithmetic relation based solely on familiarity with the numbers themselves. To be clear, we find that performance on ordered number sequences *predicts performance on arithmetic problems that do not necessarily involve these numbers or sequences*. Because this result cannot be attributed to either familiarity or counting (as noted earlier), and given the several lines of argument already discussed, we argue that the most straightforward explanation is that the single-digit i1 trials tap an underlying principle of numerical processing—ordinal understanding of number symbols—that drives performance on both the ordering and arithmetic tasks.

On a more applied note, it is worth highlighting the fact that the single-digit i1 subcondition comprised only five trials. That so much (unique) arithmetic variance can be accounted for by just a handful of trials is certainly remarkable. In particular, the ordering task described here may prove useful in educational settings, for instance as an easily administered and interpreted marker of children who might be in need of further math assistance. On the other hand, this study no doubt benefited from a large sample with the resources to test each child in one-on-one sessions. Hence, it remains to be seen whether the current results generalize to situations with fewer individuals and perhaps less well-controlled testing environments. In other words, for practical purposes, the current results appear promising for educators, but caution must be urged pending more rigorous field testing.

An important open question at the moment is how exactly ordinality is assessed during symbolic number processing. One simple possibility is via sequential (i.e., left–right) comparison of number pairs. On the other hand, Lyons and Beilock (2013) indicated that, although this assumption appears to be valid for nonsymbolic ordinal judgments, this is not the case for symbolic ordinal judgments (the latter being the focus of the current article). Moreover, if this were the case, one would expect better overall performance on not-in-order relative to in-order trials because the majority of the former could be rejected after considering only the first pair (whilst one would have to compare at least two pairs and apply an associative inference rule on the in-order trials; see Table A1 in Appendix 1). However, this was not the case: The main effects of Order were nonsignificant, with performance on in-order trials in fact slightly better than not-in-order trials (single digit: $p = .105$, in-order $\mu = 2{,}933$, not-in-order $\mu = 2{,}977$; double digits: $p = .424$, in-order $\mu = 3{,}110$, not-in-order $= 3{,}132$; recall that a lower number indicates better performance). Consistent with Lyons and Beilock (2013), this suggests that sequential, pairwise comparison was not the primary strategy employed by the majority of children. Another possibility is some form of verbal rehearsal. On the other hand, we have already seen how counting is unrelated to ordering performance, and it is unclear how such an explanation could account for the greater reversed distance effect in double- relative to single-digit trials. Still other possibilities include visuo-spatial processing (Knops & Willmes, 2014) or other types of ordinal processing such as serial-order working memory (e.g., Abrahamse, van Dijck, Majerus, & Fias, 2014). The bottom line, at present however, is that we simply do not know. That said, we believe the current results clearly demonstrate that ordinality is key to understanding how even elementary-age children process number symbols, and we hope that this in turn will spur further critical research into answering this question.

In conclusion, the current data strongly support the notion that ordinality is a fundamental aspect of the development of numerical processing. It is both key to understanding how children represent numbers symbolically and it appears to serve as a fundamental building block upon which children's arithmetic abilities are formed. Although more hands-on, in-class testing is needed, these results also may prove particularly useful for math educators in need of a rapid, reliable gauge of children's basic numerical proficiency.

Acknowledgments—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 Daniel Ansari. Data collection costs were paid in part by Boom Test Uitgevers Amsterdam BV.

NOTES

1 Note that, strictly speaking, correcting for multiple comparisons is not necessary, given that the goal here was to verify that reversed distance-effects obtain in *all* grades (not just in one of six grades). Nevertheless, all ps < .009, which is the corrected threshold for six comparisons using the Dunn–Šidák (Šidák, 1967) method.

2 Corrected threshold: $p = .010$.

3 Note that a correlation takes into account the variability as well as the raw slope of the relation between performance and distance. In this formulation, a positive value indicates a reversed distance effect, and a negative value indicates a canonical distance effect.

4 This model had Arithmetic as the dependent variable and 16 predictors: Grade (1–6), numeral ordering (NumOrd), numeral comparison (NumComp), dot comparison (DotComp), object matching (ObjMatch), counting (Counting), dot-quantity estimation (DotEst), number-line estimation (NumLine), nonverbal intelligence (Ravens), reading scores (Reading), and age (Age, centered on each grade); in addition to four interaction terms: Grade × NumOrd, Grade × ObjMatch, Grade × Count, and Grade × NumLine (see Lyons et al., 2014, for details on how this model was selected).

5 The next highest subcondition produced a unique F-value of 32.54.

REFERENCES

Abrahamse, E., van Dijck, J. P., Majerus, S., & Fias, W. (2014). Finding the answer in space: The mental whiteboard hypothesis on serial order in working memory. *Frontiers in Human Neuroscience*, *8*, 932.

Baguley, T. (2012). *Serious stats: A guide to advanced statistics for the behavioral sciences*. New York, NY: Palgrave Macmillan.

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, The Netherlands: Boom Test Publishers.

Bourassa, A. (2014). *Numerical sequence recognition: Is familiarity or ordinality the primary factor in performance?* (Master's thesis). Carleton University, Ottawa, Ontario, Canada.

Brannon, E. M. (2002). The development of ordinal numerical knowledge in infancy. *Cognition*, *83*, 223–240.

Buckley, P. B., & Gillman, C. B. (1974). Comparisons of digits and dot patterns. *Journal of Experimental Psychology*, *103*, 1131–1136.

Colomé, A., & Noël, M. P. (2012). One first? Acquisition of the cardinal and ordinal uses of numbers in preschoolers. *Journal of Experimental Child Psychology*, *113*, 233–247.

De Vos, T. 2010. *Tempo Test Automatiseren*. Amsterdam, The Netherlands: Boom Test Publishers.

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

Delazer, M., & Butterworth, B. (1997). A dissociation of number meanings. *Cognitive Neuropsychology*, *14*, 613–636.

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.

Fias, W., Lammertyn, J., Caessens, B., & Orban, G. A. (2007). Processing of abstract ordinal knowledge in the horizontal segment of the intraparietal sulcus. *The Journal of Neuroscience*, *27*, 8952–8956.

Franklin, M. S., & Jonides, J. (2009). Order and magnitude share a common representation in parietal cortex. *Journal of Cognitive Neuroscience*, *21*, 2114–2120.

Franklin, M. S., Jonides, J., & Smith, E. E. (2009). Processing of order information for numbers and months. *Memory and Cognition*, *37*, 644–654.

Gebuis, T., & Reynvoet, B. (2012). The interplay between nonsymbolic number and its continuous visual properties. *Journal of Experimental Psychology*, *141*, 642–648.

Gilmore, C., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., Simms, V., & Inglis, M. (2013). Individual differences in inhibitory control, not non-verbal number acuity, correlate with mathematics achievement. *PloS One*, *8*(6), e67374.

Inglis, M., & Gilmore, C. (2014). Indexing the approximate number system. *Acta Psychologica*, *145*, 147–155.

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.

Jou, J. (2003). Multiple number and letter comparison: Directionality and accessibility in numeric and alphabetic memories. *The American Journal of Psychology*, *116*, 543–579.

Knops, A., & Willmes, K. (2014). Numerical ordering and symbolic arithmetic share frontal and parietal circuits in the right hemisphere. *NeuroImage*, *84*, 786–795.

LeFevre, J. A., & Bisanz, J. (1986). A cognitive analysis of number-series problems: Sources of individual differences in performance. *Memory and Cognition*, *14*, 287–298.

LeFevre, J.-A., Kulak, A. G., & Bisanz, J. (1991). Individual differences and developmental change in the associative relations among numbers. *Journal of Experimental Child Psychology*, *52*, 256–274.

Lovelace, E. A., & Snodgrass, R. D. (1971). Decision times for alphabetic order of letter pairs. *Journal of Experimental Psychology*, *88*, 258–264.

Lyons, I. M., & Beilock, S. L. (2009). Beyond quantity: Individual differences in working-memory and the ordinal understanding of numerical symbols. *Cognition*, *113*, 189–204.

Lyons, I. M., & Beilock, S. L. (2011). Numerical ordering ability mediates the relation between number-sense and arithmetic competence. *Cognition*, *121*, 256–261.

Lyons, I. M., & Beilock, S. L. (2013). Ordinality and the nature of symbolic numbers. *The Journal of Neuroscience*, *33*(43), 17052–17061.

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.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgments of numerical inequality. *Nature*, *215*, 1519–1520.

Neter, J., Wasserman, W., & Kutner, M. H. (1989). *Applied linear regression models*. Homewood, IL: Irwin.

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.

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

Rubinsten, O., & Sury, D. (2011). Processing ordinality and quantity: The case of developmental dyscalculia. *PloS One*, *6*(9), e24079.

Rundus, D. (1971). An analysis of rehearsal processes in free recall. *Journal of Experimental Psychology*, *89*, 63–77.

Šidák, Z. K. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, *62*(318), 622–633.

Turconi, E., Campbell, J. I., & Seron, X. (2006). Numerical order and quantity processing in number comparison. *Cognition*, *98*, 273–285.

Turconi, E., & Seron, X. (2002). Dissociation between order and quantity meanings in a patient with Gerstmann syndrome. *Cortex*, *38*, 911–914.

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, The Netherlands: Swets Test Services.

Zorzi, M., Di Bono, M. G., & Fias, W. (2011). Distinct representations of numerical and non-numerical order in the human intraparietal sulcus revealed by multivariate pattern recognition. *NeuroImage*, *56*, 674–680.

APPENDIX 1

Table A1
ORDERING TASK ITEMS

Subcondition	*Single-digit*			*Subcondition*	*Double-digit*		
i1	2	3	4	i1	18	19	20
i1	4	5	6	i1	18	19	20
i1	4	5	6	i1	29	30	31
i1	6	7	8	i1	29	30	31
i1	6	7	8	i2	17	19	21
i2	1	3	5	i2	28	30	32
i2	1	3	5	i2	28	30	32
i2	3	5	7	i2	38	40	42
i2	3	5	7	i2	38	40	42
i2	5	7	9	i3	18	21	24
i3	2	5	8	i3	27	30	33
i3	2	5	8	i3	27	30	33
i3	3	6	9	i3	37	40	43
i3	3	6	9	i3	37	40	43
n1	4	3	2	n1	18	20	19
n1	4	2	3	n1	20	19	18
n1	6	4	5	n1	30	29	31
n1	6	5	4	n1	39	38	40
n1	8	7	6	n1	40	39	38
n2	5	3	1	n2	32	30	28
n2	5	3	7	n2	32	28	30
n2	7	5	9	n2	40	42	38
n2	9	7	5	n2	42	40	38
n3	5	8	2	n3	24	18	21
n3	6	3	9	n3	24	21	18
n3	7	1	4	n3	33	30	27
n3	7	4	1	n3	40	37	43
n3	8	5	2	n3	43	40	37

APPENDIX 2. ADDITIONAL TASK DESCRIPTIONS

Note that task descriptions are taken from Lyons et al. (2014).

Numeral Comparison (*NumComp*). In the NumComp task, children saw two numbers presented horizontally as Arabic numerals, and their task was to decide which number represented the larger quantity. Children saw 64 trials, 32 of which were one digit and 32 of which were two digits. For both sizes, ratios ($R = \text{min/max}$) fell into one of four ranges: $R \leq .5$, $R = .5$, $.5 < R < .7$, $R \geq .7$, with eight trials in each ratio range at each size (one digit vs. two digits). Stimuli remained on the screen until the child responded. Reliability on this task was high: $\alpha = .977$.

Dot Comparison (*DotComp*). In the DotComp 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 NumComp task. Stimuli remained on the screen until the child responded. Note that strong relations between performance on this task and various measures of math ability have been reported previously when allowing for self-paced responses (e.g., Piazza et al., 2010). Owing to geometric constraints, all versions of a dot-comparison task will allow for at least some non-numerical parameter to covary with number, a problem compounded by the fact that participants switch the parameters they rely upon from trial-to-trial (Gebuis & Reynvoet, 2012). Recent work has shown that performance on dot-comparison trials where overall area and average individual dots size are *incongruent* with number is more predictive of math achievement than congruent trials (Gilmore et al., 2013). In the current data 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; note also that density/overall contour was always congruent with number). In this way, our stimuli were biased to find a positive result—that is, a significant relation between DotComp and Arithmetic. Reliability on this task was high: $\alpha = .955$.

Object Matching (*ObjMatch*). In the ObjMatch task, children were shown a sample array of common objects (various animals and pieces of fruit) and two test arrays of objects below the sample array. The children's task was to determine if the left or right test array contained the same number of objects as the sample array. Children saw 45 trials in total. On 15 trials, all objects in all arrays were the same. On 15 trials, each of the three arrays contained different types of objects, but the objects within a given array were all the same. On 15 trials, all arrays contained a mixture of object types. The number of objects in the arrays ranged from 1 to 6, and the absolute numerical distance between the two test arrays was 1 or 2. Stimuli remained on the screen until the child responded. Reliability on this task was high: $\alpha = .956$.

Counting (*Counting*). In the Counting task, children saw between 1 and 9 dots and their task was to count the number of dots on the screen as quickly and accurately as possible. Children saw five trials for each quantity. Trials were scored as correct only if the child's response was exactly correct. Verbal responses were collected by the experimenter in written fashion. Response times were estimated by having the child pressing a button as they gave their verbal response. Reliability on this task was high: $\alpha = .946$.

Numberline Estimation (*NumLine*). In the NumLine task, children were shown a horizontal line marked as 0 on the left end and 100 on the right end. On each trial, they were shown an Arabic numeral (centrally presented above the numberline) in the range 3–96 (the number was presented verbally at the same time through a pair of headphones). Children's task was to click (with a computer mouse) on the numberline where they thought the target number should be placed in terms of the relative quantity it represented. Stimuli remained on the screen until the child responded. Children saw 26 total trials. Reliability on this task was high: $\alpha = .940$.

Dot-Quantity Estimation (*DotEst*). In the DotEst task, children were shown a single array of dots presented too quickly (750 ms) for the dots to be counted individually, which was followed by a visual mask. The mask remained on the screen until the child responded. Children's task was to estimate the number of dots in the array by giving a verbal response. These responses were manually recorded by the experimenter. Children completed a total of 84 trials (12 each for quantities 1–4, 7, 11, and 16). Note that if only trials with target values 4, 7, 11, and 16 were used, results were highly similar. Reliability on this task was good: $\alpha = .824$.

Nonverbal Intelligence (*Ravens*). The Ravens task comprised a battery of colored progressive matrices. This is a normed, untimed, visuo-spatial reasoning test for children (Raven, Court, & Raven, 1995). Children saw a colored pattern and were asked to select the missing piece out of six choices. Children completed 36 items; a child's score was the total number of correctly completed items. For the Dutch version of this task, Van Bon (1986) reported reliabilities of .80 or higher.

Reading Ability (*Reading*). The Reading task was part of the normed Maastricht Dyslexia Differential Diagnosis battery (Blomert & Vaessen, 2009) and comprised three subtasks. Subtasks contained high-frequency words, low-frequency words, or pseudo-words. In each subtask, participants saw a series of up to 5 screens (advanced by the experimenter), each with up to 15 items (75 total items per task). Children's task was to read each item aloud as quickly and accurately as possible. An experimenter manually marked the accuracy of each item. A child's score on a subtask was the total number of correctly read items in 30 s. Scores on the three subtasks were summed to form a child's final Reading score. Reported test–retest reliability for this task is high (.95; Blomert & Vaessen, 2009).