Journal of Experimental Psychology: General

ISSN: 0096-3445

2021, Vol. 150, No. 10, 2100–2116
https://doi.org/10.1037/xge0001022

Rules of Order: Evidence for a Novel Influence on Ordinal Processing of Numbers

Sylvia U. Gattas[1, 3], Stephanie Bugden[2], and Ian M. Lyons[3]
[1] Department of Experimental Psychology, University of Oxford
[2] Department of Psychology, University of Pennsylvania
[3] Department of Psychology, Georgetown University

Research on how people process numerical order carries implications for our theoretical understanding of what a number means and our practical understanding of the foundation upon which more sophisticated mathematics is built. Current thinking posits that ordinal processing of numbers is linked to repeated practice with the integer count list, but the mechanisms underlying this link remain unclear. For instance, in standard ordinal verification paradigms, participants more rapidly and accurately verify that count-list sequences (e.g., 3–4–5) are "in-order" than non-count-list sequences (e.g., 2–4–6), although it remains unclear whether this is due to strong count-list processing or poor non-count-list processing. If the count list primarily *facilitates* ordinal processing of count-list sequences, then forcing participants to classify sequences like 3–4–5 as "not-in-order" should adversely affect ordinal verification performance. We found that it does, but only moderately in single-digit sequences ($d = -.26$), and not at all in the case of double-digit sequences ($d = -.02$). Alternatively, the count list may influence ordinal processing in an exclusionary manner, creating a tendency to view anything that does not match the count-list as not-in-order. If so, then allowing participants to classify ordered (but non-count-list) sequences like 2–4–6 as not-in-order should improve ordinal verification performance. It did, with strong effects for both single-digit ($d = .74$) and double-digit sequences ($d = 1.04$). Furthermore, we demonstrated that the reverse distance effect found in standard ordinal verification paradigms is driven primarily by poor non-count-list processing. Taken together, our results advance our understanding of the mechanisms by which the count list shapes ordinal processing, even in highly numerate adults.

Keywords: numerical cognition, numerical order, reverse distance effect, counting, arithmetic

A deeper understanding of how humans represent basic numerical concepts has implications for a wide range of educational, financial, and health outcomes (Bynner & Parsons, 2005; Crawford & Cribb, 2013; Duncan et al., 2007; Gerardi et al., 2013; Reyna et al., 2009). Moreover, empirical work examining the cognitive basis of foundational numerical representations can inform a long tradition of philosophical inquiry into the nature of what is meant by notions of "number", "quantity", and "order". For instance, a written number—say 6—conveys information about both cardinal value (the quantity of objects in a set) and implies relations with other numbers—for example, 6 is twice 3, a quarter of 24, and the cube-root of 216. Perhaps the most fundamental of these relations is relative order (6 comes one after 5 and one before 7). Various theoretical viewpoints on the fundamental nature of numbers have been advanced over the centuries, with some emphasizing the importance of primary cardinal value (such as Bertrand Russel), and others proposing that ordinality forms the basis of understanding numbers as a coherent system (such as Giuseppe Peano; see Coles & Sinclair, 2018). Modern empirical research on the cognitive and neural foundations of numerical processing has largely focused on cardinality (e.g., Ansari, 2008; Butterworth, 1999; Daro et al., 2011; Nieder & Dehaene, 2009). However, recent years have seen a steady uptick in work focusing on ordinality (Lyons et al., 2016). This latter upwelling of work on ordinality has been driven in part by evidence showing that basic tasks assessing individual differences in judgments of ordinal relations between numbers are particularly strong predictors of more complex forms of mathematical processing. Indeed, ordinality judgments often out-predict tasks focused on cardinal judgements of numerical value (e.g., Goffin & Ansari, 2016; Lyons & Beilock, 2011; Sasanguie et al., 2017). Hence, a deeper understanding of how individuals process ordinal relations among numbers has implications for our theoretical understanding of what it means for number symbols to convey information, as well as for the basic cognitive foundations upon which more sophisticated mathematics are built (Lyons, et al., 2016).

This article was published Online First April 5, 2021.

This work was supported by Departmental Start-Up Funds to Ian M. Lyons (Georgetown University). We thank Elizabeth Necka and Anita Grabowska for their helpful comments in revising previous versions of the article. We would also like to thank Alaina Chen, Ava Cobarrubias, Angela Foley, Subul Malik, and Caitlin Murray, for help with data collection.

Correspondence concerning this article should be addressed to Ian M. Lyons, Department of Psychology, Georgetown University, White-Gravenor Hall, 302H, Washington, DC 20057, United States. Email: ian.lyons@georgetown.edu

The Influence of the Count List on Ordinal Processing

Current thinking about the development of ordinal processing of numbers is that it is linked strongly to children's familiarity with and repeated practice rehearsing the integer count list ("one, two, three . . ."; for a review, see Lyons et al., 2016). One of the first numerical skills children acquire is the ability to recite the count list, which often precedes their understanding of the numerical meanings of the words they are saying (Wynn, 1990, 1992). By the time they enter grade school, children have substantial experience reciting the count list—certainly for numbers up to 10 (Wynn, 1992). This repeated, sustained experience is thought to create specific memory traces against which one can easily verify that sequences such as 1–2–3 and 4–5–6 are in numerical order (e.g., Bourassa, 2014; Franklin et al., 2009; Turconi et al., 2006). Moreover, the high frequency with which this list is recited is thought to strengthen the ordinal associations between adjacent numbers in the count list, thus facilitating processing of count-list-adjacent numbers (LeFevre et al., 1991; Lyons & Beilock, 2013). We refer to this hypothesis as the *facilitatory hypothesis*: greater experience with the count list strengthens ordinal associations between numerals that are adjacent in the count list. Consistent with this hypothesis, when judging the ordinality of numerical sequences, both adults and children as young as first grade show strongest performance (lowest response times and error rates) when verifying that count-list sequences (e.g., 1–2–3) are "in order" resulting in a count-list advantage (Franklin et al., 2009; Goffin & Ansari, 2016; LeFevre & Bisanz, 1986; Lyons & Ansari, 2015; Lyons & Beilock, 2009).

An alternative, albeit not mutually exclusive, hypothesis is that participants—even highly numerate, adult participants—struggle to overcome a basic, exclusionary heuristic that the only ordered sequences are count-list sequences. We refer to this hypothesis as the *exclusionary hypothesis*: greater experience with the count list constrains one's intuitions of what constitutes numerical order to just those sequences that match the integer count list to the exclusion of all other sequences that fail to match the count-list. In other words, one's default intuition is to consider non-count-list sequences like 2–4–6 to be "not-in-order." Of course, as they are defined in the world of mathematics, numerical sequences do not have to match the count list to be in order. For example, ($\sqrt{2}$, e, π), odd as it may appear, is nevertheless a perfectly valid, ordered sequence; and indeed, this broader notion of ordinality lies at the heart of some of the most famous proofs in mathematics (for instance, Cantor's work on countable and uncountable infinities). However, the issue here is not whether such sequences should be considered valid ordered sequences, but rather the notion that participants—even highly educated, numerically literate adults—may struggle to overcome the strong link between the count list and what it means for numbers to be ordered. This strong association may in turn create an exclusionary heuristic that must be overcome to extend one's notion of ordinality to include non-count-list sequences that are in fact valid ordinal exemplars.

In the current study, we probe the idea that the influence of the count list goes beyond actual count-list sequences, and forms the broader conceptual foundation of what it means for numbers to be in order in the first place. This idea in turn may have implications for how humans think about numbers in general, insofar as ordinality plays a foundational role in numerical and mathematical thinking. To our knowledge, however, this notion remains untested. To do so here, we modified an existing ordinal verification paradigm commonly used in the literature by manipulating the rules determining which sequences were to be classified as "in-order" versus "not-in-order."' We reasoned that when a new rule matched participants' underlying heuristic or intuition about what constitutes an ordered sequence, performance should improve; when a rule created a mismatch, performance should suffer. Moreover, by comparing the relative magnitude of facilitation and exclusionary effects, we could thus also determine whether the count list influences ordinal processing of numbers primarily by privileging count-list sequences, by disadvantaging non-count-list sequences, or both.

Explaining the Reverse Distance Effect

One implication of an exclusionary heuristic that count-list sequences are the only ordered sequences is an alternative explanation for a now classic behavioral and neural effect seen in ordinality judgment tasks. When judging the ordinality of (symbolic) numbers, participants tend to be slower and less accurate when numbers are numerically farther apart: 2–4–6 is harder than 3–4–5 (e.g., Franklin & Jonides, 2009; Franklin et al., 2009; Goffin & Ansari, 2016; Lyons & Ansari, 2015; Lyons & Beilock, 2013; Turconi et al., 2006). This phenomenon is referred to as a *reverse distance-effect* (RDE) because it reverses the pattern typically seen for cardinality judgments (determine which quantity is numerically greater) in which numerically closer pairs (7 vs 8) are harder to judge than numerically farther pairs (6 vs 9), (Buckley & Gillman, 1974; Moyer & Landauer, 1967).

A leading explanation for the RDE, consistent with the facilitatory hypothesis, rests on the notion that repeated practice with and exposure to the integer count list instills strong memory traces for these sequences (e.g., 3–4–5), and so participants are particularly adept at verifying the ordinality of sequences that match these memory traces. The root cause of the RDE is thus privileged, efficient processing of these sequences (e.g., Bourassa, 2014; Franklin et al., 2009; Turconi et al., 2006). An alternative view, per the exclusionary hypothesis, is that participants must overcome an exclusionary heuristic that the only ordered sequences are count-list sequences. Being forced to inhibit this default response to classify sequences such as 2–4–6 as in-order leads to poor performance on these trials.

Data from existing paradigms cannot distinguish between these hypotheses, however, because the RDE is typically calculated by computing the difference in performance on count-list trials (e.g., 3–4–5) versus ordered, non-count-list trials (e.g., 2–4–6), making it fundamentally ambiguous whether the difference is driven by "good" performance in one Condition or "poor" performance in the other (or both). In other, similar situations, such as computing facilitation and interference effects in the Stroop effect, one often compares performance on congruent and incongruent trials, respectively, to a neutral condition (e.g., one where all words are in white font on a black screen). In ordinality verification paradigms no such neutral condition is immediately forthcoming. For instance, nonordered trials (e.g., 3–5–4, 2–6–4) are not a truly neutral comparison condition as they differ in terms of both ordinality and pairwise-distance.

The novel paradigm developed here allowed us to overcome this limitation by comparing performance on the same sequences under different rule conditions. The different rule conditions were designed to independently test the hypothesized influences of a count-list advantage (as predicted by the facilitatory hypothesis) and a non-count-list disadvantage (predicted by the exclusionary hypothesis). If one were to eliminate the count-list advantage, this should eliminate the RDE. Similarly, if one were to eliminate the non-count-list disadvantage, this should also eliminate the RDE. Finally, because this approach tests each hypothetical source of the RDE separately, it also allows for the possibility that both a count-list advantage and a non-count-list disadvantage contribute to the RDE (i.e., the two hypotheses are not mutually exclusive).

Current Study

Current thinking posits that ordinal processing of numbers in children and adults is linked strongly to repeated practice with the integer count list. Here we suggest that this influence goes beyond count-list sequences themselves, and forms the broader conceptual foundation of people's understanding of what it means for numbers to be ordered in the first place. Specifically, we tested two different—albeit not mutually exclusive—hypotheses on how the count list shapes ordinal thinking (namely, the facilitatory and exclusionary hypotheses, as described above). To do so, we modified a standard ordinality verification task (in which participants judge whether sequences of three numbers are in numerical order), by manipulating the rules by which participants had to classify different types of number sequences as in-order versus not-in-order. As a secondary consequence, this approach also allows us to probe competing explanations for the source of the RDE. Furthermore, we probe whether the count list influences highly familiar, overlearned sequences (e.g., 1–2–3) in a different manner than less familiar sequences (e.g., 29–30–31).

Testing Different Influences of the Count List on Ordinal Processing

By definition, numerical order does not require that sequences match the count list to be considered "in order" (e.g., 3–4–5, 2–4–6 and $\sqrt{2}$–e–π are all in-order; Figure 1, gold [light gray] column), and that is how the numerical ordering task is typically administered (e.g., Franklin et al., 2009; Lyons & Beilock, 2009). However, this approach does not allow one to distinguish between the facilitatory and exclusionary hypotheses outlined above. To overcome this limitation, we modified the rules for what sequences should be classified as in-order versus not in-order.

The facilitatory hypothesis states extensive experience with the count list has created strong memory traces for count-list sequences (e.g., 3–4–5), allowing one to readily verify such sequences as being in-order. To test this idea, we modified the rules of the standard ordinality verification task such that count-list sequences must now be classified as "not-in-order." Per the facilitatory hypothesis, participants should have a strong default inclination to view count-list sequences like 3–4–5 as in-order, which they must now inhibit, and so performance should be adversely affected. In the incongruent-rule condition (Figure 1, red [white] column), trials that do match the count list such as 3–4–5 were to be classified as not-in-order (note that non-count-list trials like 2–4–6 were classified as in-order in this condition, which is the same as in the standard-rule condition). Hence, according to the facilitatory hypothesis, relative to the standard condition, we expected performance to be worse (higher response times and errors) for count-list sequences in the incongruent condition.

According to the exclusionary hypothesis, participants' default (or initial) intuition is to see non-count list sequences like 2–4–6 as not-in-order, even if they in fact are. Adults must overcome an initial heuristic that the only ordered sequences are count-list sequences to correctly complete the standard version of the task. With this in mind, modifying the rules of the task so that participants are permitted to classify sequences like 2–4–6 as not-in-order (Figure 1, blue [dark gray] column), should improve performance relative to the standard version of the task. This is because sequences like 2–4–6 should now match participants' default intuition that these sequences are not in order. Furthermore, participants no longer need to distinguish between different types of sequences that they see as nonordered (2–4–6 and 4–6–2

Figure 1

Overview of the Various Stimulus Types and the Three Different Ordinality Rule Conditions

			Ordinality Rule		
Ordinality	**Distance**	**Examples**	**Standard**	**Incongruent**	**Congruent**
Count-List	1	(3–4–5) (28–29–30)	✓	✗	✓
Non-Count-List	2, 3	(2–4–6) (26–28–30)	✓	✓	✗
Mixed	1	(4–5–3) (29–30–28)	✗	✗	✗
Mixed	2, 3	(4–6–2) (28–30–26)	✗	✗	✗

Note. Example stimuli are provided in parentheses. A ✓ indicates that stimulus-type was to be counted as "in-order" for that ordinality-rule condition; an ✗ indicates that stimulus-type was to be counted as "not-in-order" for that condition. See the online article for the color version of this figure.

can now both be classified as not-in-order). Hence, for the congruent-rule condition (Figure 1, blue [dark gray] column), we predicted an improvement in performance on non-count-list (but still ordered) sequences, relative to the standard condition (Figure 1, gold [light gray] column).

Finally, this approach allows for the possibility that both the facilitatory and exclusionary hypotheses are correct. It may be that the influence of the count list on ordinal processing is twofold: adults are predisposed to view count-list sequences as in-order and tend to view all other sequences as not-in-order even when they actually are. Including both the congruent- and incongruent-rule conditions allowed us to test for both influences independently. Furthermore, we can also compare the magnitudes of the two hypothetical influences to determine which exerts the strongest effect on adult participants' intuitions about numerical order.

Testing Explanations of the Reverse Distance Effect

The reverse distance effect (RDE) refers to better performance (lower response times and errors) on count-list trials (e.g., 3–4–5) relative to ordered, non-count-list trials (e.g., 2–4–6) in a standard ordinality verification paradigm. However, it is unclear whether the RDE is driven by relatively good performance on the count-list trials, poor performance on the non-count-list trials, or both. On the one hand, according to the facilitatory hypothesis, the RDE may be driven by good performance on sequences that do match the count-list (e.g., 3–4–5). In that case, we would expect the RDE to be reduced or even eliminated in the incongruent-rule condition because participants must judge these count-list trials to be not-in-order, and the added difficulty of having to inhibit the tendency to see these items as being in-order should reduce performance. This reduction in performance should thus reduce or eliminate the gap between count-list and non-count-list trials, thereby reducing or eliminating the RDE. On the other hand, according to the exclusionary hypothesis, the RDE may be driven by poor performance on sequences that do not match the count-list (e.g., 2–4–6). In that case, we would expect the RDE to be reduced or even eliminated in the congruent-rule condition because participants are allowed to judge these non-count-list trials as not-in-order, and the increased ease of yielding to the tendency to see these items as being not-in-order should improve performance. This improvement in performance should thus reduce or eliminate the gap between count-list and non-count-list trials, hence reducing or eliminating the RDE. Finally, it is important to note that the two mechanisms proposed here are not mutually exclusive; both effects may be present, which would imply that RDEs arise from both sources.

Generalizability and Mechanism

A major part of the story here is that rehearsal of the count list leads to increased familiarity with numerical sequences that match the count list (LeFevre et al., 1991). It is unclear whether this familiarity is in terms of representing specific number sequences (1–2–3, 4–5–6, etc.) in memory (Bourassa, 2014; Sella et al., 2020); or with a more general process, such as reciting the count list itself. Because participants are expected to have less overall exposure to (Dehaene & Mehler, 1992) and less experience reciting (Wynn, 1990) the count list at higher (double-digit) numbers, we can probe these mechanisms by including both single- and double-digit sequences as stimuli. Specifically, we can assess whether the presence of a count-list advantage and/or a non-count-list disadvantage obtains only in highly familiar, overlearned sequences (single-digits) or generalizes to sequences where ordinal processing is less able to rely on direct memory retrieval (double-digits). If an effect is specific to overlearned single-digit sequences, this would suggest it is driven more by retrieval of specific ordinal representations; if an effect generalizes to less familiar double-digit sequences, this would suggest a more general process is required to explain the effect, such as verbal recitation or pairwise comparison. In this way, we can test the specific mechanisms —representation-based or process-based—by which the count-list exerts different influences on how we process the ordinality of numbers.

Method

Participants

Data were collected from 61 Georgetown University students and members of the Georgetown community. One participant did not comply with instructions and so was omitted from analysis. We thus proceeded with a total N of 60 participants (M_{age} = 21.8 years, SD = 4.7; 37 female).

Procedure

All tasks and procedures were approved for use with human subjects by the Georgetown University Institutional Review Board (IRB). The total study took approximately 1 hour to complete; and participants were compensated either $10 or one research credit for their time. The ordinality judgment task that is the main focus here was presented as part of a larger experiment, which included a mental arithmetic and an antisaccade task. As these latter two tasks are part of a different study and are not relevant to the theoretical questions we are addressing here, they are omitted from analysis in the main text. Due to reviewer interest, correlations between each trial type in the ordinal verification task and performance on the mental arithmetic task are provided in Appendix D. Importantly, all tasks were presented in randomized order to prevent context or sequence effects from biasing the results discussed here.

Stimuli were presented on a 24 Dell flat-screen monitor with a native 1920 × 1080 resolution, a 60 Hz refresh rate, situated at approximately 18 inches from the seated participant.

Numerical Ordinality Judgment Task

Stimulus Details. The overall task design was modeled after the ordering task in Lyons and Beilock (2011). In this task, participants judged whether three numbers presented horizontally should be considered in-order or not-in-order (from left to right). Participants pressed one of two keys (*C* or *M* on a standard U.S. English keyboard) to indicate their judgment. The meaning of each button (in-order or not-in-order) was randomized across participants; once assigned a given configuration, it was kept consistent across the study for that participant. Stimuli were presented in white font on a black background in 27-point fixed-width (Consolas) font, with three spaces between each number. The three numbers together subtended to about 7.1° of visual angle for single-digit stimuli, and about 8.7° for double-digits.

For each trial, the three numbers were presented for 3 seconds or until the participant responded. If the 3 seconds elapsed, the program moved on automatically. Participants were informed in the instructions section that these trials would be counted as incorrect and thus encouraged to respond within the 3-second time frame. Trials with no response were omitted from analysis (.8% of all trials). The intertrial interval (fixation) randomly varied between 800 and 1,200 ms. In addition, as overall mean response times were 1,076 ms (SD = 437), we semiarbitrarily deemed response times less than 200 ms to be implausible reflections of actual on-task performance, and so omitted these trials (.1% of all trials). In sum, less than 1% of all trials were omitted from analysis.

Trials were divided equally into single- and double-digits (56 trials of each per condition), which were randomly intermixed. On single-digit trials, all three numbers were in the range 1–9. On double-digit trials, all three numbers were two-digits in the range: 17–52. In keeping with prior research, the three double-digit numbers always crossed a decade (e.g., 28–30–32, 48–49–50) to prevent participants from simply ignoring the digit in the decade position (Franklin & Jonides, 2009; Franklin et al., 2009; Lyons & Ansari, 2015).

Trials were divided into three different ordinality conditions: count-list, non-count-list and mixed. See Appendix A (Table A1) for a complete list of trials.

Count-list trials comprised sequences of three numbers that matched the count list (e.g., 3–4–5, 28–29–30). More precisely, the numerical distance between the left and the middle and the middle and the right numbers was held constant at 1. Count-list trials were always presented in their count-list permutations (smallest-median-largest), from left to right.

Non-count-list trials were trials that, by most mathematical definitions were in-order, in that all three numbers increased in magnitude from left to right. However, sequences were not a direct match with the count list in that the sequences skipped either one or two numbers (e.g., 2–4–6, 27–30–33). More precisely, the numerical distance between the minimum and median, and between the median and maximum numbers in a given sequence was always held constant, and the numerical distance between the left and middle and middle and right numbers was either 2 or 3.

Mixed trials were also included as these are a customary aspect of the standard configuration of the ordinality verification task and provide a necessary foil to maintain the validity of the task. For instance, because one responds in-order to both count-list and non-count-list trials in the standard-rule condition, it is necessary to have trials where the correct response is not-in-order so that all trials do not have the same correct response. To this end, mixed trials comprised the same sequences from the count-list and non-count-list trials, but with their order permuted such that they no longer were increasing from left to right (e.g., 4–5–3, 30–33–27). In tables and figures, Mixed_1 refers to permutations of count-list trials with a numerical distance of 1, and $\text{Mixed}_{2,3}$ refers to permutations of non-count-list trials with numerical distances of 2 or 3. Mixed trials also serve the useful role of preventing participants from focusing on any pair of numbers, and instead encourage them to process the overall ordinality of the entire sequence (Lyons & Beilock, 2009). Thus, mixed permutations were chosen so that no single pair of positions was guaranteed to be either increasing or decreasing.

Ordinality Rule Manipulation. Participants completed three different conditions for this task. The main manipulation distinguishing each condition concerned what types of trials were to be classified as in-order versus not-in-order (specifics of each condition follow below). See Appendix B for complete instructions for each condition. Condition order was randomized across participants. There were 112 trials of each condition, divided into two blocks, with a brief rest provided between blocks. Prior to starting the 112 main trials, participants first received instructions and examples, followed by 24 practice trials. Feedback (correct or incorrect) was given for the practice but not the main trials. The probability of an in-order or not-in-order (C or M key) response was held equal (50/50) across all conditions. Finally, note that accuracy was always defined as adherence to the rules for that condition.

Standard Ordinality Rule. In the Standard-Rule condition, the rule governing what counted as in-order was based purely on numerical value. Thus, any sequence that was in fact numerically increasing (left-to-right) was to be considered in-order regardless of the numerical distance between numbers in the sequence. Thus, both count-list (e.g., 3–4–5) and non-count-list trials (e.g., 2–4–6) were to be classified as in-order. This condition mimics that used widely throughout the literature, which is why we refer to it here as the standard ordinality rule condition. To maintain equal probability of yes/no responses, there were 56 trials that counted as in-order: 28 count-list and 28 non-count-list trials. There 56 trials that counted as not-in-order were all drawn from the mixed set of trials, with an equal representation of permutations of the count-list trials (Mixed_1) and permutations of non-count-list trials ($\text{Mixed}_{2,3}$). All of the above subconditions were divided equally between single- and double-digit trials. All trial types comprised 112 total trials for this condition, randomly intermixed and presented over two blocks.

Congruent Rule. In the congruent-rule condition, only count-list trials (e.g., 3–4–5) were to be considered in-order. All other trials, including non-count-list trials (e.g., 2–4–6) were to be considered not-in-order. To maintain equal probability of yes/no responses, there were 56 trials that were considered in-order (all drawn from the count-list set), 56 trials that were considered not-in-order (28 drawn from the non-count-list set and 28 drawn from the mixed set—with equal representation of count-list and non-count-list permutations among mixed trials). All of the above subconditions were divided equally between single- and double- digit trials. All trial types comprised 112 total trials for this condition, randomly intermixed and presented over two blocks.

Incongruent Rule. In the incongruent-rule condition, only non-count-list trials (e.g., 2–4–6) were to be considered in-order. All other trials, including count-list trials (e.g., 3–4–5), were to be considered not-in-order. To maintain equal probability of yes/no responses, there were 56 trials that counted as in-order (all drawn from the non-count-list set), 56 trials that counted as not-in-order (28 drawn from the count-list set and 28 drawn from the mixed set —with equal representation of count-list and non-count-list permutations among mixed trials). All of the above subconditions were divided equally between single- and double-digit trials. All trial types comprised 112 total trials for this condition, randomly intermixed and presented over two blocks.

Methodological Considerations

Note that the probability of a correct in-order and a not-in-order was the same (50/50) across all conditions, meaning that differences between conditions cannot be explained by differences in the prevalence of "yes" vs "no" distributions. Similarly, because the mapping between a given hand and a given classification (left = in-order, right = not-in-order; or vice-versa) was counterbalanced across participants, differences across conditions cannot be explained by handedness. However, achieving a 50/50 distribution of in-order and not-in-order responses required an asymmetrical number of close- and far distance trials in the congruent-rule and incongruent-rule conditions. Importantly therefore, beyond testing key hypotheses of interest, the congruent and incongruent conditions also serve as important controls for one another. For instance, one might argue that differences between the standard and congruent or incongruent conditions could be driven by an asymmetry between response-demands and stimulus categories. However, this asymmetry was present in both the congruent and incongruent conditions (only one category is associated with the in-order option, and three with the not-in-order option). Thus, if the two conditions alter performance in different ways, it is difficult to see how an asymmetry simply in how many categories were assigned to each response option. Instead, results would be better explained by which categories are assigned to which response option, as is the case for the facilitatory and exclusionary hypotheses.

Similarly, two of our key comparisons are between count-list trials in the standard and incongruent conditions (red [dark gray] bars, Figure 2), and between non-count-list trials in the standard and congruent conditions (blue [light gray] bars, Figure 2). One might object that, in each case, one is comparing a Yes response (in-order) with a No response (not-in-order). Importantly, our key predictions, from the facilitatory and exclusionary hypotheses, are that these contrasts should yield effects *in* opposite directions: a negative difference for $\text{Count-List}_{\text{Standard}} - \text{Count-List}_{\text{Incongruent}}$, and a positive difference for $\text{Non-Count-List}_{\text{Standard}} - \text{Non-Count-List}_{\text{Congruent}}$. Because both contrasts involve comparing between Yes/No responses in the same manner, an explanation based solely on response-Type could not explain the predicted results.

Measurement

Performance on this task was measured in terms of response times (ms) and error rates (proportion wrong), which were combined into a single composite measure. First, this approach halved the number of statistical tests that need to be performed, thus reducing the likelihood of Type I errors without introducing the need to correct for multiple comparisons (thus protecting somewhat against Type II errors as well). Second, this approach implicitly controlled for speed–accuracy trade-offs. Response times and error rates were combined here using the combined performance measure (*CP*), introduced by Lyons et al. (2014): $CP = RT(1 + 2ER)$, where RT = response time and ER = error rate for a given participant, on a given condition/trial type. As chance performance on this task corresponds to an ER of .5, this formula linearly reweights RT as a function of ER, running from $CP = RT$ for perfect accuracy, to $CP = 2*RT$ for chance accuracy. Note that this formulation makes no assumptions about underlying distributions (as one does for instance in the case of combining z-scores), and it combines the two measures in a linear manner (in contrast to the nonlinear weighting implied by metrics such as inverse efficiency, which introduces assumptions about which side of the error distribution should be weighted as more consequential). In the interest of full transparency, however, response times and error rates (broken down into relevant condition means) can be found in Appendix C.

Bayesian Statistics

We conducted Bayesian *t*-tests using the default prior of .707 (Faulkenberry et al., 2020; Morey & Rouder, 2015) to quantify the probability of support in favor or against our hypotheses. The Bayes Factor (BF_{10}) is reported for each test, which is the ratio of the likelihood of data fitting the alternative hypothesis relative to the null hypothesis (BF_{01} is the inverse and provides support for the null relative to the alternative hypothesis). For example, a BF_{10} of 1 would provide equal support for both the null and alternative hypothesis, whereas a BF_{10} of 3 suggests that the data are 3 times more likely in favor of the alternative hypothesis relative to the null. Bayes Factors greater than 10 are considered "strong evidence," and Bayes Factors greater than 30 are considered "very strong evidence" (Jeffreys, 1961).

Data Availability

Data are publicly available via the Open Science Framework: https://osf.io/z4ngj/.

Results

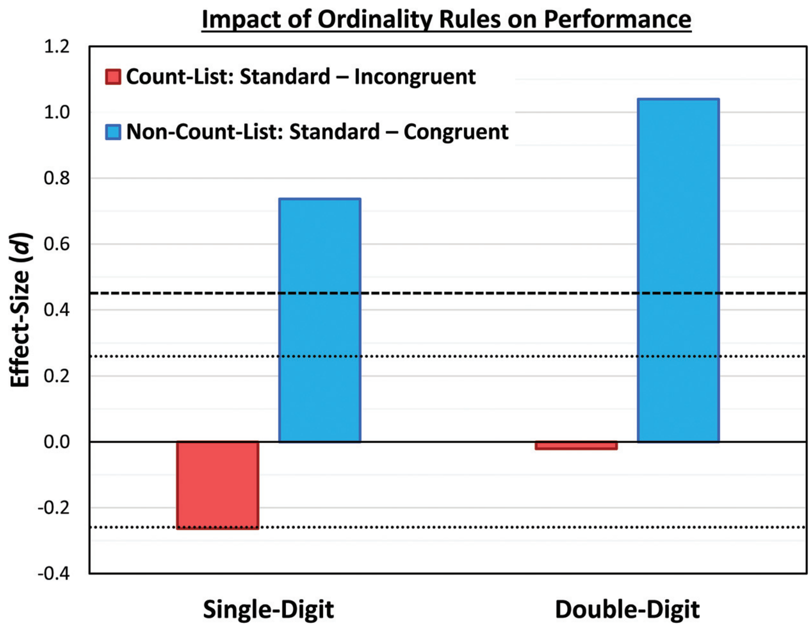
Analysis 1: The Impact of Different Ordinality Rules on Count-List and Non-Count-List Performance

Our primary question was whether altering the rules by which count-list and non-count-list trials impacted the ease or difficulty with which participants classified these trials as in-order or not-in-order. In particular, we tested two (nonmutually exclusive) hypotheses regarding how the count list influences ordinal processing of numbers. Furthermore, we tested for these influences separately in single- and double-digit numbers to assess the specific mechanisms —representation-based or process-based—by which the count list exerts different influences on how individuals process the ordinality of numbers. If an effect is specific to overlearned single-digit sequences, this would suggest it is driven more by retrieval of specific ordinal representations; if an effect generalizes to less familiar double-digit sequences, this would suggest the influence of the count list in this case operates via the more general procedure or process of rehearsing the count list.

Facilitatory Hypothesis

We first tested the facilitatory hypothesis of whether the count list influences ordinality by facilitating recognition of count-list sequences as in-order. In the incongruent-rule condition, participants were told to classify count-list sequences (e.g., 3–4–5) as though they are not-in-order. From this perspective, performance on count-list trials should be worse in the incongruent condition relative to the standard-rule condition because participants were forced to classify these sequences in a manner that contradicted their (hypothesized) intuition about what it means for numbers

Figure 2
Impact of the Different Ordinality Rules (Incongruent and Congruent) on Performance, Relative to Performance in the Standard-Rule Condition



Note. Recall that the incongruent-rule condition altered only the rule for count-list trials, and the congruent-rule condition altered only the rule for non-count-list trials; hence, only the relevant trial type is shown here in each case (for all trials, see Table 1). A positive value indicates better performance on that trial type than in the standard condition; a negative value indicates worse performance on that trial type than in the standard condition. The graph shows results from paired *t*-tests, converted to effect-sizes (*d*s), where the null (no difference) corresponds to $d = 0$. Dotted lines indicate an effect-size corresponding to $p = .05$; the thicker dashed line indicates an effect-size corresponding to $p = .001$. See the online article for the color version of this figure.

to be in-order. Specifically, we computed the difference as (standard – incongruent), so a significant negative effect would indicate worse performance in the latter condition relative to the former. From Figure 2 (red [dark gray] bars), we see a relatively small but significant effect for single-digits [$t(59) = -2.04$, $p = .046$, $d = -.26$, $BF_{10} = .97$; left red [dark gray] bar < 0], but no significant effect for double-digits [$t(59) = -.16$, $p = .872$, $d = -.02$, $BF_{10} = .14$; right red (dark gray) bar not different from 0]. The effect for single-digits was marginally greater than that for double-digits [$t(59) = 1.97$, $p = .054$, $d = .25$, $BF_{10} = .85$; the left red (dark gray) bar was marginally more negative than the right red (dark gray) bar in Figure 2].

Exclusionary Hypothesis

We next tested the exclusionary hypothesis of how the count list influences ordinal processing. Recall that in the congruent-rule condition, participants were told to classify non-count-list sequences (e.g., 2–4–6) as though they are not-in-order. From this perspective, we should see performance on non-count-list trials improve in the congruent relative to the standard condition because participants were, in theory, allowed to classify these sequences in a manner in keeping with their (hypothesized) intuition about what it means for numbers to be not-in-order. Specifically, we computed the difference as (standard − congruent), so a significant positive effect would indicate better performance in the latter condition relative to the former. From Figure 2 (blue [light gray] bars), we can see relatively large effects in the predicted direction for both single-digit [$t(59) = 5.71$, $p = 4.0E - 07$, $d = .74$, $BF_{10} = 3.8E + 04$; left blue (light gray) bar > 0] and double-digits [$t(59) = 8.05$, $p = 4.4E\text{-}11$, $d = 1.04$, $BF_{10} = 2.2E + 08$; right blue (light gray) bar > 0]. In fact, the effect was significantly greater for double- relative to single-digit sequences [$t(59) = 4.76$, $p = 1.3E - 05$, $d = .61$, $BF_{10} = 1.4E + 03$; the right blue (light gray) bar was significantly higher than the left blue (light gray) bar in Figure 2].

In sum, we found relatively weak and inconsistent evidence to support the facilitatory hypothesis—that participants are predisposed to see count-list sequences as in-order. This effect obtained only in single-digit sequences, indicating it operates primarily via matching specific count-list sequences with representations of those sequences stored in memory. By contrast, we found strong and consistent evidence in support of the exclusionary hypothesis—that participants' default intuition is to see non-count-list sequences as not-in-order. This effect extended to double-digit sequences (and in fact was slightly stronger), suggesting the exclusionary influence of the count-list on ordinal processing operates via a more general process, such as verbal recitation or pairwise comparison.

Table 1
All Subcondition Means

Trial type	Single-digit Standard	Single-digit Incongruent	Single-digit Congruent	Double-digit Standard	Double-digit Incongruent	Double-digit Congruent
Count-list	**1,025 (39)**	**1,102 (42)**	**911 (37)**	**1,229 (45)**	**1,235 (45)**	**1,078 (37)**
Non-count-list	**1,145 (48)**	**1,134 (40)**	**889 (36)**	**1,474 (61)**	**1,412 (46)**	**1,015 (41)**
Mixed (1)	1,284 (48)	1,259 (60)	1,115 (47)	1,506 (46)	1,504 (65)	1,324 (47)
Mixed (2, 3)	1,166 (41)	1,369 (57)	805 (25)	1,505 (48)	1,726 (66)	940 (33)

Note. Table 1 gives all subcondition means and standard errors (the latter in parentheses). A higher value indicates poorer performance. The two theoretically central subconditions in Analyses 1 and 2 (Count-List and Non-Count-List) are shown in bold. The specific trial sub-types isolated in Analysis 1 are also underlined. See Analysis 3 for treatment of Mixed trials.

Comparing Count-List and Non-Count-List Effects

From the previous section (and from Figure 2), it appears that the tendency to see non-count-list sequences (e.g., 2–4–6) as not-in-order may be stronger than the tendency to see count-list sequences (e.g., 3–4–5) as in-order—that is, the blue (light gray) bars in Figure 2 appear to be further from 0 than are the red (dark gray) bars. To test this idea more formally, it was important to account for the fact that the two effects were in opposite directions (one with a positive and the other with a negative arithmetic sign). To preserve the correlational structure between conditions (which is crucial for properly estimating effect sizes in within-subjects designs), one must compute the direct (nonreflected), row-wise subtraction between variables. However, because we know the mean differences are positive in one case and negative in the other, a comparison against 0 will lead to an inflated estimate. Hence, we instead compared difference scores against twice the absolute mean of the smaller effect. Taking single-digits as an example, for non-count-list trials, the mean difference between congruent-rule and standard-rule conditions was 1145.0 – 889.3 = 255.3; for count-list trials, the mean difference between incongruent-rule and standard conditions was 1025.1 − 1102.2 = −77.1 (values are taken from Table 1). This computation results in a vector of difference scores with a mean value of 255.3 − −77.1 = 332.1, which, we argue, is biased because of the opposing signs of the mean effects (255.3 and −77.1). To account for this bias, rather than compare these difference scores against 0, we compare them against 255.3 – |−77.1| = 2*|−77.1| = 154.2. Note that this is both the more conservative test, and, we argue, provides the more accurate estimate of the true difference between the magnitudes of each effect, by taking into account both the correlational structure and their expected opposing signs.

Using this approach, results demonstrated that the tendency to see non-count-list sequences (e.g., 2–4–6) as not-in-order was indeed stronger than the tendency to see count-list sequences (e.g., 3–4–5) as in-order. This was true for both single-digits [$t(59)$ = 4.43, p = 4.2E − 05, d = .57, BF_{10} = 489] and for double-digits [$t(59)$ = 6.65, p = 1.1E − 08, d = .86, BF_{10} = 1.2E + 06].

Analysis 2: Explaining the Reverse Distance Effect (RDE)

Here we assessed whether the RDE is driven by especially poor performance on ordered sequences that do not match the count list (e.g., 2–4–6), by especially good performance on ordered sequences that match the count list (e.g., 3–4–5), or both. In the first case, consistent with the facilitatory hypothesis, we would expect the RDE to be eliminated in the incongruent-rule condition due to worse performance on count-list trials. In the second case, consistent with the exclusionary hypothesis, we would expect the RDE to be eliminated in the congruent-rule condition due to improvement in performance on non-count-list trials. In the third case, we would expect the RDE to be eliminated in both conditions (support for both hypotheses).

Distance effects were calculated by subtracting non-count-list (i.e., "far" numerical distance) from count-list (i.e., "close" numerical distance) performance. Hence, in Figure 3, a positive value indicates a canonical distance effect and a negative value indicates a RDE. Note that the RDE is typically seen only in ordered trials (i.e., not in mixed trials), so as with Analysis 1, here we focus on count-list and non-count-list trials. From the gold (light gray) bars in Figure 3, we see medium to large RDE effects in the standard-rule condition for both single-digits [$t(59)$ = −4.13, p = 1.1E − 04, d = −.53, BF_{10} = 192; left gold (light gray) bar significantly < 0] and for double-digits [$t(59)$ = −6.27, p = 4.5E − 08, d = −.81, BF_{10} = 2.9E + 05; right gold (light gray) bar significantly < 0], which replicates previous work (Franklin & Jonides, 2009; Franklin et al., 2009; Goffin & Ansari, 2016; Lyons & Ansari, 2015; Lyons & Beilock, 2013; Turconi et al., 2006).

Facilitatory Hypothesis

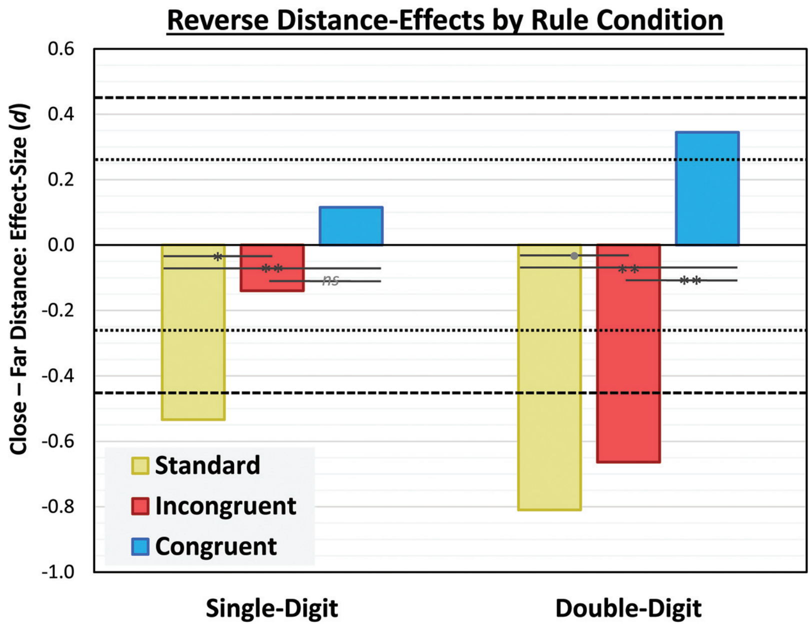
In the incongruent-rule condition, the RDE was also eliminated (i.e., was nonsignificant) for single-digits [$t(59)$ = −1.08, p = .284, d = −.14, BF_{10} = .25; left red (light gray) bar in Figure 3 was not different from 0], and the magnitude of the RDE was significantly reduced in the incongruent relative to the standard condition [$t(59)$ = 2.25, p = .028, d = .29, BF_{10} = 1.45; left red (dark gray) bar vs left gold (light gray) bar in Figure 3]. For double-digits, the RDE remained highly significant [$t(59)$ = −5.14, p = 3.3E − 06, d = −.66, BF_{10} = 5.2E + 03; right red (dark gray) bar in Figure 3 remained significantly < 0], and the magnitude of the RDE was only marginally reduced in the incongruent relative to the standard condition [$t(59)$ = 1.75, p = .085, d = .23, BF_{10} = .60; right red (dark gray) bar vs right gold (light gray) bar in Figure 3].

Exclusionary Hypothesis

In the congruent-rule condition, the RDE was eliminated (i.e., was nonsignificant) for single-digits [$t(59)$ = .89, p = .375, d = .12, BF_{10} = .21; left blue (black) bar in Figure 3 was not different from 0], and the magnitude of the RDE was significantly reduced in the congruent relative to the standard condition [$t(59)$ = 3.54, p =

Figure 3

Impact of the Different Rule Conditions on Reverse Distance Effects (RDEs)



Note. RDEs were computed as count-list − non-count-list (i.e., "close" minus "far" numerical distance), so that a negative value corresponds to a reverse distance effect. The graph shows results from paired *t*-tests, converted to effect-sizes (*d*s). Effects are plotted as effect-sizes, where the null (no difference) corresponds to $d = 0$. Dotted lines indicate an effect size corresponding to $p = .05$; thicker dashed lines indicate an effect-size corresponding to $p = .001$. Gray lines compare distance effects across conditions: * $p < .05$. ** $p < .001$. ● marginal ($p < .10$). *ns* = not significant. See the online article for the color version of this figure.

7.9E − 04, $d = .46$, $BF_{10} = 33.00$; left blue (black) bar vs left gold (light gray) bar in Figure 3]. For double-digits, the RDE in fact flipped to become a canonical distance effect (better performance on non-count-list, 2–4–6, relative to count-list, 3–4–5, trials) [$t(59) = 2.67$, $p = .010$, $d = .35$, $BF_{10} = 3.57$; right blue (black) bar in Figure 3 was significantly >0]. The distance effect in the congruent condition was significantly different from that observed in the standard condition [$t(59) = 6.52$, $p = 1.7E-08$, $d = .84$, $BF_{10} = 7.3E+05$; right blue (black) bar vs right gold (light gray) bar in Figure 3].

In sum, we found that for highly overlearned sequences (single-digits), the RDE typically seen in the literature (akin to the standard condition) appears to result from a combination of poor performance on non-count-list trials and good performance on count-list trials. In other words, effects predicted by both the facilitatory and exclusionary hypotheses account for the RDE (the third case noted above). For less familiar sequences (double-digits), the RDE appears to be driven primarily by poor performance on non-count-list trials (providing sole support for the exclusionary hypothesis).

Analysis 3: Additional Effects of Altering Ordinality Rules

Though not of direct theoretical interest, it may be of value to examine how altering ordinality rules affected performance (relative to the standard-rule condition) on the other trial types.

The incongruent-rule condition altered the rule specifically for count-list trials (e.g., 3–4–5). From Analysis 1 above, we saw this rule change exerted only a small impact on count-list performance, which was limited to single-digit sequences. Consistent with a limited effect, performance did not differ between standard and incongruent conditions for non-count-list or mixed$_1$ trials across both single- and double-digit sequences (all *p*s > .15, all *d*s < .20, all $BFs_{10} < .39$). Interestingly, performance was significantly worse for mixed$_{2,3}$ (far distance) trials for both single-digits [$t(59) = -4.43$, $p < .001$, $d = -.57$, $BF_{10} = 493.66$] and double-digits [$t(59) = -3.76$, $p < .001$, $d = -.49$, $BF_{10} = 62.32$]. Indeed performance in the incongruent-mixed$_{2,3}$ trials was worse than any of the other three incongruent trial types (incongruent-count-list, incongruent-non-count-list, incongruent-mixed$_1$) for both single-digit (*p*s < .010, *d*s > .34, $BFs_{10} > 3.56$) and double-digit sequences (*p*s < .00, *d*s > .51, $BFs_{10} > 128$). One possibility is that this condition isolated non-count-list sequences as the only trials that counted as in-order, which may have placed greater processing demands on rejecting permutations of these trials (i.e., mixed$_{2,3}$ trials). This result is broadly consistent with what we saw in the congruent condition, where performance on mixed$_1$ trials—permutations of count-list trials—was worse than any other trial type in that condition for both single-digits (*p*s < .001, *d*s > .69, $BFs_{10} > 1.3E+04$) and double-digits (*p*s < .001, *d*s > .81, $BFs_{10} > 3.8E+05$).

Although the congruent-rule condition altered the rule specifically for non-count-list trials (2–4–6 was to be treated as not-in-order, with rules unchanged for the other trial types), this rule change may have affected performance on other trials. Indeed, it appears that performance in the congruent condition improved for all trial types (i.e., comparing the standard column with the congruent column in Table 1 all *p*s $< .001$, all *d*s $> .50$, all $\mathrm{BFs}_{10} > 680$). In other words, it appears that aligning ordinality rules to be more in keeping with individuals' intuitions about what constitutes numerical order had both a strong (large effect sizes) and a pervasive effect on overall ordinality verification performance. Namely, when participants are permitted to treat only count-list sequences as in-order and all other sequences as not-in-order, performance improved across the board, which further underscores the potential impact this heuristic has on basic conceptions of numerical order.

In sum, analysis of the remaining trial types (i.e., those not immediately central to our main hypotheses, and thus omitted from Analyses 1 and 2) is broadly consistent with our conclusions from the previous analyses. Namely, altering ordinality rules to conform to the notion that only count-list trials are in-order (congruent condition) has a strong and pervasive effect on processing of numerical order. Conversely, altering ordinality rules that contradict the notion that count-list trials must be in order (incongruent condition) has a more limited influence on numerical order processing.

Discussion

There has been a steady increase in work focusing on how people process ordinal relations between numbers, in part because this work carries implications for our basic theoretical understanding of what a number means, as well as our practical understanding of the foundations upon which more sophisticated mathematics are built. A leading explanation is that ordinal processing of numbers is linked to repeated practice with the integer count list (Lyons et al., 2016); but the mechanisms underlying this link remain unclear. For instance, in standard ordinal verification paradigms, participants more rapidly and accurately verify that count-list sequences like 3–4–5 are in-order than non-count-list sequences such as 2–4–6 (aka, the reverse distance effect, RDE). However, it is unclear whether this effect is due to strong count-list processing or poor non-count-list processing, or whether the count list influences ordinal processing in an exclusionary manner, creating a tendency to view anything that does not match the count-list as not-in-order. Here we modified the standard ordinal verification paradigm in terms of the rules determining which sequences were to be classified as in-order versus not-in-order, which allowed us to test each of these hypotheses in a nonmutually-exclusive manner. In turn, this approach allowed us to probe the ultimate source of the RDE, a hallmark of ordinal processing. Finally, examining both highly familiar, overlearned single-digit sequences and less familiar double-digit sequences, we probed the specific mechanisms—representation-based or process-based—by which the count list exerts different influences on how we process the ordinality of numbers.

Facilitatory Versus Exclusionary Hypotheses

We reasoned that if participants default to seeing ordered, non-count-list sequences (e.g., 2–4–6) as not-in-order, then creating a task condition that allowed them to respond in agreement with this heuristic—to be able to consider these sequences as indeed not-in-order, as was the case in the congruent-rule condition—would facilitate performance (relative to the standard-rule condition). That is exactly what we found: participants responded substantially (effect-sizes $> .7$) more efficiently (lower combined response times and error rates) in the congruent than in the standard condition. Importantly, this result obtained for both single-digit ($d = .74$, $p < .001$) and double-digit sequences ($d = 1.04$, $p < .001$), indicating it more likely stems from the process of reciting the count list itself rather than from memorizing specific, overlearned sequences. Instead, an exclusionary heuristic by which all non-count-list sequences are initially perceived as not-in-order (even when they actually are) seems to comprise both a substantial and ubiquitous aspect of how literate, adult participants think about the ordinality of numbers (see Figure 2).

By contrast, we found only minimal evidence to support the facilitatory hypothesis that participants' ordinality processing was influenced by directly matching count-list sequences with overlearned representations stored in memory. Forcing participants to treat count-list sequences as though they were not-in-order (incongruent-rule condition) did incur a significant processing cost (relative to the standard condition), but only in the case of single-digits ($d = .26$, $p = .046$), and not for double-digits ($d = .02$, $p = .872$). Moreover, even in the case of single-digits, the absolute magnitude of this effect was significantly smaller ($p = 4.2\mathrm{E} - 05$) than the boost in performance observed when participants were "permitted" to classify ordered, non-count-list sequences (e.g., 2–4–6) as not-in-order ($d = .74$). Thus, while we did find some evidence that ordinal processing is tied to specific memory traces associated with the count-list (Franklin et al., 2009; Sella et al., 2020; Turconi et al., 2006); this link is relatively weak, and limited only to highly overlearned single-digit quantities (Figures 2–3).

One of the first times that children encounter numbers in an ordered context—indeed, one of the first times many children encounter verbal or symbolic numbers in general—is when hearing and reciting the list of count integers ("one, two, three, four . . ."; Wynn, 1990, 1992). Coupled with the observation that both children and adults are particularly adept at recognizing ordered sequences of numbers that match the count list (e.g., 1–2–3, 29–30–31; Franklin & Jonides, 2009; Lyons & Ansari, 2015), this has led researchers to speculate that our sense of numerical order is strongly tied to our experience reciting the count list (Lyons et al., 2016). What has remained unclear is precisely how the count list shapes our sense of numerical order. Our results suggest that, rather than a *facilitatory* heuristic to see count-list sequences as in-order, instead we saw a much stronger *exclusionary* heuristic to see any type of other sequence—ordered or not—as being not-in-order. That this result was observed in highly numerate adults (university students) is all the more striking, as it suggests that even a decade or more experience working with increasingly complex mathematics does not completely dispel the initial impulse to see a sequence like (2–4–6) as being not in order.

Deconstructing the Ordinal Verification Task

Over the past decade, the ordinal verification task has become a popular tool in the toolkit of numerical cognition researchers (Franklin & Jonides, 2009; Lyons & Beilock, 2009; Turconi et al., 2006). This is in part due to rising interest in understanding numerical ordinality as distinct from numerical cardinality (Lyons et al., 2016). Moreover, the ordinal verification task is simple enough that it can be used in both adults and children (Lyons & Ansari, 2015; Morsanyi et al., 2018). On a theoretical level, the task provides a reliable marker of ordinal processing in the RDE (Goffin & Ansari, 2016); and it consistently predicts more complex numerical skills such as mental arithmetic over and above the contributions of other basic numerical tasks (e.g., Lyons & Beilock, 2011; Morsanyi et al., 2017; Sasanguie et al., 2017). We replicated both of these effects here (gold [light gray] bars in Figure 3, and Table D1, respectively). Furthermore, in the present study, we manipulated the ordinality rules of the ordinal verification task, which allowed us to probe the different ways in which the count list influences ordinal processing. An additional consequence of this approach was that it shed light on the underlying processes that drive task performance. Below, we outline our interpretation of how the current results shape our understanding of the relevant factors in this increasingly popular experimental technique.

Our interpretation is that the rule conditions primarily altered performance by impacting the procedures by which input stimuli are checked. We propose this checking procedure follows two steps. First, one checks the set of three input stimuli together against a preexisting heuristic for what constitutes numerical order. Second, if necessary, one engages a secondary check, based on "brute force" processing (such as verbal recitation and/or pairwise comparison of stimuli). Our main take-aways are (a) that the first check does not include (but in fact excludes—per the exclusionary hypothesis) ordered trials that do not perfectly match the integer count list; (b) this exclusion has a large impact on verification performance; (c) the inclusionary benefit—to actual count-list trials—is relatively small and specific to high-frequency (single-digit) numbers.

In the standard-rule condition, we saw rapid verification that count-list stimuli are in order because one does not need to engage the second level of checks. Performance on all other sequences—including non-count-list (but ordered) trials—was worse because one must engage the second round of checks. This account does not make it clear, however, whether better performance on count-list sequences is due—at least in part—to facilitation during the first check. The incongruent condition gives us our answer: for single-digits, there is some facilitation; for double-digits, there is none. Namely, the incongruent rule change, which reverses the ordinality rule only for count-list trials, should impair performance on count-list trials to the extent that any facilitation these trials might otherwise have enjoyed must now be overridden. However, comparing count-list trials across standard and incongruent conditions showed only a small decrement in performance for single-digits, and no significant decrement for double-digits. This led us to conclude there is only minimal support for the facilitatory hypothesis. Moreover, it is important to emphasize we saw evidence favoring the facilitatory hypothesis only for single-digits, which are encountered more frequently (Dehaene & Mehler, 1992); and thus are more likely to generate memorized sequences (Lyons & Beilock, 2009; Sella et al., 2020). We interpreted this to mean facilitation of count-list trials—during the first round of checks—is likely driven by memory-based retrieval of specific representations. More broadly, this suggests that count-list trials in the standard condition may be indicative of memory-based factors in numerical order processing, though this conclusion is perhaps limited to single-digit numbers.

To understand the potential negative impact of resorting to a second round of checks for non-count-list trials, we must turn to the congruent-rule condition. This condition essentially obviates the need for a second round of checks because all trials that fail to pass the first check can—correctly in this case—be classified as not in order. This simplification of the checking procedure should lead to an overall improvement in performance, including for non-count-list trials, which is exactly what we saw (see Table 1). Focusing on how the count list impacts how we think about trials that are in fact in-order, when we compared non-count-list performance between the standard and congruent condition, we found effect sizes roughly 3 times larger than the facilitatory effect in single-digits, and still for double-digits. Recall that in this interpretation, the second round of checks is necessary for non-count-list trials only because they were excluded from the first round of checks. The negative impact on non-count-list trial processing of this exclusion thus appears to be both substantial and pervasive.

To infer the nature of the second round of checks, we can turn to single- versus double-digit results in the congruent condition. If the second round operates via checking a second set of memorized triplet representations, then we should again see an exclusionary effect primarily for single-digits, or at minimum the effect should be stronger for single- relative to double-digits. Instead, we saw the opposite: (the right blue bar was significantly higher than the left blue [dark gray] bar in Figure 1). This suggests the need for a more general mechanism that could apply to both single- and double-digits (and in fact apply more strongly to the latter). Hence, we propose either verbal recitation of sequences ("2 . . . 4 . . . 6 . . . ," "28 . . . 30 . . . 32"), or "brute force" pairwise comparison (e.g., 2 vs 4, 4 vs 6; 28 vs 30, 30 vs 32), which can be implemented for any numerical sequence. To get at this final distinction (recitation or comparison), we can turn to the mixed trials (Analysis 3). In particular, we saw especially poor performance for the mixed$_{23}$-incongruent and mixed$_{1}$-congruent trials (see Table 1). In each case, these trials are the strongest foils to the only type of sequence that should be classified as in-order in that condition. For instance, in the incongruent condition, only non-count-list trials (e.g., 2–4–6, 28–30–32) are in-order; hence, mixed$_{23}$-incongruent trials (e.g., 2–6–4, 28–32–30) are the only type of sequence that is a direct permutation of the only type of sequence that is a valid ordered sequence (for that rule condition). Note that a similar situation applies to mixed$_{1}$-congruent in the congruent condition. We argue this result suggests participants tend to engage a verbal recitation processes at the second stage of checks, which is what makes these permutation especially challenging. That is, permuted foils are more likely to be mistaken for in-order trials (higher error rates) and take longer to reject (longer RTs), as is evidenced from Tables C1 and C2.

To summarize, we argue that the standard ordinal verification task proceeds by first checking the input set against the integer count list, followed by a more general verbal recitation process. The congruent obviates the need for the second verbal recitation check, thus substantially improving performance across the board. The incongruent-rule provides an additional challenge in inhibiting the first round of checks, though this effect is much smaller and limited to sequences that can be checked via direct retrieval mechanisms. The RDE is thus driven by a combination of count-list facilitation and non-count-list impediment (due to being excluded from the first round of checks) for single-digit sequences, and seemingly entirely by a non-count-list impediment for double-digit sequences.

Broader Implications

On a theoretical level, these results suggest our intuitive sense of a fundamental principle of mathematics—ordinality—is fundamentally shaped by overrepresentation of the count list, which may steer one toward mathematically incorrect conclusions (the sequence 2–4–6 is of course very much in numerical order). This idea is reminiscent of how our intuitions based on more familiar natural numbers (positive integers) can lead us astray when reasoning about less familiar, more counterintuitive types of numbers, such as negative numbers ($-1 > -2$) or fractions (½ > ¼), (Siegler & Lortie-Forgues, 2014). What is striking about the current data is they suggest we should add the notion that 2–4–6 is in numerical order to the list of counterintuitive numerical concepts.

A second implication—for which we also provide evidence here—is that the counterintuitive nature of non-count-list ordered sequences can help explain an increasingly well-known phenomenon in the numerical cognition literature—the reversal of the distance effect (RDE) when making ordinal judgments (Franklin & Jonides, 2009; Franklin et al., 2009; Goffin & Ansari, 2016; Lyons & Ansari, 2015; Lyons & Beilock, 2013; Turconi et al., 2006). In particular, we showed (see Figure 3) that an exclusionary heuristic against sequences like 2–4–6 and 28–30–32 plays a consistent role in producing the RDE (single-digits: d = .42, double-digits: d = .65), and a proclivity toward sequences like 1–2–3 and 29–30–31 contributes less so, obtaining significance only for more familiar overlearned number sequences (single-digits: d = .28, double-digits: d = .22). Thus, we propose that the more pervasive influence responsible for the RDE in standard ordinality verification paradigms is relatively poor performance on ordered, non-count-list sequences (e.g., 2–4–6, 28–30–32), rather than relatively good performance on count-list sequences (e.g., 1–2–3, 29–30–31).

A third implication concerns the question of when and how in development the heuristic to view only numerical sequences that match the count list as ordered. As noted above, the count list is one of the first numerical procedures children learn—often preceding even their understanding of what the number words they are saying mean with respect to specific numerical quantities, or that the order in which they are saying those words is of any particular import (Wynn, 1990, 1992). Furthermore, the RDE in ordinal judgments is present in children as early as first grade (Lyons & Ansari, 2015). Thus, it seems possible the heuristic to see sequences like 2–4–6 as being not-in-order may be linked to the earliest stages of children's acquisition of number symbols. However, we must stress that the current data are drawn from adult participants, so any developmental inferences must be taken with a healthy dose of skepticism. Instead, we present the idea that children as young as 4 or 5 years of age will display a tendency to exclude non-count-list sequences from their idea of what it means for numbers to be in-order (even when they are) as a speculative hypothesis that may warrant future testing.

A final question to consider is whether the exclusionary heuristic in ordinal processing revealed here is specific to numerical stimuli, or whether it generalizes to other types of ordinal sequences, such as letters of the alphabet, months, days of the week, and so on. Each of these ordinal lists is often recited according to a standard verbal order—indeed, the list for the letters of the alphabet has even been canonized as a childhood song in many languages. Moreover, non-numerical ordered sequences show RDEs (Franklin et al., 2009; Morsanyi et al., 2017); and numerical and non-numerical ordering appear to share overlapping neural substrates (Fulbright et al., 2003; Ischebeck et al., 2008). Thus, we hypothesize that the tendency to initially see ordered sequences that are not part of a standard verbal recitation list as not-in-order may be just as strong in non-numerical lists. For instance, we hypothesize that participants will have to overcome an initial impulse to see sequences like B–D–F, February–April–June, or Tuesday–Thursday–Saturday as not-in-order. Of course, as with the developmental hypothesis discussed above, the current data cannot answer the question; rather, our data here give reason to pose these questions in the first place, and, we hope, the impetus for future work to address them.

Conclusion

To our knowledge, we provide the first evidence that literate adults find the notion that sequences like 2–4–6 are in order to be counterintuitive, which is driven by a heuristic to initially view any sequences that do not match the count list as being not in numerical order. This tendency exerts a relatively large and consistent influence on ordinal verification performance (effect sizes $\geq$ .74), and it helps explain the reversal of the distance effect. We speculate that this heuristic may have early developmental origins tied to how children learn to intuitively think about the ordinal arrangement of numbers via the integer count list. More broadly, this work helps reveal key processes that shape the cognitive rules by which people think about one of the foundational principles of the broader system of mathematical relations—numerical order.

Context of Research

Previous work has established that ordinality is a foundational principle of numerical representation, that it is distinct from cardinality, and that it is a strong predictor of more complex forms of mathematical processing in both children and adults. But where does our sense of ordinality come from? When children learn to recite the count list, they are learning the specific order in which integers, in the form of number words, should be arranged. It thus seems intuitive that our sense of order is linked to our extensive experience counting integers. The question, though, is how. Does one memorize specific sequences (1–2–3, 7–8–9) that make one very good at recognizing these specific sequences as ordered sets? Does one form a more general heuristic: any sequence that does not match the count list is not in order? Or both? Here we modified a popular paradigm used in the literature to assess ordinal processing, and using this novel approach, we demonstrate that the answer is "both" for highly familiar, overlearned sequences of single-digit numbers. However, the data also showed that an exclusionary heuristic (anything that does not match the count list is not in order) is both the stronger and more widespread way in which our experience with the count list influences our sense of how numbers should—and should not—be ordered. From prior work, we know that intuitions about numbers can lead us astray when reasoning about less familiar, less intuitive types of numbers, such as negative numbers ($-1 > -2$) or fractions (½ > ¼). The current data suggest another such example, but this time in the realm of ordinality: Even educated adults find it counterintuitive to extend their sense of order to include sequences that deviate only slightly from the typical count-list (e.g., 3–5–7).

References

Ansari, D. (2008). Effects of development and enculturation on number representation in the brain. *Nature Reviews Neuroscience*, *9*(4), 278–291. https://doi.org/10.1038/nrn2334

Bourassa, A. (2014). *Numerical sequence recognition: Is familiarity or ordinality the primary factor in performance?* [Master's thesis, Carleton University, Ottawa, Canada]. https://curve.carleton.ca/282e5db5-4f36-4e5f-93fa-7a47bbaf5640

Buckley, P. B., & Gillman, C. B. (1974). Comparisons of digits and dot patterns. *Journal of Experimental Psychology*, *103*(6), 1131–1136. https://doi.org/10.1037/h0037361

Butterworth, B. (1999). *The mathematical brain*. Macmillan.

Bynner, J., & Parsons, S. (2005). *Does numeracy matter more?* National Research and Development Centre for Adult Literacy and Numeracy, Institute of Education.

Coles, A., & Sinclair, N. (2018). Re-thinking 'normal' development in the early learning of number. *Journal of Numerical Cognition*, *4*(1), 136–158. https://doi.org/10.5964/jnc.v4i1.101

Crawford, C., & Cribb, J. (2013). *Reading and maths skills at Age 10 and earnings later in life: A brief analysis using the British Cohort Study* (Report No. 3). Center for Analysis of Youth Transitions, Department of Education (United Kingdom).

Daro, P., Mosher, F., & Corcoran, T. (2011). *Learning trajectories in mathematics: A foundation for standards, curriculum, assessment, and instruction (CPRE Research Report #RR 68)*. Consortium for Policy Research in Education. https://doi.org/10.12698/cpre.2011.rr68

Dehaene, S., & Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition*, *43*(1), 1–29. https://doi.org/10.1016/0010-0277(92)90030-L

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428–1446. https://doi.org/10.1037/0012-1649.43.6.1428

Faulkenberry, T. J., Ly, A., & Wagenmakers, E. (2020). Bayesian inference in numerical cognition: A tutorial using JASP. *Journal of Numerical Cognition*. Advance online publication. https://doi.org/10.5964/jnc.v6i2.288

Franklin, M. S., & Jonides, J. (2009). Order and magnitude share a common representation in parietal cortex. *Journal of Cognitive Neuroscience*, *21*(11), 2114–2120. https://doi.org/10.1162/jocn.2008.21181

Franklin, M. S., Jonides, J., & Smith, E. E. (2009). Processing of order information for numbers and months. *Memory & Cognition*, *37*(5), 644–654. https://doi.org/10.3758/MC.37.5.644

Fulbright, R. K., Manson, S. C., Skudlarski, P., Lacadie, C. M., & Gore, J. C. (2003). Quantity determination and the distance effect with letters, numbers, and shapes: A functional MR imaging study of number processing. *American Journal of Neuroradiology*, *24*(2), 193–200.

Gerardi, K., Goette, L., & Meier, S. (2013). Numerical ability predicts mortgage default. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(28), 11267–11271. https://doi.org/10.1073/pnas.1220568110

Goffin, C., & Ansari, D. (2016). Beyond magnitude: Judging ordinality of symbolic number is unrelated to magnitude comparison and independently relates to individual differences in arithmetic. *Cognition*, *150*, 68–76. https://doi.org/10.1016/j.cognition.2016.01.018

Ischebeck, A., Heim, S., Siedentopf, C., Zamarian, L., Schocke, M., Kremser, C., Egger, K., Strenge, H., Scheperjans, F., & Delazer, M. (2008). Are numbers special? Comparing the generation of verbal materials from ordered categories (months) to numbers and other categories (animals) in an fMRI study. *Human Brain Mapping*, *29*(8), 894–909. https://doi.org/10.1002/hbm.20433

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford University Press.

LeFevre, J. A., & Bisanz, J. (1986). A cognitive analysis of number-series problems: Sources of individual differences in performance. *Memory & Cognition*, *14*(4), 287–298. https://doi.org/10.3758/BF03202506

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*(2), 256–274. https://doi.org/10.1016/0022-0965(91)90062-W

Lyons, I. M., & Ansari, D. (2015). Numerical order processing in children: From reversing the distance-effect to predicting arithmetic. *Mind, Brain and Education*, *9*(4), 207–221. https://doi.org/10.1111/mbe.12094

Lyons, I. M., & Beilock, S. L. (2009). Beyond quantity: Individual differences in working memory and the ordinal understanding of numerical symbols. *Cognition*, *113*(2), 189–204. https://doi.org/10.1016/j.cognition.2009.08.003

Lyons, I. M., & Beilock, S. L. (2011). Numerical ordering ability mediates the relation between number-sense and arithmetic competence. *Cognition*, *121*(2), 256–261.

Lyons, I. M., & Beilock, S. L. (2013). Ordinality and the nature of symbolic numbers. *The Journal of Neuroscience*, *33*(43), 17052–17061. https://doi.org/10.1523/JNEUROSCI.1775-13.2013

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*(5), 714–726.

Lyons, I. M., Vogel, S., & Ansari, D. (2016). On the ordinality of numbers: A review of neural and behavioral studies. *Progress in Brain Research*, *227*, 187–221. https://doi.org/10.1016/bs.pbr.2016.04.010

Morey, R. D., & Rouder, J. N. (2015). *BayesFactor 0.9.11-1*. Comprehensive R Archive Network.

Morsanyi, K., O'Mahony, E., & McCormack, T. (2017). Number comparison and number ordering as predictors of arithmetic performance in adults: Exploring the link between the two skills, and investigating the question of domain-specificity. *Quarterly Journal of Experimental Psychology*, *70*(12), 2497–2517. https://doi.org/10.1080/17470218.2016.1246577

Morsanyi, K., van Bers, B. M. C. W., O'Connor, P. A., & McCormack, T. (2018). Developmental dyscalculia is characterized by order processing deficits: Evidence from numerical and non-numerical ordering tasks. *Developmental Neuropsychology*, *43*(7), 595–621. https://doi.org/10.1080/87565641.2018.1502294

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, *215*(5109), 1519–1520. https://doi.org/10.1038/2151519a0

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

Reyna, V. F., Nelson, W. L., Han, P. K., & Dieckmann, N. F. (2009). How numeracy influences risk comprehension and medical decision making. *Psychological Bulletin*, *135*(6), 943–973. https://doi.org/10.1037/a0017327

Sasanguie, D., Lyons, I. M., De Smedt, B., & Reynvoet, B. (2017). Unpacking symbolic number comparison and its relation with arithmetic in adults. *Cognition*, *165*, 26–38. https://doi.org/10.1016/j.cognition.2017.04.007

Sella, F., Sasanguie, D., & Reynvoet, B. (2020). Judging the order of numbers relies on familiarity rather than activating the mental number line. *Acta Psychologica*, *204*, 103014. https://doi.org/10.1016/j.actpsy.2020.103014

Šidák, Z. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, *62*(318), 626–633. https://doi.org/10.2307/2283989

Siegler, R. S., & Lortie-Forgues, H. (2014). An integrative theory of numerical development. *Child Development Perspectives*, *8*(3), 144–150. https://doi.org/10.1111/cdep.12077

Turconi, E., Campbell, J. I. D., & Seron, X. (2006). Numerical order and quantity processing in number comparison. *Cognition*, *98*(3), 273–285. https://doi.org/10.1016/j.cognition.2004.12.002

Wynn, K. (1990). Children's understanding of counting. *Cognition*, *36*(2), 155–193. https://doi.org/10.1016/0010-0277(90)90003-3

Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cognitive Psychology*, *24*(2), 220–251. https://doi.org/10.1016/0010-0285(92)90008-P

Appendix A

Complete Ordinality Task Trial List

Table A1
All Trials

Single-digit					Double-digit				
Left	Center	Right	Trial type	Distance	Left	Center	Right	Trial type	Distance
1	2	3	Count-list	1	18	19	20	Count-list	1
2	3	4	Count-list	1	19	20	21	Count-list	1
3	4	5	Count-list	1	28	29	30	Count-list	1
4	5	6	Count-list	1	29	30	31	Count-list	1
5	6	7	Count-list	1	38	39	40	Count-list	1
6	7	8	Count-list	1	39	40	41	Count-list	1
7	8	9	Count-list	1	49	50	51	Count-list	1
1	3	5	Non-count-list	2, 3	18	20	22	Non-count-list	2, 3
2	4	6	Non-count-list	2, 3	27	29	31	Non-count-list	2, 3
4	6	8	Non-count-list	2, 3	37	39	41	Non-count-list	2, 3
5	7	9	Non-count-list	2, 3	48	50	52	Non-count-list	2, 3
1	4	7	Non-count-list	2, 3	17	20	23	Non-count-list	2, 3
2	5	8	Non-count-list	2, 3	26	29	32	Non-count-list	2, 3
3	6	9	Non-count-list	2, 3	37	40	43	Non-count-list	2, 3
1	3	2	Mixed	1	18	20	19	Mixed	1
2	4	3	Mixed	1	19	21	20	Mixed	1
3	5	4	Mixed	1	28	30	29	Mixed	1
4	6	5	Mixed	1	29	31	30	Mixed	1
5	7	6	Mixed	1	38	40	39	Mixed	1
6	8	7	Mixed	1	39	41	40	Mixed	1
7	9	8	Mixed	1	49	51	50	Mixed	1
1	5	3	Mixed	2, 3	18	22	20	Mixed	2, 3
2	6	4	Mixed	2, 3	27	31	29	Mixed	2, 3
4	8	6	Mixed	2, 3	37	41	39	Mixed	2, 3
5	9	7	Mixed	2, 3	48	52	50	Mixed	2, 3
1	7	4	Mixed	2, 3	17	23	20	Mixed	2, 3
2	8	5	Mixed	2, 3	26	32	29	Mixed	2, 3
3	9	6	Mixed	2, 3	37	43	40	Mixed	2, 3
2	1	3	Mixed	1	19	18	20	Mixed	1
3	2	4	Mixed	1	20	19	21	Mixed	1
4	3	5	Mixed	1	29	28	30	Mixed	1
5	4	6	Mixed	1	30	29	31	Mixed	1
6	5	7	Mixed	1	39	38	40	Mixed	1
7	6	8	Mixed	1	40	39	41	Mixed	1
8	7	9	Mixed	1	50	49	51	Mixed	1
3	1	5	Mixed	2, 3	20	18	22	Mixed	2, 3
4	2	6	Mixed	2, 3	29	27	31	Mixed	2, 3
6	4	8	Mixed	2, 3	39	37	41	Mixed	2, 3
7	5	9	Mixed	2, 3	50	48	52	Mixed	2, 3
4	1	7	Mixed	2, 3	20	17	23	Mixed	2, 3
5	2	8	Mixed	2, 3	29	26	32	Mixed	2, 3
6	3	9	Mixed	2, 3	40	37	43	Mixed	2, 3

Note. Table A1 provides a complete list of trials used in the numerical ordinality judgment task.

(*Appendices continue*)

Appendix B

Ordinality Task Instructions

Appendix B provides verbatim instructions given to participants for each of the rule conditions. During instructions, participants were also provided with several examples for each condition, which are noted below. In addition, following instructions for a given condition, participants completed a block of 24 practice trials. During the practice block, feedback was provided after each trial (no feedback was given during the main blocks of trials—i.e., those used for analysis here).

Standard Condition Instructions

For this part of the experiment, to count as IN-ORDER, a set of numbers must simply be INCREASING (left–right). Note that whether or not numbers are adjacent in the count list does not matter for this part of the experiment. Several examples of IN-ORDER sets are listed below:

{1 2 3}
{7 8 9}
{3 5 7}
{1 4 7}
{13 16 19}
{29 30 31}
{37 40 43}
{48 49 50}

For this part of the experiment, all other sets of numbers should be considered NOT-IN-ORDER. Several examples of NOT-IN-ORDER sets are listed below:

{2 1 3}
{9 7 8}
{6 4 2}
{3 9 5}
{16 14 15}
{12 20 16}
{32 30 28}
{47 53 50}

[The following was shown as a reminder on the screen immediately before a participant started a block of trials (practice or main experiment).]

Remember: For this part of the experiment, ALL sets that are increasing (e.g., 1–2–3, 2–4–6) count as in-order. All other sets count as not-in-order.

Congruent Condition Instructions

For this part of the experiment, to count as IN-ORDER, a set of numbers must be INCREASING (left–right) like the count-list. Several examples of IN-ORDER sets are listed below:

{1 2 3}
{7 8 9}
{14 15 16}
{28 29 30}
{49 50 51}

For this part of the experiment, all other sets of numbers should be considered NOT-IN-ORDER. Note this means that increasing, nonadjacent sets (like 2–4–6) should be considered NOT-IN-ORDER. Several examples of NOT-IN-ORDER sets are listed below:

{2 1 3}
{9 8 7}
{2 4 6}
{3 5 9}
{16 14 15}
{12 16 20}
{28 30 32}
{47 53 50}

[The following was shown as a reminder on the screen immediately before a participant started a block of trials (practice or main experiment).]

Remember: For this part of the experiment, ONLY sets that are increasing and adjacent (e.g., 1–2–3) count as in-order. All other sets count as not-in-order.

Incongruent Condition Instructions

For this part of the experiment, to count as IN-ORDER, a set of numbers must be INCREASING (left–right), but NOT ADJACENT.

Several examples of IN-ORDER sets are listed below:

{2 4 6}
{3 5 7}
{2 5 8}
{15 17 19}
{18 20 22}
{27 30 33}
{38 41 44}
{46 49 52}

For this part of the experiment, all other sets of numbers should be considered NOT-IN-ORDER. Note this means that increasing, adjacent sets (like 1–2–3) should be considered NOT-IN-ORDER.

Several examples of NOT-IN-ORDER sets are listed below:

{1 2 3}
{2 1 3}
{7 8 9}
{2 6 4}
{9 3 6}
{14 15 16}
{21 20 19}
{27 33 33}
{47 53 50}

[The following was shown as a reminder on the screen immediately before a participant started a block of trials (practice or main experiment).]

Remember: For this part of the experiment, ONLY sets that are increasing and NOT adjacent (e.g., 2–4–6) count as in-order. All other sets count as not-in-order.

(*Appendices continue*)

Appendix C

Response Times and Error Rates

Table C1
Response Times

Trial type	Single-digit Standard	Single-digit Congruent	Single-digit Incongruent	Double-digit Standard	Double-digit Congruent	Double-digit Incongruent
Count-list	973 (33)	848 (29)	1,012 (32)	1,151 (37)	997 (31)	1,150 (33)
Non-count-list	1,039 (36)	857 (28)	1,069 (34)	1,278 (37)	978.7 (32.5)	1,299 (37)
Mixed (1)	1,128 (37)	943 (31)	1,143 (38)	1,294 (37)	1,111 (32)	1,319 (38)
Mixed (2, 3)	1,054 (30)	800 (25)	1,087 (34)	1,324 (36)	929 (33)	1,391 (39)

Note. Table C1 gives all subcondition means and standard errors (the latter in parentheses) in terms of response times (RTs).

Table C2
Error Rates

Trial type	Single-digit Standard	Single-digit Congruent	Single-digit Incongruent	Double-digit Standard	Double-digit Congruent	Double-digit Incongruent
Count-list	3.3 (0.9)	3.9 (0.9)	4.5 (1.0)	3.7 (1.0)	4.8 (1.0)	4.4 (1.0)
Non-count-list	5.7 (1.2)	2.4 (0.8)	3.3 (0.5)	8.6 (1.5)	2.0 (0.9)	5.8 (0.8)
Mixed (1)	7.7 (1.4)	9.9 (1.7)	5.5 (1.9)	9.5 (1.2)	10.8 (1.9)	8.5 (2.1)
Mixed (2, 3)	5.7 (1.2)	1.9 (0.7)	13.9 (2.1)	8.2 (1.5)	1.6 (1.0)	13.4 (2.2)

Note. Table C2 gives all subcondition means and standard errors (the latter in parentheses) in terms of error rates (ERs, % incorrect).

Appendix D

Correlations With Arithmetic

Table D1 shows correlations between each trial type in the ordinal verification task and the mental arithmetic task. The mental arithmetic task was the same as that described in Lyons and Beilock (2011). A higher score on the mental arithmetic task indicated better performance, so a negative correlation indicates better ordinal verification performance was associated with better mental arithmetic performance. Note also that one participant was removed from this analysis because they did not follow instructions on the mental arithmetic task, and so the correlations below proceeded with $N = 59$. Finally, mental arithmetic data are included in the open-source online data repository (https://osf.io/z4ngj/), should the interested reader wish to explore these correlations in greater detail.

(*Appendices continue*)

Table D1
Correlations Between Each Trial Type in the Ordinal Verification Task and the Mental Arithmetic Task (N = 59)

Ordinality	Digits	Rule condition	Trial type	Correlation (p-val)
Ordered	1-Digit	Standard	Count-list	$r = -.53$ (p = 2E − 05)
			Non-count-list	$r = -.45$ (p = 3E − 04)
		Incongruent	Count-list	$r = -.55$ (p = 7E − 06)
			Non-count-list	$r = -.53$ (p = 2E − 05)
		Congruent	Count-list	$r = -.47$ (p = 2E − 04)
			Non-count-list	$r = -.36$ ($p = .005$)
	2-Digit	Standard	Count-list	$r = -.48$ (p = 1E − 04)
			Non-count-list	$r = -.47$ (p = 2E − 04)
		Incongruent	Count-list	$r = -.40$ ($p = .002$)
			Non-count-list	$r = -.50$ (p = 6E − 05)
		Congruent	Count-list	$r = -.43$ (p = 7E − 04)
			Non-count-list	$r = -.40$ ($p = .002$)
Mixed	1-Digit	Standard	Mixed (1)	$r = -.54$ (p = 1E − 05)
			Mixed (2, 3)	$r = -.46$ (p = 2E − 04)
		Incongruent	Mixed (1)	$r = -.53$ (p = 2E − 05)
			Mixed (2, 3)	$r = -.55$ (p = 8E − 06)
		Congruent	Mixed (1)	$r = -.55$ (p = 8E − 06)
			Mixed (2, 3)	$r = -.39$ ($p = .002$)
	2-Digit	Standard	Mixed (1)	$r = -.44$ (p = 4E − 04)
			Mixed (2, 3)	$r = -.47$ (p = 2E − 04)
		Incongruent	Mixed (1)	$r = -.50$ (p = 6E − 05)
			Mixed (2, 3)	$r = -.48$ (p = 1E − 04)
		Congruent	Mixed (1)	$r = -.52$ (p = 3E − 05)
			Mixed (2, 3)	$r = -.48$ (p = 1E − 04)

Note. Table D1 shows the correlation between performance on each trial-type of the ordinality verification task and the mental arithmetic task. Combined Performance (CP) values were used for the former, meaning a lower value indicates better performance. Net correct in the allotted time (see Lyons & Beilock, 2011, for details) was used to score the arithmetic task, meaning a higher value indicated a higher score. Hence, a negative correlation (r-value) indicates better performance on the ordinality task was associated with better performance on the arithmetic task. Zero-order r-values ($df = 57$) are given in the rightmost column, with associated p-values in parentheses.

Received May 17, 2019
Revision received October 24, 2020
Accepted November 15, 2020 ■