Numeral order and the operationalization of the numerical system

David Muñez[1], Josetxu Orrantia[2], Laura Matilla[2], & Maria Rosario Sanchez[2]

[1]Center for Research in Child Development. National Institute of Education. Nanyang Technological University, Singapore.

[2]Department of Developmental and Educational Psychology. Faculty of Education. Universidad de Salamanca, Spain.

Abstract

Recent years have witnessed an increase in research on how numeral ordering skills relate to children's and adults' mathematics achievement both cross-sectional and longitudinally. Nonetheless, it remains unknown which core competency numeral ordering tasks measure, which cognitive mechanisms underly performance on these tasks, and why numeral ordering skills relate to arithmetic and math achievement. In the current study, we focused on the processes underlying decision-making in the numeral order judgment task with triplets to investigate these questions. A drift-diffusion model for two-choice decisions was fit to data from ninety-seven undergraduates. Findings aligned with the hypothesis that numeral ordering skills reflected the operationalization of the numerical system, where small numbers provide more evidence of an ordered response than large numbers. Furthermore, the pattern of findings suggested that arithmetic achievement was associated with the accuracy of the ordinal representations of numbers.

1. INTRODUCTION

Albeit the ontogenesis of numeral ordinality is not clear yet, numeral ordering is becoming a rather influential line of research on children's and adults' numerical understanding. The ability to discriminate whether a string of Arabic numbers is in order has emerged as a robust predictor of arithmetic and math achievement in both children (Lyons & Ansari, 2015; Lyons, Price, Vaessen, Blomert, & Ansari, 2014; Morsanyi, O'Mahony, & McCormack, 2017) and adults (Goffin & Ansari, 2016; Lyons & Beilock, 2011; Vogel et al., 2017; Vos, Sasanguie, Gevers, & Reynvoet, 2017). There is evidence that such ability is a stronger explanatory variable of later math achievement, and plays a more relevant role in early screening of mathematical learning disabilities, than other basic numeracy skills such as number comparison or the ability to discriminate between the magnitudes of two numbers (Morsanyi, van Bers, O'Connor, & McCormack, 2018). This has spurred empirical and theoretical work about the nature of the numerical representations that are activated during numeral order judgments and whether item information (cardinality) and order information (ordinality) can be separated to some level. Nonetheless, it remains unclear which cognitive mechanisms underpin performance on numeral order judgments, which core competency numeral ordering skills measure, whether the ordinality of numbers (or the position of each numeral in the counting sequence) is activated in the context of numeral order judgments, and how the ability to judge the order of a string of numerals relates to math and arithmetic proficiency. In other words, what do numeral ordering skills reflect?

1.1. The representation of numeral order in memory: symbol-symbol associations

Although a long tradition of studies have claimed that ordinal processing is little more than a special case of cardinal processing, numeral order judgments that require participants to assess the order of three numerals usually render behavioral evidence that is not in line with that typically hypothesized in the context of cardinality processing tasks. For instance, behavioral studies have found that a *numerical distance effect* (a signature of cardinality processing) is rarely revealed in numeral order judgments (e.g., Goffin & Ansari, 2016; Lyons & Beilock, 2013; Turconi, Campbell, & Seron, 2006; Turconi & Seron, 2002; Zorzi, Di Bono, & Fias, 2011). The *numerical distance effect* shows, for example, that deciding whether “5” is larger than “4” is more difficult than deciding whether “7” is larger than “2”. In fact, numeral order judgments are characterized by the opposite effect— a *reverse distance effect*. Participants are faster and less error-prone on ordered trials showing adjacent numbers (e.g., [2-3-4]). Thus, the *reverse distance effect* has been formulated as the behavioral signature of ordinal processing and a consequence of how numeral order is represented in memory (Franklin, Jonides, & Smith, 2009; Goffin & Ansari, 2016; Lyons & Beilock, 2013; Turconi et al., 2006).

In contrast to a magnitude-based account in which the representation of numeral order in memory would reflect the cardinal value of each number, it is thought that the *reverse distance effect* reflects the strength of symbol-symbol associations that are stored in memory, which are independent of cardinal representations (Goffin & Ansari, 2016; Lyons & Beilock, 2013; Rubinsten & Sury, 2011). Although the processes underlying the acquisition of these associations are unknown, it has been suggested that chaining mechanisms, where each element becomes a stimulus for the next element, might contribute to store the string 1-to-9 as an ordered sequence of elements. Symbol-symbol associations between numbers would originate from an iterative process where number symbols are connected to verbal numbers of the counting sequence during childhood. Such connections among symbols would create memory traces that benefit the settlement of number symbols as an ordered sequence in long-term memory (Lyons & Beilock, 2009, 2013). These ordinal associations would trigger memory retrieval processes during numeral order judgments (Logan and Cowan, 1984).

1.2. Numeral order mechanisms and numeral ordering skills

Although the *reverse distance effect* is conceptualized as the behavioral consequence of how numeral order is encoded and represented in memory, the ability to recognize the order of a string of numerals (i.e., numeral ordering skills) does not just relate to participants' proficiency to retrieve symbol-symbol associations from long-term memory. For instance, it has been suggested that serial-search mechanisms (similar to those that are presumed in more general order contexts) might be activated during order judgments (Franklin, et al., 2009; Jou, 2003; Turconi et al., 2006; Turconi. Jemel, Rossion, & Seron, 2004). Participants would mentally scan the number line item by item to find the digits that are shown in each stimulus (pairs, triplets) before a response is triggered. Increasing the distance between numerals would expand the size of the hypothetical to-be-scanned vector of information and affect serial-search performance (Franklin et al., 2009). Both a memory-retrieval mechanism (based on symbol-symbol associations) and a serial-search mechanism explain the *reverse distance effect* and render similar predictions regarding performance on ordered trials that are presented in ascending order.

There is also behavioral evidence that suggests that numeral order processing, as well as broader ordinal processing abilities (Botvinick & Watanabe, 2007; Marshuetz & Smith, 2006), might involve magnitude-based mechanisms. For instance, a canonical *numerical distance effect* has been reported when the distance between constituents is larger than three[1] (e.g., [1-5-9]) and in non-ordered trials (e.g., [5-2-7]). These findings have led some authors to suggest that certain ordering task parameters may activate magnitude-based mechanisms in numeral order processing (Lyons & Ansari, 2016). It has been argued that a magnitude comparison strategy can be used for all types of trials, but can be bypassed by an association-based solution when the associations between numerals

[1] The circumstances under which the reverse distance effect (RDE) is revealed are not universal. Whereas some studies have not found an RDE for distances larger than two (Goffin and Ansari, 2016), others have found that is revealed for distances up to three (Lyons & Ansari, 2015), and some authors have not found this effect (Morsanyi et al., 2017; Vogel, Remark, & Ansari, 2015).

are strong (Logan & Cowan, 1984; Vos et al., 2017; for similar theoretical positions see Leth-Steensen & Marley, 2000; Marshuetz & Smith, 2006).

These findings also suggest that numeral ordering skills may relate to both the ability to retrieve information from memory and the ability to compare the magnitude of symbol numbers, as suggested by neuroscience evidence (e.g., Franklin et al., 2009; Matejko, Hutchison, & Ansari, 2019; Zorzi, et al., 2011; although see Lyons & Beilock, 2013, for different results). In turn, this presumes a shared representation of ordinal and cardinal information (Caplan, 2015), or at least a common cognitive architecture for both magnitude and ordinal processing (van Dijck & Fias, 2011). This possibility—that different types of mechanisms are activated during numeral order judgments—challenges the interpretation of the skills that are measured with numeral order judgment tasks as well as the interpretation of interindividual differences (i.e., the magnitude of the *reverse distance effect* depends on two different mechanisms).

Behavioral studies have also found that the *reverse distance effect* correlates with measures of numeral cardinality (Attout & Majerus, 2015; Orrantia et al., 2019, but see Goffin & Ansari, 2016) and that both cardinality skills and numeral ordering skills jointly explain variance in math skills. Nonetheless, it is unclear why numeral ordering skills relate to math and arithmetic proficiency. Some authors have suggested that numeral ordering skills may reflect the ability to temporarily store numerical information in memory, for instance, during mental subtraction (O'Connor, Morsanyi, & McCormack, 2017). Others have focused on the mechanisms involved in numeral order judgments. For instance, Lyons and Ansari (2015) found that performance on trials that showed sequential trials explained a larger amount of variance in arithmetic performance than other types of trials (see also Goffin & Ansari, 2016), which suggests that the rate of access to symbol-symbol associations may reflect, for instance, differences in arithmetic facts' retrieval. In contrast, Morsanyi et al. (2018) found that performance on non-ordered trials (which likely rely on magnitude-based mechanisms) was a better indicator of arithmetic proficiency than performance on ordered trials. Another group of authors has suggested that domain-general ordering skills rather than numeral ordering skills are responsible for the association between numeral order and math proficiency (O'Connor et al., 2017; Vos et al., 2017). Although this myriad of findings reflects different approaches to the understanding and conceptualization of numeral ordering skills, all of them focus on the differences rather than on the underpinning mechanisms of such differences. In other words, do such disparities in arithmetic/math achievement reflect different representations of numeral order? Or different mechanisms during numeral order judgments? Why high achievers process the order of a sequence of numerals more efficiently?

1.3. The current study: numeral ordinality and the operationalization of the numerical system

In the current study, we argue that the *reverse distance effect* is not the only pattern that emerges during numeral order judgments and that the representation of numeral order in memory likely includes other traces directly related to the positional value or location of each number symbol in the counting sequence[2] or how the numerical system is operationalized. The bounded nature of the counting list means that small numbers are more likely to cause an ascending sequence. For instance, the probability that a number "1" can cause an ascending sequence is 8/8 or 100% (i.e., 2, 3, 4, 5, 6, 7, 8, and 9), whereas such probability decreases to 50% if the number is "5". Put differently, if we were presented with a random number *x* and asked to bet on whether another randomly drawn number may be in ascending order, then, our bet would be framed by the distance from *x* to the upper bound, or the likelihood that *x* can cause an ascending sequence. This a priori probability is a by-product of the positional value of each numeral on the counting list. If X and Y are both discrete random and uniformly distributed with support {1,2,...,n}, then:

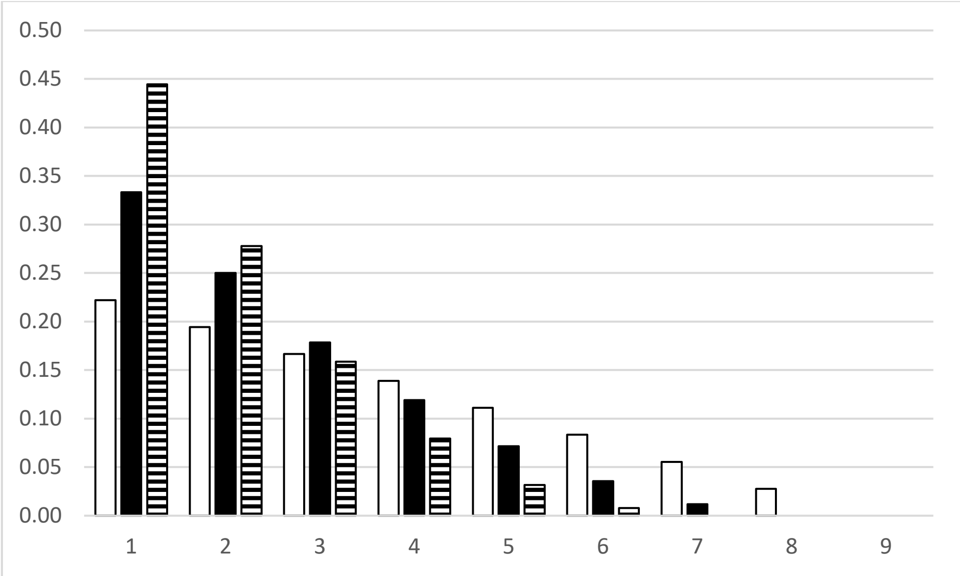
[2] In the current paper, counting sequence and counting list will be used indistinctly to refer to how numbers 1-to-9 are ordered.

$$P(X < Y \mid X \neq Y \text{ AND } X = x) \; = \frac{n-x}{n-1} \qquad \text{(Eq.1)}$$

Thus, the numerator reflects ordered outcomes for a number X (or distance to the upper bound) and the denominator reflects the total number of outcomes assuming that there is no replacement. For instance, if there are 9 numbered balls from 1 to 9 in a bag and we draw ball #4, the probability that the next ball that we pick up is in ascending order is well over 50% [.625= (9-4) / (9-1)]. If the second ball that is drawn is ball #7, then, the probability that a third ball that we pick up is in ascending order is substantially smaller ~ 29% [.286= (9-7) / (9-2)]. This means that each numeral in the counting sequence has a different probability associated with an ordered response and that the strength of such probability varies monotonically as a function of the position of a number in the counting sequence (numeral ordinality).

Note that this is simply a consequence of how the numerical system is operationalized. Indeed, the odds that a sequence of numbers is in ascending order are higher for smaller numbers. Figure 1 shows the skewness of the probability distribution of all possible combinations of ordered responses (ascending order) according to the first number and how the skewness of the distribution increases as a function of the number of elements in the sequence. For instance, whereas the probability that a pair of ordered numbers starts with 1 is 22%, that probability increases to 33% when it comes to three ordered numbers.

Figure 1: Probability distribution of ordered responses as a function of first number and elements in a list



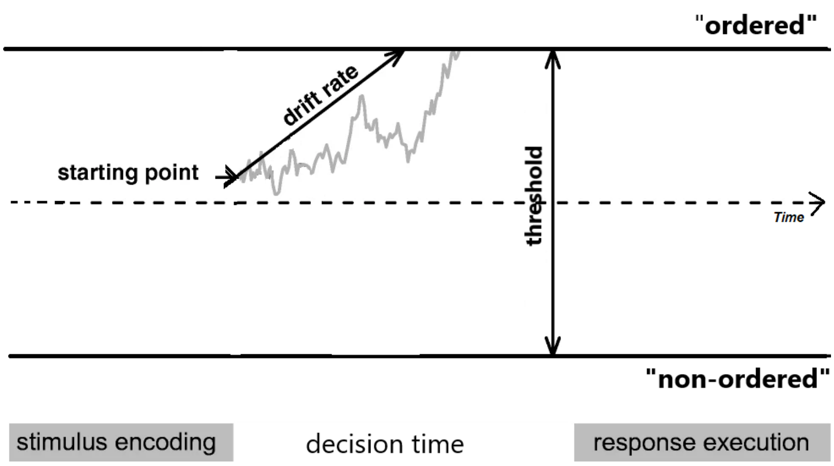
Note: Probability distribution of ordered 2-, 3-, and 4-digit combinations (white column, black column, and stripped column, respectively), without repetition, according to the first numeral.

On these grounds, it is feasible that positional values in the counting sequence and the corresponding a priori probabilities that are associated with an ordered response are activated when a numeral x in a triplet is processed in the context of numeral order judgments. Thus, the likelihood that a triplet with a first numeral *one* is ordered would be higher than that of a triplet with a first numeral *four* (i.e., [1-x-x] vs., [4-x-x]). Similarly, the likelihood that a triplet [2-4-x] is ordered would be higher than that of [2-7-x]. To test this hypothesis, we can investigate the latent processes involved in deciding whether a string of numerals is an ordered sequence. In other words, we can focus on the mechanisms underlying the differences (i.e., the decision-making process) rather than on the differences themselves.

One of the most popular models of decision-making is Ratcliff's (1978) diffusion model for 2-choice decisions, which has been recently used to uncover the processes involved in other numerical tasks (Iuculano et al., 2020; Park & Starns, 2015; Ratcliff, Thompson, & McKoon, 2015; Szardenings, Kuhn, Ranger, & Holling, 2017; Thompson, Ratcliff, & McKoon, 2016). In the diffusion model, the mechanism underlying two-choice decisions is the accumulation of noisy information from a stimulus over time. Information accumulates toward one or the other of two decision criteria until one of the criteria is reached; then, the response associated with that criterion is initiated (Ratcliff, 1978). In the numerical order judgment task with triplets, one of the criteria would be associated with an ordered response, the other with a non-ordered response.

This model is a prototypical example of an evidence-accumulation model and considers both response time and accuracy data. The model transforms these data into latent constructs (see Figure 2), hence, teasing apart different aspects involved in decision-making (encoding time, decision time, and execution time). These latent variables usually refer to four parameters: threshold (amount of information that needs to be accumulated for the decision), drift-rate (speed or rate of evidence-accumulation over time), non-decision time (representing stimulus encoding and motor response), and response-bias or starting point (quantifies the bias towards either of the response options). The drift-rate is commonly interpreted as ability or performance, and the threshold is usually referred to as response caution (Voss, Nagler, & Lerche, 2013).

Figure 2: Drift-diffusion model for 2-choice decisions



Assuming an unbiased observer and sequential processing, each numeral in a triplet would contribute to that accumulation of evidence. If the representation of numeral order in memory includes traces related to the operationalization of the numerical system, where the odds that a small number can cause an ascending sequence are higher, then, the rate of accumulation of evidence of an ordered response would depend on the position of the numerals in the counting sequence. The uptake of information would be faster for triplets consisting of numbers that are closer to the lower bound.

It is possible that the level of representation that is formulated here changes over development and varies as a function of one's knowledge and proficiency with the numerical system. In the context of cardinality processing, there is evidence that the *numerical distance effect* decreases over development and that both children and adults with better arithmetic skills also exhibit smaller numerical distance effects than those with poorer arithmetic proficiency (e.g., De Smedt, Verschaffel, & Ghesquiere, 2009), probably, because their numerical representations become more specific with experience. Similarly, the representation of the positions of numbers on the number line undergoes a refining process that spans the school years. For instance, Siegler and colleagues (Siegler & Booth, 2004; Siegler & Opfer, 2003) found that such representation shifted from a logarithmic pattern to a linear pattern over development (and reflected different degrees of numerical proficiency, Friso-van den Bos et al., 2015).

In the context of numeral ordinality, the a priori probabilities that are associated with an ordered response are a by-product of the positional value of each numeral in the counting list and how the numerical system is operationalized, hence, the odds that participants with a better understanding of the properties and operationalization of the numerical system access positional values (and related a priori probabilities of an ordered response) during numeral order judgment tasks are higher. These participants would have faster rates of accumulation of evidence of an ordered response. It is also feasible that the decision-making process involves having larger thresholds and non-decision times if that representation of how the numerical system is operationalized is weak. This means that more accumulation of evidence of an ordered response is needed—i.e., being more cautious before a response is triggered. Indeed, the differences found in the literature and the robust association of numeral ordering skills with arithmetic and math achievement in children also suggest an alternative possibility—that high- and low-achievers have different representations of numeral order and that the mechanisms involved in numeral order judgments are consequently different. Failing to activate such positional values (and prior expectations of an ordered response) upon processing each numeral in a triplet means that the uptake of information may relate to another type of evidence or that different mechanisms are activated.

To investigate this possibility and the hypothesis that the representation of numeral order in memory includes traces related to the operationalization and structure of the numerical system, we will pit two diffusion models against each other. In one of the models, we will specify the rate of accumulation of evidence of an ordered response as a function of the position of the numerals of each triplet in the counting sequence (i.e., small numerals would provide more evidence of an ordered response). In the alternative model, we will specify the accumulation of evidence as a function of the distance between the numerals in each triplet—i.e., the uptake of evidence would be faster for triplets presenting adjacent numbers. This reflects the possibility that symbol-symbol associations that are stored in memory explain the ability to judge the order of a sequence of numbers.

2. METHOD

2.1 Sample

Ninety-seven adult participants (64 female, all right-handed, M_{age} = 21.44 years, SD_{age} = 3.82, age-range = 19-39 years) took part in the experiment. All participants provided informed consent prior to their participation and received class-credit compensation. All procedures contributing to the study complied with the ethical standards of the relevant institutional committee on human experimentation and with the Helsinki Declaration.

2.2. Stimuli, procedure, and data preparation

Numeral order judgment task:

In the numeral order judgment task participants were asked to indicate as quickly as possible whether three numerals were in increasing order from left to right (for details on task procedure, see Goffin & Ansari, 2016). The arrangement of numerals in each trial aligned with other studies tackling numeral order, and included sequential (adjacent numerals), balanced (non-adjacent numerals and even distance between integers), and unbalanced stimuli (non-adjacent numerals and uneven distance between integers). There were seven possible sequential combinations and nine possible balanced combinations; each combination was presented four times in order (in-order trials) and out of order (non-ordered trials). Out of 68 possible unbalanced combinations 42 were selected (and presented twice, both in order and out of order) in such a way that the difference between numerical distances (first and second integers, and second and third integers) was not larger than two (e.g., [1-4-9]→ 3-5=2). This selection of unbalanced stimuli aimed at reducing differences with balanced and sequential triplets for which the difference between such numerical distances is zero (e.g., [2-4-6]→ 2-2=0). Therefore, there were 296 experimental trials and 10 practice trials (the full list of stimuli that were presented may be found in the Supplementary Material). Participants were presented with four blocks of 74 experimental trials each, which were administered in two separate sessions with a one-week gap between sessions and 20 minutes gap between blocks within the session. This was because we received feedback in a pilot study that the numeral order judgment task may be cognitively demanding and exhausting. In each block, 14 sequential, 18 balanced, and 42 unbalanced trials were presented. In each block (and across types of stimuli), half the trials showed the digits in increasing order. Response selection (key pressing) was counterbalanced for each participant across blocks.

In the analyses that follow, we used only ordered trials because responding to non-ordered trials probably involves other than just an accumulation of evidence of an ordered response until a response is triggered. Indeed, the response would be triggered when such accumulation is interrupted. Furthermore, there is robust evidence that magnitude-based processes are more likely in non-ordered trials. For instance, triplets with smaller ratios between the non-ordered numeral and the preceding numeral in the triplet are usually responded faster than those with ratios closer to one (i.e., [1-6-2] faster than [1-6-5]). A trimming procedure excluded all responses with RTs faster than 200 ms or slower than 3000ms; then, we discarded any trial with logarithmized RTs exceeding +- 3SDs of the mean on an intra-individual level (totaling 3.3% of ordered responses).

Arithmetic proficiency:

Arithmetic performance was measured using a task that included single-digit addition and subtraction. We selected this simple task to minimize (as much as possible) the role of domain-general skills (such as updating capacity) that may be involved in other tasks that are more complex and include multi-digit trials or present trials verbally. Stimuli included all possible combinations of single-digit integers excluding tie problems (e.g. 3 + 3; 6 – 3) and problems containing 0 or 1 as operands or answers. This arrangement resulted in a set of 56 problems per operation. Four practice trials displaying combinations not used as experimental trials (i.e., tie problems and problems containing 1 as an operand) were used to familiarize participants with the task. For subtraction, problems were formed by reversing the addition problems (i.e., the sum became the minuend and both operands became the subtrahend). Each operation was presented in separate blocks. The order of problems was randomized with the constraint that the same operand or answer did not appear in succession. If a problem was presented in the first half of the block (e.g., 3 + 5), its inverse problem was displayed in the second half (i.e., 5 + 3). Block order was counterbalanced across participants. Problems were presented horizontally in white characters on a black background. Each trial began with participants pressing the space bar key when a 1,000 ms ready signal ("***") appeared in the center of the screen. After 200 ms the problem appeared and remained on the screen until participants' verbal response. A microphone connected to a voice-activated relay, and interfaced with the computer, registered the latency of the responses. An experimenter registered the participants' responses. Participants were asked to perform both accurately and fast. The inverse efficiency score (IES; Townsend & Ashby, 1983)—a measure that combines RT and accuracy data—was used as the measure of arithmetic proficiency. The administration protocol of the study may be found in the Supplementary Material.

2.3. Statistical approach

In the current study, we used a measurement model for response time and accuracy data—Ratcliff's (1978) diffusion model for 2-choice decisions. The drift-diffusion model naturally accounts for the right-skew of response time (RT) distributions and the fact that the right-skew increases with difficulty. Given the small number of trials in numeral order judgment tasks, we estimated the parameters of the diffusion model with the Python package HDDM (v. 0.6.0; Wiecki, Sofer, & Frank, 2013; see description in Supplementary Material). In the current study, we focused on the rate of accumulation of evidence (drift-rate). Regarding the remaining parameters in the diffusion model, we assumed a symmetric starting point (not modeled), and we did not allow the threshold and non-decision time to vary across types of trials.

We first estimated three different models to investigate whether the decision mechanisms underlying numeral order judgments reflected that the representation of order in memory was based on symbol-symbol associations or, in contrast, it was based on the operationalization of the numerical system. In the first model (Model 1), the drift-rate was estimated independently of the type of trial. This model assumes that the accumulation of evidence is uniform across trials (i.e., numeral order judgments are independent of the three numerals that are presented in each triplet). In Model 2, we investigated whether the rate of accumulation of evidence of an ordered response was a function of the strength of symbol-symbol associations (i.e., adjacent numerals should provide more evidence of an ordered response). The numerical distance was calculated as the difference between the smaller and larger numeral (e.g., [2-3-4] = distance 2). The inter-integer distance ranged between 1 and 3. For each participant, we estimated five drift-rates that corresponded to distance-2 trials (sequential trials), distance-3, distance-4, distance-4(B) (balanced trials), and distance-5 trials. Distance-4(B) and distance-5 trials did not include adjacent numerals that could activate symbol-symbol associations. We also estimated two additional variations of Model 2 (Model 2b and 2c) to test the symbol-symbol association account at different levels (reflecting different degrees of association by combining triplets of different distances; see below). In Model 3, we investigated whether the rate of accumulation of evidence of an ordered response was a function of the position of the numerals of each triplet in the counting sequence (i.e., small numerals should provide more evidence of an ordered response). The first numeral in each triplet served as a proxy for that positional value. Trials starting with *one* have numerals that are closer to the lower bound, higher probability of an ordered response, and afford a larger number of 3-digit ordered combinations. We also estimated an additional variation of this model (Model 3b) that reflected more precisely the role of positional values on the accumulation of evidence of an ordered response.

Finally, we looked at whether arithmetic achievement was associated with different decision-making mechanisms by fitting the above-described models to data from low- and high-achievers. To determine the ability groups (or to optimize the cut-off points regarding arithmetic proficiency), we submitted the IES data to a cluster analysis that revealed four distinct groups with different means [$F(3, 93) = 359.73$] and similar variances [$F(3, 93) = .454$] (see Supplementary Material).

Across models, we used the same dataset. All observations for every participant were entered in the analysis, with the following constraints: we limited the analysis to ordered triplets starting with *one* to *five* because the models including the remaining trials did not converge (due to the smaller number of trials in ordered triplets starting with *six* and *seven*); we also limited the analysis to triplets for which the distance between the first and last numeral did not exceed five to achieve a homogeneous distribution of trials of different distances per positional value, and vice versa. We did not include [1-2-3] because this stimulus is known to trigger familiarity traces that might not directly involve ordinal processing (Sella, Sasanguie, & Reynvoet, 2020). The complete list of trials that were included in the analyses and the number of trials per condition and model may be found in Appendix A.

We used 20,000 samples and 2000 burn-in to estimate the parameters for each model. Model convergence was determined visually (investigating the trace, the autocorrelation, and the marginal posterior) as well as using the Gelman-Rubin statistic ($\hat{R}$) that compares within-chain to between-chain variability (Gelman & Rubin, 1992). To assess convergence, we ran five chains (values below 1.1 indicate successful converge). Model comparison was based on deviance information criterion values (DIC) (Spiegelhalter et al., 2002). A DIC difference of 10 is strong evidence for a model (Kass & Raftery, 1995). We also checked the posterior predictive values of the best-fit model to confirm that the model predictions are in line with the observations (see Supplementary Material). To investigate whether the drift-rates differed between conditions, we used Bayesian hypothesis-testing and calculated the proportion of overlap between the posterior distributions.

3. RESULTS

Descriptive analyses

Table 1 shows the mean RTs (and corresponding accuracy in the numeral order judgment task) of individual median RTs for correct and error responses per type of condition (and model that was estimated). Table 2 shows descriptive statistics regarding performance on the arithmetic task.

Table 1: Descriptive statistics for response time data in milliseconds (and the proportion of correct and error responses and SD) for the models that were investigated

	Model 1		Model 2 (Distance)										Model 3 (Digit 1)									
			Two (S)		Three		Four		Four (B)		Five		One		Two		Three		Four		Five	
	E	C	E	C	E	C	E	C	E	C	E	C	E	C	E	C	E	C	E	C	E	C
%	0.03	0.97	0.02	0.98	0.03	0.97	0.04	0.96	0.04	0.96	0.04	0.96	0.01	0.99	0.02	0.98	0.02	0.98	0.05	0.95	0.06	0.94
%SD	-	2.70	-	4.02	-	4.23	-	4.66	-	4.87	-	4.75	-	2.55	-	3.10	-	3.41	-	6.60	-	6.33
M	1088	916	1152	886	1166	941	1063	935	990	908	1090	927	1013	791	1035	901	1092	929	1090	983	1083	1003
SD	375	184	468	176	389	205	354	198	277	187	483	210	269	153	296	194	365	203	436	197	378	201
Min	621	588	740	552	584	570	578	566	596	563	437	579	461	507	523	582	629	575	453	598	596	591
Max	2873	1505	2563	1340	2607	1506	2177	1587	2048	1380	2883	1703	1482	1198	1736	1548	2646	1644	2873	1488	2607	1500

Note. E: error; C: correct. (S) stands for sequential trials (e.g., [2-3-4]) and (B) for balanced trials (e.g., [2-4-6]).

Table 2: Descriptive statistics (IES: inverse efficiency score; RT; and accuracy) in the arithmetic task (by ability group)

	IES					RT					Accuracy				
	Overall	Group 1	Group 2	Group 3	Group 4	Overall	Group 1	Group 2	Group 3	Group 4	Overall	Group 1	Group 2	Group 3	Group 4
n	97	10	21	31	35	97	10	21	31	35	97	10	21	31	35
M	2085	3389	2690	2064	1368	1960	3123	2509	1940	1316	0.95	0.92	0.93	0.94	0.96
SD	695	165	198	189	216	636	279	231	208	204	0.05	0.05	0.05	0.04	0.04
Min	954	3185	2426	1745	954	909	2611	1997	1591	909	0.81	0.82	0.81	0.86	0.82
Max	3689	3689	3081	2384	1680	3620	3620	2971	2361	1634	1.00	0.99	0.99	1.00	1.00

In line with other studies with adults, it can be observed that the percentage of errors in the numeral order judgment task was small (totaling 3% of the data). This percentage increased as a function of the positional value of the numerals in each triplet (see columns corresponding to Model 3). It can also be observed that the pattern of RTs also aligned with that of accuracy (slower RTs as number size increased), which suggests that more accumulation of evidence is needed to trigger a response. In contrast, the pattern of RTs and errors in Model 2 was noisier and did not align with the predictions of the symbol-symbol account. Faster responses were observed in triplets without symbol-symbol associations (distance-4(B) and distance-5 triplets). The corresponding response-accuracy function may be found in Appendix B.

Drift-diffusion model

Across models, for each participant i, we estimated the posterior distributions of three different parameters: drift-rate, threshold, and non-decision time. All of the models converged successfully (see HDDM estimation and convergence information in Supplementary Material). Parameter estimates and the corresponding DIC values of the models that were estimated are shown in Table 3.

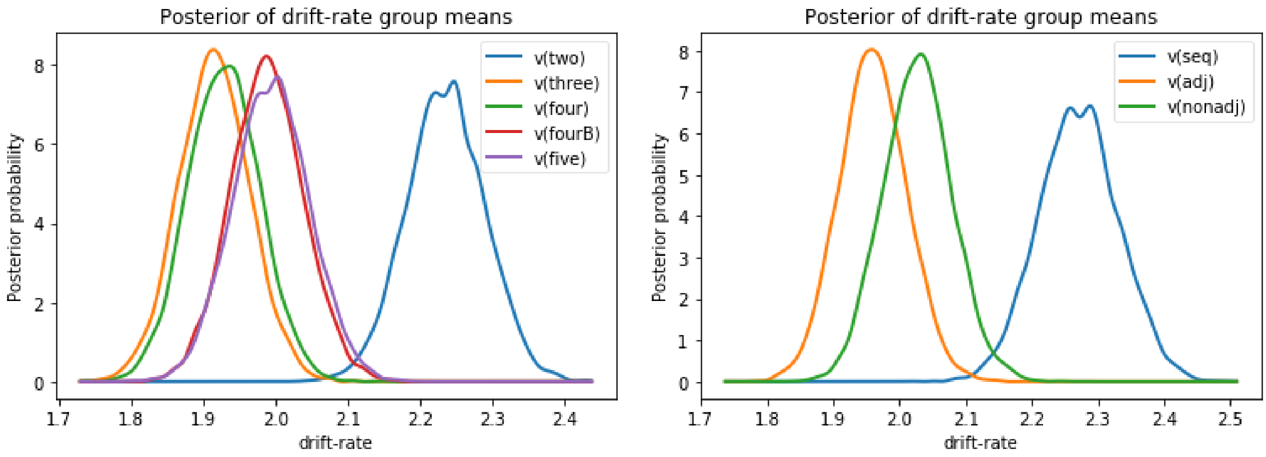
Table 3: Parameter estimates and DIC values

Model	Drift-rate	Threshold	Non-decision time	DIC
Model 1	2.07	2.12	.47	2206
Model 2 (Distance)	-	2.09	.47	2397
Two	2.23	-	-	-
Three	1.92	-	-	-
Four	1.93	-	-	-
Four (B)	1.99	-	-	-
Five	1.99	-	-	-
Model 2b (Distance)	-	2.11	.47	2271
Sequential	2.27	-	-	-
Adjacent	1.96	-	-	-

Non-adjacent	2.03	-	-	-
Model 2c (Distance)	-	2.12	.47	2196
Sequential	2.29	-	-	-
Non-sequential	2.02	-	-	-
Model 3 (Digit 1)	-	2.16	.47	1821
One	2.87	-	-	-
Two	2.25	-	-	-
Three	2.12	-	-	-
Four	1.76	-	-	-
Five	1.68	-	-	-
Model 3b (Number size)	-	2.16	.47	1781
One	2.76	-	-	-
Two	2.34	-	-	-
Three	2.09	-	-	-
Four	1.86	-	-	-
Five	1.55	-	-	-

Figure 3 (left panel) shows the Kernel density estimate of the posterior distribution for each drift-rate in Model 2. It can be observed that participants' rate of accumulation of information was faster on distance-2 trials (highest degree of symbol-symbol associations), compared to that of triplets that showed larger inter-integer distances (lowest degree of symbol-symbol associations). Nonetheless, the posterior distributions of non-sequential triplets showing adjacent numerals (i.e., distance-3 and distance-4 trials) as well as that of triplets showing non-adjacent numerals (distance-4(B) and distance-5 trials) overlapped substantially. Furthermore, the DIC value was larger than that of Model 1, hence, suggesting a worse fit.

Figure 3: Kernel densities of posterior distributions of drift-rates of Model 2 (left) and 2b (right)



We formulated another model (Model 2b) that included only three drift-rates that corresponded to three different degrees of association between numerals: sequential triplets (distance-2 triplets), adjacent (distance-3 and distance-4 triplets), and non-adjacent triplets (distance-4(B) and distance-5 triplets). The DIC value suggested that this model fitted the data better than Model 2. Nonetheless, the DIC value of Model 1 was still smaller. Figure 3 (right panel) shows the Kernel density estimate of the posterior distribution for each drift-rate that was estimated in Model 2b. It can be observed substantial overlap in the posterior distribution of the drift-rate of non-sequential triplets showing non-adjacent numbers and that of non-sequential triplets showing adjacent numbers (p_{Bayes} [*non-adj* > *adj*] = .16) [3]. Note that this also suggests that magnitude-based mechanisms are not involved in numeral order judgments of ordered triplets.

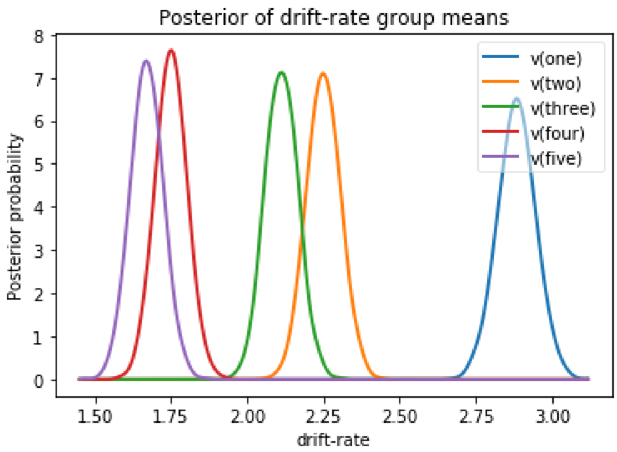
It might be argued that processing non-sequential triplets that have a pair of adjacent numerals does not involve associative mechanisms in the sense that these trials trigger weaker symbol-symbol associations. Thus, we formulated an additional model (Model 2c) to investigate whether the accumulation of evidence of an ordered response aligned with the *reverse distance effect*. In this model, a single drift-rate was modeled across non-sequential trials and a different drift-

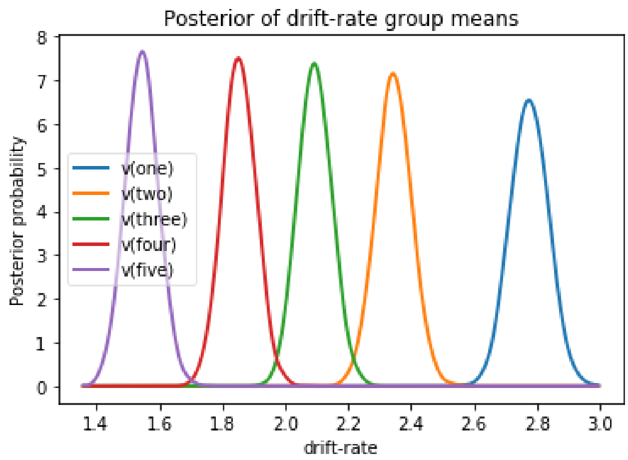
[3] Bayesian p values quantify the degree to which the difference in the posterior distribution is consistent with the hypothesis that, for instance, the parameter is greater for *Non-adjacent* than *Adjacent* triplets. Bayesian p of .05 indicates that 95% of the posterior distribution supports the hypothesis.

rate was modeled for sequential trials. This model would reflect more precisely the hypothesis that the mechanisms underlying numeral order judgments are based on symbol-symbol associations. The DIC value indicated that this model fitted the data better than Model 2b. Assessment of model parameters showed that the drift-rate of sequential triplets was larger/faster, (p_{Bayes} [*seq* > *non-seq*] < .001), which aligns with the *reverse distance effect* that has been reported in the literature. Nonetheless, the DIC value was not substantially different to that of Model 1, which was a more parsimonious model.

In Model 3, the DIC value was substantially smaller than that of Model 1, suggesting a better fit to the data. Assessment of the drift-rates showed that triplets with numerals closer to the lower bound of the counting sequence had faster rates of information uptake. The kernel density estimates showed some overlap between some configurations (see Figure 4, left panel). This is probably because the arrangement of stimuli was based on the first numeral in each triplet, which provides a raw estimation of the positional value of the three numerals in each triplet (and the corresponding probability of an ordered response). For instance, there were overlaps across the five groups of triplets with regards to the range of number size in each triplet—(digit *One* [7-11], digit *Two* [9-14]; digit *Three* [12-17]; digit *Four* [15-20]; digit *Five* [18-21]). Given that the number size of the numerals in each triplet may summarize more precisely the positional value of the numerals, we formulated another model (Model 3b) where the triplets were arranged in five groups that did not overlap in terms of number size (while maintaining the same distribution of distances per group; see Appendix A). The DIC value of this model was smaller than that of Model 3. Figure 4 (right panel) shows the Kernel density estimate of the posterior distribution for each drift-rate in Model 3b.

Figure 4: Kernel densities of posterior distributions of drift-rates of Model 3 (left) and 3b (right)





Finally, we investigated potential differences in the mechanisms underpinning numeral order judgments in low and high arithmetic achievers (groups 1 and 4 in Table 2). DIC values of the equivalent Model 1, Model 2c, and Model 3b are shown in Table 3 (top panel). The best fit model in both groups was that in which the rate of accumulation of evidence of an ordered response was modeled as a function of the position of the numerals in the counting sequence (Model 3b).

Table 3: DIC values of Model 1, 2c, and 3b for Low and High achievers (top panel) and parameter estimates of Model 3b per ability group (bottom panel)

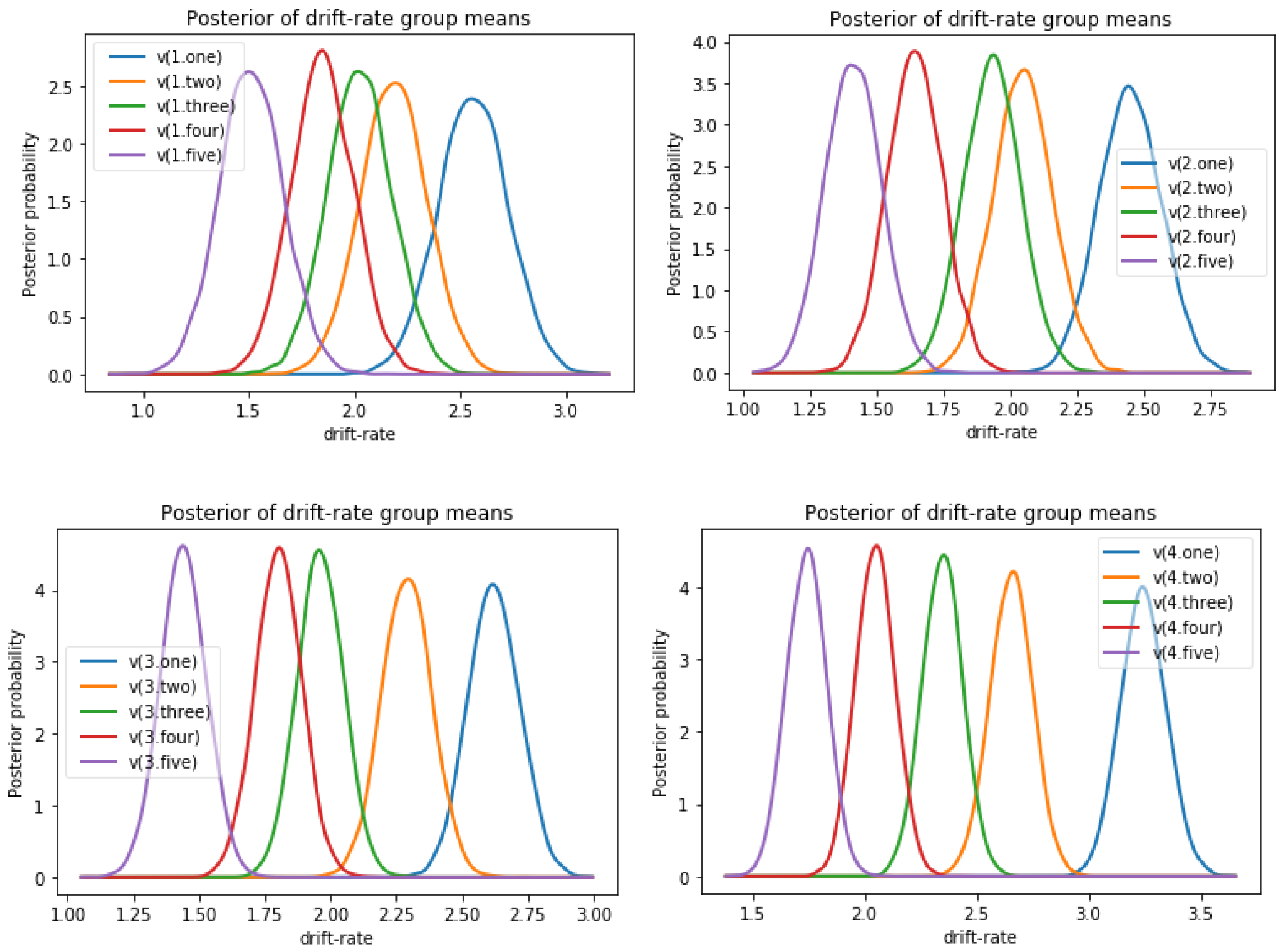
Model	**DIC**			
	Group 1 (Low achievers)	**Group 2**	**Group3**	**Group 4 (High achievers)**
Model 1	560	-	-	-596
Model 2c	558	-	-	-622
Model 3b	531	-	-	-809
Threshold	2.56	2.23	2.18	1.97
Non-response time	.47	.52	.48	.44
Drift-rate	-	-	-	-
One	2.55	2.44	2.61	3.23
Two	2.18	2.05	2.29	2.65
Three	2.03	1.93	1.96	2.35
Four	1.85	1.65	1.81	2.06
Five	1.52	1.42	1.45	1.76

Since the same mechanisms were found in low- and high-achievers, we modeled data from all of the groups together to explore how the parameters of the best fit model varied as a function of arithmetic ability[4]. In this model, for each participant i in group g (1-to-4), we estimated his/her threshold, non-decision time, and five different drift-rates. Parameter estimates may be found in Table 3 (bottom panel). Across groups, trials that provided more evidence of an ordered response (numerals closer to the lower bound) were associated with faster information uptake. The kernel densities of posterior distributions of drift-rates per ability group are shown in Figure 5. Assessment of model parameters showed that high-achievers (Group 4) accumulated evidence at a faster rate, independently of the position of the numerals on the counting list. The proportion of overlap between the posterior distribution of each drift-rate in Group 4 and the corresponding posteriors of Group 3, 2, and 1, ranged between 0% and 4% (all $p_{Bayes} < .05$). Nonetheless, we found substantial overlap in the posterior distribution of each drift-rate across groups 3, 2, and 1. For instance, the drift-rate corresponding to trials that provided more evidence of an ordered response (“*One”* triplets) was similar across these groups (the proportion of overlapped ranged between 20% and 50%).

A closer look at the posterior distributions of the drift-rates of each ability group revealed that the proportion of overlap decreased as a function of arithmetic ability. For instance, whereas the overlap between the posterior distributions of the drift-rates in Group 4 (high achievers) approached 0% (all $p_{Bayes} < .001$; see Figure 5, bottom right panel), there was a slight overlap in Group 3 (p_{Bayes} [*Three* > *Four*] = .09), some overlap in Group 2 (p_{Bayes} [*Two* > *Three*] > .05 and p_{Bayes} [*Four* > *Five*] = .06), and substantial overlap in the posterior distributions of Group 1 (see Figure 5, top left panel).

[4] There were convergence problems leading to difficulty in confidently estimating the effect in each individual participant when the drift-rates were regressed on the IE scores (i.e., arithmetic as a continuous predictor of the drift-rate). This was likely due to the limited number of trials for each participant and condition (trial type) and arithmetic scores.

Figure 5: Kernel densities of posterior distributions of drift-rates of Model 3b per ability group



Note: From left to right and top to bottom (Group 1, 2, 3, and 4).

The threshold or amount of evidence that was needed to decide whether a sequence of three numerals was in order showed a similar pattern—i.e., smaller thresholds as arithmetic ability increased (p_{Bayes} [*Group_4* < *Group_3*] = .03, p_{Bayes} [*Group_3* < *Group_2*] > .05, p_{Bayes} [*Group_2* < *Group_1*] = .04). No clear evidence of the association between arithmetic ability and the non-decision time parameter (stimulus encoding) was found.

3. DISCUSSION

In the current study, we formulated a drift-diffusion model to investigate the mechanisms underlying decision-making in the numeral order judgment task with triplets. In particular, we looked at whether the accumulation of evidence of an ordered response (or numeral order ability) reflected a representation of numeral order based on symbol-symbol associations or, in contrast, such ability responded to the operationalization of the numerical system. Although it is feasible that connections among symbols create memory traces that benefit the settlement of number symbols as an ordered sequence in long-term memory (Lyons & Beilock, 2009, 2013), our findings suggest that other traces related to the operationalization of the numerical system (and the ordinality of numbers) may be stored in parallel and characterize more precisely the ability underlying numeral order judgments.

Numeral ordinality and numeral order judgments

The possibility that different levels of representation are interconnected and affect recognition memory has received support from a variety of studies and theoretical accounts. For instance, there is evidence that, even when a "pure" ordering process is required (e.g., participants are tasked to reorder a collection of given elements according to a previously studied sequence), a parallel activation (i.e., "automatic process") of item information occurs (Neath, 1997). Thus, it is feasible that the positional values of the numerals in each triplet are activated during numeral order judgments. These positional values or ordinal representations would reflect the operationalization of the numerical system where small numbers provide more evidence of an ordered response—orderliness. Since item (symbol) and order (position) are isomorphic, each number symbol would be linked to the corresponding a priori probabilities of an ordered response in the representation of order in memory.

Hypothetically, any connectionist model of numerical representation with a continuum architecture where each number is a node or unit would account for this representation of numeral ordinality. The ordinal representation of each number would relate to the propagation of activation towards ascending numbers (or elements that are located towards the upper bound). Thus, the spread of that activation (i.e., the strength of ordinal representations) would vary monotonically as a function of the position of a number on the counting sequence. For instance, the ordinal representation of *one* would be linked to all of the numbers in the counting sequence whereas that of *five* would be linked to nodes *six, seven eight, and nine.* In the context of numeral order judgments, 1-x would trigger a stronger ordinal representation than 5-x. This would reflect the hierarchical nature of ordinal representations. Arguably, this propagation of activation may be considered a sort of item-item association since, for instance, 1-x would activate nodes *two* to *nine*. Nonetheless, and in contrast to the nature of the symbol-symbol association that is assumed in the literature of numeral ordinality, the propagation of activation across nodes would not vary as a function of the distance between nodes. This is because the likelihood that any of the nodes that are activated is in order is uniform across nodes (e.g., both #9 and #2 have identical status as ordered outcomes upon processing a first numeral #1 in the context of numeral order judgments, 1-x).

This representation of numeral ordinality is consistent with most of the findings that have been reported in the context of numeral order judgments. For instance, it would account for differences between ascending and descending order or why processing descending order is more challenging (and probably involves different mechanisms to those outlined here). It would explain why non-sequential trials may be responded faster than some sequential trials (e.g., [2-4-6] faster than [4-5-6]), and why the *reverse distance effect* is not universal. It would also explain why in the context of 2-digit order judgments a *numerical distance effect* has been reported. For instance, Turconi et al. (2006), found no differences in RTs between distance-4 pairs (e.g., [2-6]) and sequential pairs.

Note that the findings of the current study can also be explained according to magnitude-based models that have been formulated to explain how order (in non-numeric contexts) is encoded in memory. Botvinick and Watanabe (2007) formulated a neuro-computational model of memory for serial order in which representation of order in memory was based on rank information. Based on behavioral and neuroscience evidence, it is argued that encoding the order of a sequence of elements would involve a conjunctive code that integrates items with order (rank) information. This model is a prototype of magnitude-based ordering and assumes a compressive representation where the strength or activation of ordinal information decreases as a function of the position of an item on the sequence that is encoded. Indeed, it explicitly links such encoding to the activation of neural areas that are usually engaged in magnitude processing (Nieder, 2012; for a review, see Dehaene, Meyniel, Wacongne, Wang, & Pallier, 2015). Thus, when participants are tasked to memorize a novel list of items, the order of an item in that list would be defined by abstract ordinal information (presumably conveyed by IPS neurons). Although this model aims at explaining the mechanisms underlying broader ordering abilities, it also predicts that ordinal representations of small numbers in the counting sequence are "stronger".

Indeed, the pattern of findings cannot be fully distinguished from that predicted by the activation of the mental number line—a model of numerical magnitude representation that assumes that each number magnitude is represented by a distribution of activation on a logarithm compressed mental number line, and predicts poorer performances as a function of number magnitude/size (Piazza et al., 2004; Van Oeffelen & Vos, 1982). Larger numbers would activate noisier (overlapping) representations that would constraint the identification of whether a numeral X precedes a numeral Y in the counting sequence (the ratio of the magnitude of X and Y is closer to 1 for numerals closer to the upper bound). Note that the average ratio of adjacent numerals in an ordered triplet XYZ highly correlates with the positional value of the first numeral that has been investigated in the current study (r =. 89). Nonetheless, we are hesitant to suggest that such activation of the number line occurs during numeral order judgments and reflects differences in the drift rates between triplets showing small numerals and those showing numerals closer to the upper bound. We found that the rate of accumulation of evidence for triplets showing adjacent numerals—i.e., those showing more representational overlap on the mental number line—was similar to that of triplets without adjacent numerals (for a similar argument, see Sella et al., 2020). Although some studies have reported behavioral effects that are compatible with the activation of magnitude-based mechanisms in non-ordered trials (Lyons & Beilock, 2013; Morsanyi et al., 2017; Voss et al., 2017), more evidence is needed regarding the conditions under which these effects are observed.

It might be argued that familiarity with small numerals rather than the operationalization of the numerical system underpins the pattern of findings reported in this study. For instance, Verguts and van Opstal (2005) showed that the *numerical size effect* in symbolic comparison of small numbers is the result of the skewed frequency distribution of numbers, with smaller numbers being more frequent than larger numbers (see also, Krajcsi, Lengyel, & Kojouharova; 2016; Verguts & van Opstal, 2014). Indeed, a frequency-based account would render similar predictions to those formulated in the current study, and the pattern of results would be hardly discernible. These arguments stem, mainly, from Dehaene and Mehler's (1992) work on the frequency of numbers. These authors specifically tackled this question and explored whether the fact that small numerals are more frequent than large numerals is due to environmental circumstances (as smaller numerosities are more frequent in a natural environment), notational regularities in the numerical system (as the distribution of frequencies of first numbers in natural data follows Benford's law; Benford, 1983), or to the particularities of our numerical representation system and its capacity to process numerical information from the environment. They found that the frequency with which one encounters written numerical symbols correlated with numerical size in an inverse power-law manner (both cardinal and ordinal numbers). Although this work is usually cited as supporting evidence of the frequency-based account, the conclusions of this work are usually neglected—"The frequency of numerals can be used as an indicator of the mental organization of number concepts" (Dehaene & Mehler, 1992, p. 20; for a similar argument, see Brysbaert, 1995). Put differently, it is not that the organization of the numerical system stems from the frequency with which numbers are encountered in real life but rather the opposite. Indeed, the *reverse distance effect*, which would reflect such familiarity traces during numeral order judgments (i.e., sequential trials have higher co-occurrence in everyday language), is not universal (e.g., Morsanyi et al., 2017; Vogel, Remark, & Ansari, 2015). Furthermore, it is unlikely that the frequency of small numbers may affect decisions that rely exclusively on the structure of the numerical system. As mentioned in the Introduction section, if we were presented with a random number x and asked to bet on whether another randomly drawn number may be in ascending order, then, our decision would not depend on the frequency with which numbers are encountered in life.

Numeral ordering skills and arithmetic achievement

In the current study, we also investigated how the processes involved in decision-making during numeral order judgments related to arithmetic proficiency. We found that the best-fit model in high- and low-achievers was that in which the uptake of information was a function of the position of the numerals in the counting sequence. The analyses also revealed that high achievers accumulated evidence at a faster rate (independently of the positional values that were activated) and that they were less cautious during numeral order judgments, however, no clear evidence of a linear association between that accumulation of evidence and arithmetic proficiency was found. Note that arithmetic ability was treated as a categorical "predictor" of the drift-rate and threshold, hence, it does not involve the same level of heterogeneity that might emerge with continuous measures. In any case, the pattern of findings did reveal that the accuracy of the positional representations or the effect that the operationalization of the numerical system may have on numeral order judgments became more evident and specific as arithmetic ability increased.

These positional values are essential for the understanding of the properties of the numerical system—in particular, for proportional reasoning. Participants with a better understanding of the structure of the numerical system (and more accurate representations of the positional values of symbols) would be able to navigate the numerical system more efficiently and disentangle item and order information (if needed). That is, knowing that i) the odds that the next ball that is picked up is in ascending order are 50% after drawing ball #5 from a bag with numbered balls #1 to #9, ii) that such odds decrease substantially if the balls in the bag are balls #1 to #6, or iii) that such odds are well over 50% if the balls in the bag are balls #4 to #8. Indeed, it is feasible that the same processes that contribute to the refinement of the positions of numbers on the number line also contribute to strengthening the link between each numeral in the counting sequence and the corresponding positional value in that sequence. Although the number line is usually linked to the representation of numerical magnitudes, the number line is a system that naturally represents positional and ordinal information. Furthermore, there is evidence that number line estimation skills require proportional reasoning (Barth & Paladino, 2011; Slusser, Santiago, & Barth, 2013) and some studies have found that numeral ordering skills share a larger amount of variance with number line estimation skills than with other numeric skills (Morsanyi et al., 2018).

Whilst proportional reasoning is foundational to understanding complex concepts such as fractions (Boyer & Levine, 2012), it is also involved in the understanding of part-whole relations that are relevant for basic arithmetic (Baroody, Ginsburg, & Waxman, 1983). The possibility that numeral order skills may reflect this higher-level structural knowledge of the numerical system has been suggested by some studies that have reported that the effect of magnitude processing on arithmetic/mathematics skills is mediated by order processing skills (Lyons & Beilock, 2011; Morsanyi, van Bers, O'Connor, & McCormack, 2020; Sasanguie, Lyons, De Smedt, & Reynvoet, 2017).

Theoretical and practical implications for future studies

Given that our findings suggest that numeral ordering skills may be better captured by measures that consider how performance varies according to the position of the numerals of each triplet in the counting sequence, future studies should investigate the adequacy of extant task parameters. For instance, i) using trials that reduce the possibility that symbol-symbol associations are activated (and consequently affect numeral order judgments), ii) providing a balanced distribution of trials according to the position of the numerals in the counting list, and iii) computing numeral ordering skills according to the strength of the ordinal representations that are hypothetically activated—i.e., our findings suggest that the performance pattern may reflect a linear association between behavioral data and the average position of the numerals in the counting sequence. Nonetheless, more research is needed to investigate the discrimination properties of each trial. Alternative tasks where item and order information are dissociated or where the ordinality of each number is explicitly activated may also provide valuable information about numeral order abilities and the relevance of numeral ordinality.

Although the mechanisms underpinning numeral order judgments as well as the representations that have been formulated in the current study are plausible, they are also speculative and warrant further investigation. More direct evidence of the activation of positional values during numeral order judgments is needed to formulate an adequate neural and computational model. It is also important to adapt this hypothetical representation of the ordinality of numbers to extant frameworks on the nature of number representations. Similarly, more research is needed to disentangle how the ordinal representations that are suggested here are activated in non-numeric contexts. Whilst these traces are probably stronger in the counting sequence than in other non-numeric stimuli where item and order information are not isomorphic, similar effects may be found when it comes to familiar sequences (e.g., days of the week, months, alphabet). Encoding the order of the days of the week and the months in a year also involves assigning positional values that may inform about the likelihood that an element may cause an ascending sequence. In other words, if we have 12 balls in a bag (representing the months January to December) and the first ball that is drawn reads July, then, the probability that the next ball corresponds to a month between August and December is smaller than that of picking up a ball corresponding to months January to June. Indeed, there is evidence that non-numeric ordering skills are associated with other skills that involve proportional reasoning, such as number line estimation (Morsanyi et al., 2017).

Finally, future studies should also explore the developmental implications of the rationale outlined here. Given that children's numerical representations change over time, and younger children may have limited experience with the numerical system, a different pattern of results (and mechanisms) may be expected. For instance, Lyons et al. (2014) found that from around Grade 6, numerical order processing was the best predictor of children's arithmetic performance, which suggests that the operationalization of the numerical system is not yet mature in younger children and that the representation of numeral ordinality that has been suggested here is not probably developed.

Acknowledgments: This work was supported by Ministry of Economy and Enterprise in Spain to Josetxu Orrantia [Grant number PSI2015-66802-P]

4. REFERENCES

Attout, L., & Majerus, S. (2015). Working memory deficits in developmental dyscalculia: The importance of serial order. *Child Neuropsychology*, *21*(4), 432-450.

Baroody, A. J., Ginsburg, H. P., & Waxman, B. (1983). Children's use of mathematical structure. *Journal for Research in Mathematics Education, 14*(3), 156-168.

Barth, H. C., & Paladino, A. M. (2011). The development of numerical estimation: Evidence against a representational shift. *Developmental Science, 14*(1), 125-135. doi:10.1111/j.1467-7687.2010.00962.x

Benford, F. (1938) The law of anomalous numbers. *Proc. Amer. Philos. Soc.78*, 551{572

Botvinick, M., & Watanabe, T. (2007). From numerosity to ordinal rank: a gain-field model of serial order representation in cortical working memory. *Journal of Neuroscience, 27*(32), 8636-8642.

Boyer, T. W., & Levine, S. C. (2012). Child proportional scaling: Is 1/3= 2/6= 3/9= 4/12? *Journal of Experimental Child Psychology, 111*(3), 516-533.

Brysbaert, M. (1995). Arabic number reading: On the nature of the numerical scale and the origin of phonological recoding. *Journal of Experimental Psychology: General*, *124*(4), 434.

Caplan, J. B. (2015). Order-memory and association-memory. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale, 69*(3), 221.

Colomé, À., & Noël, M. P. (2012). One first? Acquisition of the cardinal and ordinal uses of numbers in preschoolers. *Journal of Experimental Child Psychology, 113*(2), 233-247.

De Smedt, B., Verschaffel, L., & Ghesquière, P. (2009). The predictive value of numerical magnitude comparison for individual differences in mathematics achievement. *Journal of Experimental Child Psychology*, *103*(4), 469-479.

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

Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., & Pallier, C. (2015). The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees. *Neuron, 88*(1), 2-19.

Franklin, M. S., Jonides, J., & Smith, E. E. (2009). Processing of order information for numbers and months. *Memory & Cognition, 37*(5), 644-654.

Friso-van den Bos, I., Kroesbergen, E. H., Van Luit, J. E., Xenidou-Dervou, I., Jonkman, L. M., Van der Schoot, M., & Van Lieshout, E. C. (2015). Longitudinal development of number line estimation and mathematics performance in primary school children. *Journal of Experimental Child Psychology, 134*, 12-29.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 7*(4), 457-472.

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.

Iuculano, T., Padmanabhan, A., Chen, L., Nicholas, J., Mitsven, S., de Los Angeles, C., & Menon, V. (2020). Neural correlates of cognitive variability in childhood autism and relation to heterogeneity in decision-making dynamics. *Developmental Cognitive Neuroscience,* 100754.

Jou, J. (2003). Multiple number and letter comparison: Directionality and accessibility in numeric and alphabetic memories. *American Journal of Psychology*, *116*(4), 543-580.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773-795.

Krajcsi, A., Lengyel, G., & Kojouharova, P. (2016). The source of the symbolic numerical distance and size effects. *Frontiers in Psychology, 7*, 1795.

Leth-Steensen, C., & Marley, A. A. J. (2000). A model of response time effects in symbolic comparison. *Psychological Review*, *107*(1), 62.

Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review, 91*(3), 295.

Lyons, I. M., & Ansari, D. (2015). Foundations of children's numerical and mathematical skills: the roles of symbolic and nonsymbolic representations of numerical magnitude. In *Advances in Child Development and Behavior* (Vol. 48, pp. 93-116). JAI.

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., 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.

Marshuetz, C., & Smith, E. E. (2006). Working memory for order information: multiple cognitive and neural mechanisms. *Neuroscience*, *139*(1), 195-200.

Matejko, A. A., Hutchison, J. E., & Ansari, D. (2019). Developmental specialization of the left intraparietal sulcus for symbolic ordinal processing. *Cortex, 114*, 41-53.

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.

Morsanyi, K., van Bers, B. M., 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.

Morsanyi, K., van Bers, B. M., O'Connor, P. A., & McCormack, T. (2020). The role of numerical and non-numerical ordering abilities in mathematics and reading in middle childhood. *Contemporary Educational Psychology, 62*, 101895.

Neath, I. (1997). Modality, concreteness, and set-size effects in a free reconstruction of order task. *Memory & Cognition, 25*(2), 256-263.

Nieder, A. (2012). Supramodal numerosity selectivity of neurons in primate prefrontal and posterior parietal cortices. *Proceedings of the National Academy of Sciences, 109*(29), 11860-11865.

O'Connor, P. A., Morsanyi, K., & McCormack, T. (2018). Young children's non-numerical ordering ability at the start of formal education longitudinally predicts their symbolic number skills and academic achievement in maths. *Developmental Science, 21*(5), e12645.

Orrantia, J., Muñez, D., Matilla, L., Sanchez, R., San Romualdo, S., & Verschaffel, L. (2019). Disentangling the Mechanisms of Symbolic Number Processing in Adults' Mathematics and Arithmetic Achievement. *Cognitive Science*, *43*(1).

Park, J., & Starns, J. J. (2015). The approximate number system acuity redefined: A diffusion model approach. *Frontiers in Psychology, 6*, 1955.

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron*, *44*(3), 547-555.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*(2), 59.

Ratcliff, R., Thompson, C. A., & McKoon, G. (2015). Modeling individual differences in response time and accuracy in numeracy. *Cognition, 137*, 115-136.

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

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.

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.

Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development, 75*(2), 428-444.

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science, 14*(3), 237-250.

Slusser, E. B., Santiago, R. T., & Barth, H. C. (2013). Developmental change in numerical estimation. *Journal of Experimental Psychology: General, 142*, 193-208. doi:10.1037/a0028560

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series b (statistical methodology), 64*(4), 583-639.

Szardenings, C., Kuhn, J. T., Ranger, J., & Holling, H. (2018). A diffusion model analysis of magnitude comparison in children with and without dyscalculia: Care of response and ability are related to both mathematical achievement and stimuli. *Frontiers in Psychology, 8*, 1615.

Thompson, C. A., Ratcliff, R., & McKoon, G. (2016). Individual differences in the components of children's and adults' information processing for simple symbolic and non-symbolic numeric decisions. *Journal of Experimental Child Psychology, 150*, 48-71.

Townsend, J. T., & Ashby, F. G. (1983). Stochastic modeling of elementary psychological processes. CUP Archive.

Turconi, E., & Seron, X. (2002). Dissociation between order and quantity meaning in a patient with Gerstmann syndrome. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*.

Turconi, E., Campbell, J. I., & Seron, X. (2006). Numerical order and quantity processing in number comparison. *Cognition, 98*(3), 273-285.

Turconi, E., Jemel, B., Rossion, B., & Seron, X. (2004). Electrophysiological evidence for differential processing of numerical quantity and order in humans. *Cognitive Brain Research, 21*(1), 22-38.

van Dijck, J. P., & Fias, W. (2011). A working memory account for spatial–numerical associations. *Cognition, 119*(1), 114-119.

Van Oeffelen, M. P., & Vos, P. G. (1982). A probabilistic model for the discrimination of visual number. *Perception & Psychophysics*, *32*(2), 163-170.

Verguts, T., & Van Opstal, F. (2005). Dissociation of the distance effect and size effect in one-digit numbers. *Psychonomic Bulletin & Review, 12*(5), 925-930.

Verguts, T., & Van Opstal, F. (2014). A delta-rule model of numerical and non-numerical order processing. *Journal of Experimental Psychology: Human Perception and Performance, 40*(3), 1092.

Vogel, S. E., Haigha, T., Sommerauera, G., Spindlerb, M., Brunnera, C., Lyonscd, I. M., & Grabnera, R. H. (2017). Processing the Order of Symbolic Numbers: A Reliable and Unique Predictor of Arithmetic Fluency. *Journal of Numerical Cognition, 3*(2), 288-308.

Vogel, S. E., Remark, A., & Ansari, D. (2015). Differential processing of symbolic numerical magnitude and order in first-grade children. *Journal of Experimental Child Psychology, 129*, 26-39.

Vos, H., Sasanguie, D., Gevers, W., & Reynvoet, B. (2017). The role of general and number-specific order processing in adults' arithmetic performance. *Journal of Cognitive Psychology*, *29*(4), 469-482.

Voss, A., Nagler, M., & Lerche, V. (2013). Diffusion models in experimental psychology: A practical introduction. *Experimental Psychology, 60*(6), 385.

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the drift-diffusion model in Python. *Frontiers in Neuroinformatics, 7*, 14.

Zorzi, M., Di Bono, M. G., & Fias, W. (2011). Distinct representations of numerical and non-numerical order in the human intraparietal sulcus revealed by multivariate pattern recognition. *NeuroImage*, *56*(2), 674-680.

Appendix A

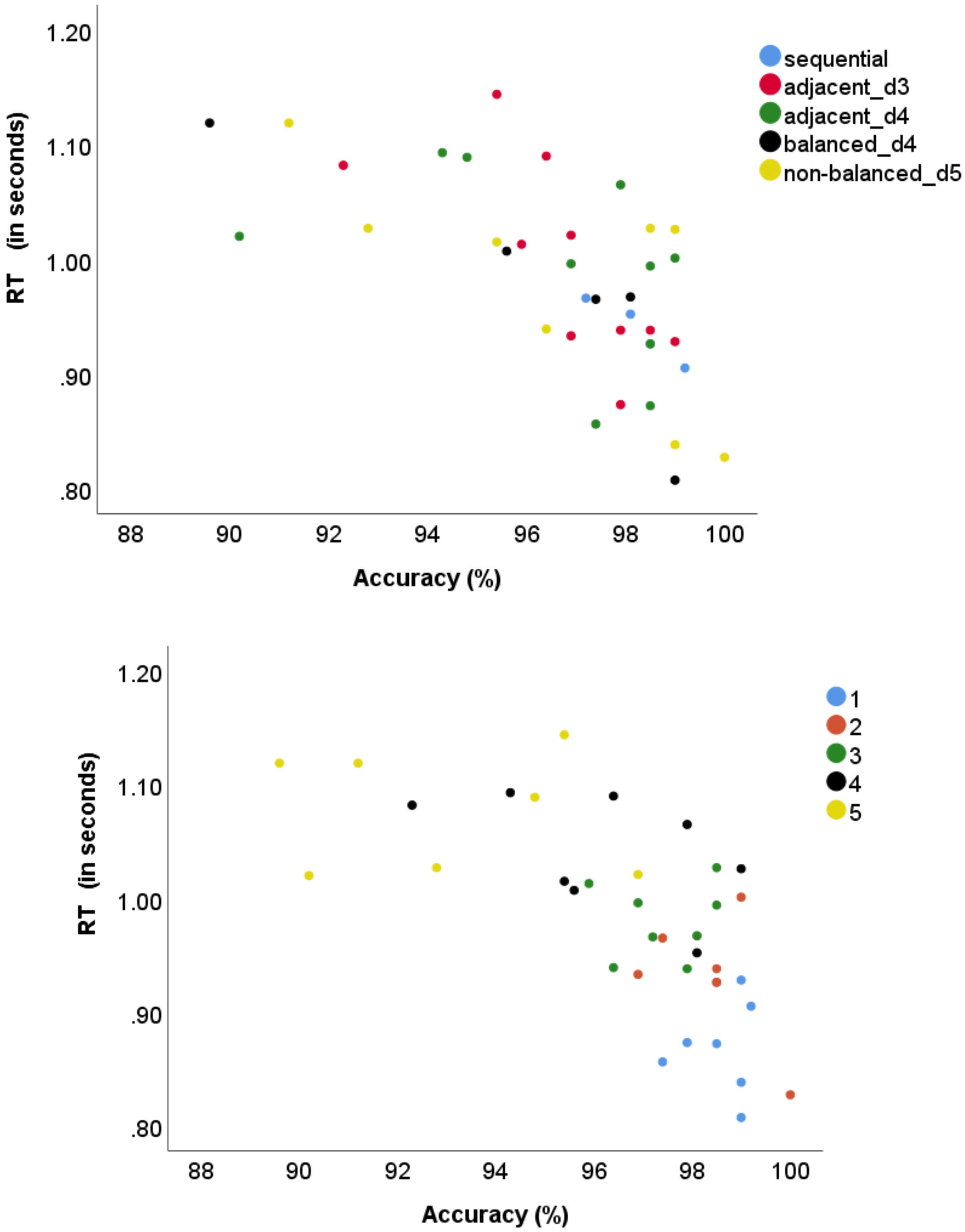
Table A1: Distribution (and number of triplets) in Model 2, 3, and 3b

Model 3/3b→ **Model 2**	**D_One** **(n=16)**	**D_Two** **(n=20)**	**D_Three** **(n=20)**	**D_Four** **(n=20**	**D_Five** **(n=16)**
Dist_Two **(n=16)**	-	234 (4)$_{g1}$	345 (4)$_{g2}$	456 (4)$_{g3}$	567 (4)$_{g4}$
Dist_Three **(n=20)**	124 (2)$_{g1}$ 134 (2)$_{g1}$	235 (2)$_{g2}$ 245 (2)$_{g2}$	346 (2)$_{g3}$ 356 (2)$_{g3}$	457 (2)$_{g4}$ 467 (2)$_{g4}$	568 (2)$_{g5}$ 578 (2)$_{g5}$
Dist_Four **(n=20)**	125 (2)$_{g1}$ 145 (2)$_{g1}$	236 (2)$_{g2}$ 256 (2)$_{g2}$	347 (2)$_{g3}$ 367 (2)$_{g3}$	458 (2)$_{g4}$ 478 (2)$_{g4}$	569 (2)$_{g5}$ 589 (2)$_{g5}$
Dist_Four (B) **(n=20)**	135 (4)$_{g1}$	246 (4)$_{g2}$	357 (4)$_{g3}$	468 (4)$_{g4}$	579 (4)$_{g5}$
Dist_Five **(n=16)**	136 (2)$_{g1}$ 146 (2)$_{g2}$	247 (2)$_{g3}$ 257 (2)$_{g3}$	358 (2)$_{g4}$ 368 (2)$_{g4}$	469 (2)$_{g5}$ 479 (2)$_{g5}$	-

Note: Subscripts indicate the distribution of items in Model 3b. The corresponding number size range is: g1 (7-10); g2 (10-12); g3 (13-15); g4 (16-18); g5 (19-21)

Appendix B

Figure B1: Response-accuracy function of correct ordered responses by type of trial. Data represent average accuracy and response time per trial across subjects.



Note: Top panel reflects the strength of symbol-symbol association (d3-d5 reflects the distance between numerals). The bottom panel reflects the position of numerals in the counting sequence (1 denotes trials with numerals closer to the lower bound)